

Nonlinear Dynamics, Stochastic Methods, And Predictive Modelling For Infectious Disease: Application To Public Health And Epidemic Forecasting

Christopher Daniel Prashad

A dissertation submitted to the Faculty of Graduate Studies
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Graduate Program in Mathematics and Statistics

YORK UNIVERSITY

December 2024

© 2024 Christopher Daniel Prashad. All rights reserved.

Nonlinear Dynamics, Stochastic Methods, And Predictive Modelling For Infectious Disease: Application To Public Health And Epidemic Forecasting

by

Christopher Daniel Prashad

A dissertation submitted to the Faculty of Graduate Studies
on December 9, 2024 in conformity with the requirements for the degree of

DOCTOR OF PHILOSOPHY

ABSTRACT

Statistical models must adapt to the evolving nature of many processes over time. This thesis introduces flexible models and statistical methods designed to infer data-generating processes that vary temporally. The primary objective is to develop frameworks for efficient estimation and prediction of both univariate and multivariate time series data. The models considered are general dynamic predictive models with parameters that change over time, featuring time-varying regression coefficients or variance components. These models are capable of accommodating time-dependent covariates and can handle situations where information is incomplete.

Several novel enhancements to existing mathematical models are introduced, with a particular focus on online learning and real-time prediction. Efficient Bayesian inference methodology is developed for analyzing the posterior of covariance components of dynamic models sequentially with a closed-form estimation algorithm for real-time online processing. Additionally, an online change detection algorithm for structural breaks is developed, combining the benefits of Kalman filters with sequential Monte Carlo methods.

A general and extensible compartmental model for the study of infectious disease data is proposed, with several innovative extensions to established probability models for the analysis of data. Next, we extend the classical SIRS (Susceptible–Infectious–Recovered–Susceptible) model by integrating innovative stochastic mean-reverting transmission processes to more accurately capture the variability observed in real-world epidemic data. Lastly, we provide a methodology that harnesses expansive data sources and feature engineering for analyzing and forecasting peak time and height of epidemic waves, crucial for the planning of public health strategies and interventions.

The performance of these inference methodologies is assessed through simulation experiments and real data from clinical, social-demographic, and epidemic domains.

Keywords: Dynamic regression models, Bayesian sequential inference, Stochastic simulation, Particle filter, Online prediction, Covariance estimation, Structural change detection, Disease transmission processes, Mean reversion, Stochastic threshold, Infectious disease modelling, Epidemic wave peak forecasting.

Thesis supervisor: Jianhong Wu

Title: Professor of Mathematics

Acknowledgments

This thesis is dedicated to the memory of Professor Hélène Massam (1949–2020), whose mentorship and passion for statistics profoundly shaped my academic journey. Hélène introduced me to advanced statistical theory and helped me navigate the rigorous mathematical foundations of the field, encouraging me to present my work with clarity and confidence. She had an extraordinary gift for turning intricate mathematical concepts into insights that felt both elegant and accessible. More than that, she fostered in me a genuine appreciation for the inherent beauty of mathematics and statistics. Her influence endures in my approach to problem-solving and in the curiosity that continues to fuel my work.

In addition to Hélène’s lasting impact, I have been fortunate to receive support from many others. I am deeply grateful to my supervisor, Professor Jianhong Wu, for his unwavering support and thoughtful guidance throughout my studies. His readiness to discuss new ideas, his encouragement to engage with the broader research community, and his calm, steady mentorship have been invaluable to the development of this thesis.

I extend my sincerest thanks to my committee members, Professor Neal Madras and Professor Xin Gao, whose constructive feedback, probing questions, and insightful suggestions not only sharpened my methodology but also broadened my perspective on the subject matter. Their careful reading and challenging commentary improved the intellectual rigour of my work. I am also thankful to my examiners, Professor Bouchra Nasri and Professor Melanie Cao, for their thorough review and thoughtful recommendations. Their critical examination ensured that this thesis stands on a solid scholarly foundation.

Finally, I owe my greatest debt of gratitude to my family: my mother, June; my father, Daniel; and my brothers, Sean and Andrew. Your love, patience, and faith in my abilities laid the foundation that made this achievement possible. I could not have reached this milestone without your steadfast support.

Table of Contents

Title page	i
Abstract	ii
Acknowledgments	iv
Table of Contents	v
List of Tables	ix
List of Figures	xii
1 Introduction	1
1.1 Background	1
1.2 Objectives and main contributions	2
1.3 Outline of the thesis	2
1.4 Exploratory methods	3
1.5 Bayesian learning	5
2 State-Space Modelling for Infectious Disease Surveillance Data: Dynamic Regression and Covariance Analysis	7
2.1 Introduction	7
2.2 Methodology	8
2.2.1 State-space model	8
2.2.2 State estimation	13
2.2.3 Sequential solution of the linear Gaussian state-space model	21
2.2.4 Component linear Gaussian state-space model	25
2.2.5 Dynamic regression models	28
2.2.6 Covariance estimation in multivariate state-space models	28
2.3 Illustrations using COVID-19 data	37

2.3.1	Application of component linear Gaussian state-space model to case surveillance data	37
2.3.2	Dynamic regression models	41
2.3.3	Application of the covariance estimation method to public health surveillance data	51
2.4	Discussion	65
2.5	Conclusion	66
3	State-Space Modelling for Infectious Disease Surveillance Data: Stochastic Simulation Techniques and Structural Change Detection	68
3.1	Introduction	68
3.2	Methodology	69
3.2.1	Markov chain Monte Carlo method	69
3.2.2	Sequential Monte Carlo method	74
3.2.3	Online structural change detection	80
3.3	Applications to understanding COVID-19 data	82
3.3.1	Bayesian inference for nonlinear dynamic regression models	84
3.3.2	Application of online structural change detection method	95
3.4	Discussion	103
3.5	Conclusion	107
4	State-Space Model for Epidemic Disease: Assessing the Impact of Vaccination and Behavioural Effects on COVID-19 Dynamics	109
4.1	Introduction	109
4.2	Methodology	110
4.2.1	Compartmental model	110
4.2.2	Mathematical analysis	113
4.2.3	Probabilistic observation model	126
4.2.4	Non-parametric Bayesian basis expansion model	130
4.2.5	Markov chain Monte Carlo implementation	137
4.3	Applications to COVID-19 mobility and vaccination data	140
4.4	Conclusion	155
5	Stochastic Threshold Analysis for SIRS model with Mean-Reverting Transmission Processes	158
5.1	Introduction	158
5.2	Methodology	159

5.2.1	Model overview	159
5.2.2	Key contributions	159
5.2.3	Stochastic model of disease transmission	161
5.2.4	The SIRS model	164
5.2.5	Reproduction number of the deterministic model	166
5.2.6	Theoretical foundations	167
5.3	Derivation of the stochastic threshold for mean-reverting processes	170
5.3.1	Stochastic integral and variance analysis	170
5.3.2	Derivation strategy for extinction threshold	172
5.3.3	Long-term behaviour	172
5.3.4	Simplification using equilibrium values	173
5.3.5	Deriving the extinction threshold	173
5.4	Preliminary results for CKLS process	174
5.4.1	CKLS process and Fokker–Planck equation	174
5.4.2	Stationary distribution of the CKLS process	175
5.4.3	Distribution from Girsanov transform	176
5.4.4	Expectation calculation for the CKLS process	178
5.5	Stochastic threshold for the CKLS process	179
5.5.1	Stochastic threshold for the CIR process	180
5.6	Additional mean-reverting diffusion processes	180
5.6.1	Mean-reverting Bessel process	180
5.6.2	Modified Bessel process with power-law drift	184
5.6.3	Logistic GBM process	189
5.6.4	Jacobi process	192
5.7	Numerical results	195
5.7.1	Parameter sets for numerical simulations	195
5.7.2	Deterministic and noisy synthetic data generation	199
5.7.3	Simulation and plotting of model fit	201
5.7.4	Illustrations of the determination of eradication/extinction or endemic- ity/persistence based on model type	204
5.7.5	Comparison of infected behaviour of deterministic and stochastic models	211
5.7.6	Transmission process simulations	217
5.7.7	Feasible extinction region analysis	221
5.8	Discussion	221
5.9	Conclusion	222

6	Multiperiod Forecasting of Epidemic Wave Peak Time and Height	224
6.1	Introduction	225
6.2	Methodology	233
6.2.1	Data collection and preprocessing	233
6.2.2	Modelling approaches	234
6.2.3	Methodology	235
6.2.4	Model training and evaluation	236
6.3	Applications to understanding the relationship between influenza case counts in the Southern and Northern hemispheres	238
6.4	Discussion	258
6.5	Conclusion	259
7	Conclusion and Future Directions	261
7.1	Novelty of the work	261
7.2	Limitations and future studies	263
7.3	Incorporating behavioural changes and public health interventions	263
7.4	Conclusion	264
A		266
A.1	Additional figures for Chapter 5	266
A.2	Key mathematical concepts for Chapter 5	310
A.2.1	Fokker–Planck equation	310
A.2.2	Brownian motion	310
A.2.3	Itô calculus	311
A.2.4	Martingales and their properties	312
A.2.5	Doob’s h -transform and its applications to stochastic processes	312
A.3	Additional figures for Chapter 6	314
	References	339

List of Tables

2.1	Model comparison via log-likelihood, MAPE and Theil's U .	41
2.2	Static sample correlations.	54
2.3	Model 1 summarized cross-correlations.	55
2.4	Comparison of Models 1 and 2 by period.	59
2.5	Comparison between the models by period.	64
3.1	Parameter estimate statistics for the inference approach with MCMC state sampling.	83
3.2	Parameter estimate statistics for the inference approach without MCMC state sampling.	84
3.3	MCMC posterior expectation approximations.	90
3.4	DFM parameter estimates.	92
3.5	Log-likelihood of the models considered.	96
4.1	Informing model parameters using clinical knowledge [67].	133
4.2	Point prediction of the major epidemic peak time and intensity.	148
5.1	Estimated parameters for different processes in Sets A and B across cases.	209
5.2	Estimated extended parameters for the Bessel and modified Bessel processes in Sets A and B.	211
5.3	Merged parameter settings for Set A and Set B with calculated $\mathcal{R}_0$ and stochastic thresholds for OU, CKLS, Logistic GBM, and Jacobi processes.	216
5.4	Empirical means and standard deviations for Set A and Set B across different processes	220
6.1	Model Performance Results for 2013-2014 Season including Southern Hemisphere covariate.	244
6.2	Model Performance Results for 2013-2014 Season excluding Southern Hemisphere covariate.	244

6.3	Model Performance Results for 2014-2015 Season including Southern Hemisphere covariate.	245
6.4	Model Performance Results for 2014-2015 Season excluding Southern Hemisphere covariate.	245
6.5	Model Performance Results for 2015-2016 Season including Southern Hemisphere covariate.	246
6.6	Model Performance Results for 2015-2016 Season excluding Southern Hemisphere covariate.	246
6.7	Model Performance Results for 2016-2017 Season including Southern Hemisphere covariate.	247
6.8	Model Performance Results for 2016-2017 Season excluding Southern Hemisphere covariate.	247
6.9	Model Performance Results for 2017-2018 Season including Southern Hemisphere covariate.	248
6.10	Model Performance Results for 2017-2018 Season excluding Southern Hemisphere covariate.	248
6.11	Model Performance Results for 2018-2019 Season including Southern Hemisphere covariate.	249
6.12	Model Performance Results for 2018-2019 Season excluding Southern Hemisphere covariate.	249
6.13	Model Performance Results for 2019-2020 Season including Southern Hemisphere covariate.	250
6.14	Model Performance Results for 2019-2020 Season excluding Southern Hemisphere covariate.	250
6.15	Performance Ranked by Prediction Error – 2013-14 including Southern Hemisphere covariate.	251
6.16	Performance Ranked by Prediction Error – 2013-14 excluding Southern Hemisphere covariate.	251
6.17	Performance Ranked by Prediction Error – 2014-15 including Southern Hemisphere covariate.	251
6.18	Performance Ranked by Prediction Error – 2014-15 excluding Southern Hemisphere covariate.	252
6.19	Performance Ranked by Prediction Error – 2015-16 including Southern Hemisphere covariate.	252
6.20	Performance Ranked by Prediction Error – 2015-16 excluding Southern Hemisphere covariate.	252

6.21	Performance Ranked by Prediction Error – 2016-17 including Southern Hemisphere covariate.	253
6.22	Performance Ranked by Prediction Error – 2016-17 excluding Southern Hemisphere covariate.	253
6.23	Performance Ranked by Prediction Error – 2017-18 including Southern Hemisphere covariate.	253
6.24	Performance Ranked by Prediction Error – 2017-18 excluding Southern Hemisphere covariate.	254
6.25	Performance Ranked by Prediction Error – 2018-19 including Southern Hemisphere covariate.	254
6.26	Performance Ranked by Prediction Error – 2018-19, Southern Hemisphere covariate excluded.	254
6.27	Performance Ranked by Prediction Error – 2019-20 including Southern Hemisphere covariate.	255
6.28	Performance Ranked by Prediction Error – 2019-20 excluding Southern Hemisphere covariate.	255
6.29	Growth metrics for the 12 weeks prior to the peak for each season.	257

List of Figures

2.1	DAG of state-space model.	8
2.2	Conditional independence of model states.	9
2.3	Conditional independence of observations.	9
2.4	Recursion diagram for posterior distribution $p(\theta_{t'} y_{1:t}) = (t' t)$. The calculation flow for the recursion is reached by moving forward in time denoted by right and down arrows.	17
2.5	Recursion processing for filtering distribution $p(\theta_t y_{1:t})$. Filtering is completed by first obtaining the one-step-ahead predictive distribution through the state transition from the filtering distribution one time before and then secondly correcting the obtained distribution based on the observations. The conceptual diagram shows the calculation follows a stepwise pattern.	18
2.6	Recursion processing for predictive distribution $p(\theta_{t+k} y_{1:t})$. The k -steps-ahead prediction can be achieved by repeating the state transition k times for the filtering distribution. The processing to obtain three-steps-ahead prediction from the current time point $t = 1$ is represented by repeating the right arrow thrice.	18
2.7	Recursion processing for smoothing distribution $p(\theta_t y_{1:T})$. At $T = 3$, the processing to obtain the smoothing distribution at $t = 0$ is represented by repeating the left arrow operation four times.	19
2.8	Decomposition of time series data.	27
2.9	COVID-19 cases of Ontario, Canada. The data are sampled on a daily basis and shown on a $\log_e$ scale. [21]	38
2.10	Local trend + periodic component plots.	39
2.11	Local trend + periodic model validation and diagnostic plots.	39
2.12	Local trend + periodic model Kalman prediction 14-day forecast.	40
2.13	Kalman prediction 14-day forecast are compared for two competing models (A) and (B).	41

2.14	Ontario COVID-19 hospitalizations and cases.[21]	42
2.15	Ontario COVID-19 hospitalizations and cases filtering result. The explanatory and dependent variables are shown in the first two panels while model level, seasonal, autoregressive and regression components are shown in the subsequent panels.	44
2.16	Time-varying regression coefficient β_t plot as defined in θ_t^{reg} .	45
2.17	Thresholded β_t plot.	45
2.18	Time-varying correlation plot.	46
2.19	Cross-correlation function between cases and hospitalizations plot.	46
2.20	Scatterplots useful for determining lag number.	47
2.21	COVID-19 cases by workday.	47
2.22	Percent change in cases associated with a weekend or holiday.	48
2.23	Standardized concentration of SARS-CoV-2 gene copies [22] and case count [21] plot.	49
2.24	Smoothing results.	50
2.25	Model diagnostics.	50
2.26	Cases, hospitalizations and deaths related to COVID-19.[21]	51
2.27	Viral load, temperature and air quality data.[22]	52
2.28	Estimated Model 1 time-varying variances.	53
2.29	Estimated Model 1 time-varying cross-correlations.	54
2.30	Estimated Model 2 time-varying variances.	58
2.31	Estimated Model 2 time-varying cross-correlations.	59
2.32	Estimated Model 3 time-varying model variances.	63
2.33	Estimated Model 3 time-varying cross-correlations.	64
3.1	Graphical depiction of the bootstrap filter.	76
3.2	Shrinkage probability distributions.	82
3.3	Local level and seasonal model for COVID-19 cases.	83
3.4	MCMC traceplots of the state noise variance of the level (W_μ) and seasonal components (W_γ) and observation noise variance V . State sampling by MCMC.	83
3.5	MCMC traceplots of the state noise variance of the level (W_μ) and seasonal components (W_γ) and observation noise variance V . State sampling by FFBS.	84
3.6	Local level and seasonal model for COVID-19 cases results. State sampling by FFBS.	85
3.7	Percent positivity by age over time.	86

3.8	Alternative view of percent positivity by age over time after the large epidemic wave.	87
3.9	OLS fitted curves at selected dates.	88
3.10	Autocorrelation (ACF) diagnostic plots.	89
3.11	Comparison of MCMC and OLS smoothing estimates.	90
3.12	Case counts and hospitalizations.	91
3.13	MCMC diagnostics for the DFM.	92
3.14	Original data with common stochastic trend.	93
3.15	Posterior means of the log volatility.	94
3.16	Original data series for the MSV model application.	94
3.17	Simulated data set for change detection.	95
3.18	Normal distribution filter results.	96
3.19	t -distribution filter results.	97
3.20	Laplace distribution filter results.	97
3.21	Horseshoe distribution filter results.	98
3.22	Parameter estimate for normal distribution model.	98
3.23	Parameter estimate for t -distribution model.	99
3.24	Parameter estimate for Laplace distribution model.	99
3.25	Parameter estimate for horseshoe distribution model.	100
3.26	ESS for normal distribution model.	101
3.27	ESS for t -distribution model.	101
3.28	ESS for Laplace distribution model.	102
3.29	ESS for horseshoe distribution model.	102
3.30	Parameter filtering and smoothing results.	105
3.31	Change points identified via τ_t thresholding.	105
3.32	The result of change detection using PELT algorithm/binary segmentation. .	106
3.33	Posterior means and probabilities of change points using Bayesian change point method.	107
4.1	Flow chart of the base model.	110
4.2	Flow chart of the model with demography.	113
4.3	Shrinkage priors for the Bayesian spline functions	132
4.4	Timeline of COVID-19 progression and model parameters.	132
4.5	Flow chart of the model with covariate inclusion.	134
4.6	Vaccination rate over time.	141
4.7	Ontario mobility trends.	142

4.8	Prediction of case counts with the original Poisson model. The analysis period is taken to be 12/14/2020 to 09/16/2022.	144
4.9	Prediction of case counts with the DLGNB model. The analysis period is taken to be 12/14/2020 to 09/16/2022.	144
4.10	Prediction of case counts with accounted vaccination using the Poisson model.	146
4.11	Prediction of case counts with accounted vaccination using the DLGNB model.	146
4.12	Prediction of hospitalization counts with accounted vaccination using the DLGNB model.	149
4.13	Prediction of death counts with accounted vaccination using the DLGNB model.	150
4.14	$\mathcal{R}_0$ from the Ontario COVID-19 Science Advisory Table.	151
4.15	Model $\mathcal{R}_0$ including mobility and vaccination data.	152
4.16	Case count prediction for the DLGNB model with a setting of 75% vaccine effectiveness.	153
4.17	Case count prediction for the DLGNB model with a setting of 90% vaccine effectiveness.	153
4.18	Reproduction number computed using the DLGNB model with a setting of 75% vaccine effectiveness.	154
4.19	Reproduction number computed using the DLGNB model with a setting of 90% vaccine effectiveness.	155
5.1	Simulation demonstrating the possibility of negative transmission rates if OU process is used. The parameter setting used are as follows: $\beta_e = 0.15$, $\theta = 0.4$, $\rho = 0.25$. The CKLS process is determined by $\gamma = \frac{1}{2}$	163
5.2	A plot of the deterministic SIRS model simulation for Set A.	199
5.3	A plot of the deterministic SIRS model simulation for Set B.	200
5.4	A plot of the synthetic data for parameter Set A.	201
5.5	A plot of the synthetic data for parameter Set B.	201
5.6	A plot of the original synthetic infected data along with the estimated SIRS model with OU transmission process. The data set and case number are A and 1, respectively.	204
5.7	A plot of the original synthetic infected data along with the estimated SIRS model with OU transmission process. The data set and case number are A and 2, respectively.	205
5.8	A plot of the original synthetic infected data along with the estimated SIRS model with OU transmission process. The data set and case number are A and 3, respectively.	205

5.9	A plot of the original synthetic infected data along with the estimated SIRS model with OU transmission process. The data set and case number are A and 4, respectively.	206
5.10	A plot of the original synthetic infected data along with the estimated SIRS model with CKLS transmission process. The data set and case number are A and 1, respectively.	206
5.11	A plot of the original synthetic infected data along with the estimated SIRS model with CKLS transmission process. The data set and case number are A and 2, respectively.	207
5.12	A plot of the original synthetic infected data along with the estimated SIRS model with CKLS transmission process. The data set and case number are A and 3, respectively.	207
5.13	A plot of the original synthetic infected data along with the estimated SIRS model with CKLS transmission process. The data set and case number are A and 4, respectively.	208
5.14	A plot of the deterministic solution along with the simulated SIRS model with the OU transmission process. The data set is A and the case number is 1. . .	212
5.15	A plot of the deterministic solution along with the simulated SIRS model with the OU transmission process. The data set is A and the case number is 2. . .	212
5.16	A plot of the deterministic solution along with the simulated SIRS model with the OU transmission process. The data set is A and the case number is 3. . .	213
5.17	A plot of the deterministic solution along with the simulated SIRS model with the OU transmission process. The data set is A and the case number is 4. . .	213
5.18	A plot of the deterministic solution along with the simulated SIRS model with the CKLS transmission process. The data set is A and the case number is 1. . .	214
5.19	A plot of the deterministic solution along with the simulated SIRS model with the CKLS transmission process. The data set is A and the case number is 2. . .	214
5.20	A plot of the deterministic solution along with the simulated SIRS model with the CKLS transmission process. The data set is A and the case number is 3. . .	215
5.21	A plot of the deterministic solution along with the simulated SIRS model with the CKLS transmission process. The data set is A and the case number is 4. . .	215
5.22	Plots of the transmission processes for Set A. Top row, from left to right: mean-reverting Bessel process, mean-reverting power law Bessel process, OU process; Bottom row, from left to right: CKLS process, Logistic GBM process, Jacobi process.	218

5.23	Plots of the transmission processes for Set B, Case 1. Top row, from left to right: mean-reverting Bessel process, mean-reverting power law Bessel process, OU process; Bottom row, from left to right: CKLS process, Logistic GBM process, Jacobi process.	219
6.1	Different fitting situations: (a) Underfitting, (b) Appropriate fitting, (c) Overfitting.	226
6.2	WHO FluNet Hemisphere Data Sets [111]: The top panel shows the natural logarithm of the actual counts of influenza viruses detected by hemisphere (data from the Global Influenza Programme). The bottom panel displays the dynamic rolling autocorrelation values of the data with a window size of 20.	238
6.3	Prediction for 2019-2020 Season by Linear Regression without Southern Hemisphere covariate.	239
6.4	Prediction for 2019-2020 Season by Linear Regression including Southern Hemisphere covariate.	240
6.5	2019-2020 Season Prediction Focused on Last 10% of Training Data without Southern Hemisphere covariate.	241
6.6	2019-2020 Season Prediction Focused on Last 10% of Training Data including Southern Hemisphere covariate.	242
6.7	2019-2020 Season Prediction on Test Set without Southern Hemisphere covariate.	243
6.8	2019-2020 Season Prediction on Test Set including Southern Hemisphere covariate.	243
A.1	A plot of the original synthetic infected data along with the estimated SIRS model with LogisticGBM transmission process. The data set and case number are A and 1, respectively.	266
A.2	A plot of the original synthetic infected data along with the estimated SIRS model with LogisticGBM transmission process. The data set and case number are A and 2, respectively.	267
A.3	A plot of the original synthetic infected data along with the estimated SIRS model with LogisticGBM transmission process. The data set and case number are A and 3, respectively.	267
A.4	A plot of the original synthetic infected data along with the estimated SIRS model with LogisticGBM transmission process. The data set and case number are A and 4, respectively.	268

A.5	A plot of the original synthetic infected data along with the estimated SIRS model with Jacobi transmission process. The data set and case number are A and 1, respectively.	268
A.6	A plot of the original synthetic infected data along with the estimated SIRS model with Jacobi transmission process. The data set and case number are A and 2, respectively.	269
A.7	A plot of the original synthetic infected data along with the estimated SIRS model with Jacobi transmission process. The data set and case number are A and 3, respectively.	269
A.8	A plot of the original synthetic infected data along with the estimated SIRS model with Jacobi transmission process. The data set and case number are A and 4, respectively.	270
A.9	A plot of the original synthetic infected data along with the estimated SIRS model with PowerLawBessel transmission process. The data set and case number are A and 1, respectively.	270
A.10	A plot of the original synthetic infected data along with the estimated SIRS model with PowerLawBessel transmission process. The data set and case number are A and 2, respectively.	271
A.11	A plot of the original synthetic infected data along with the estimated SIRS model with PowerLawBessel transmission process. The data set and case number are A and 3, respectively.	271
A.12	A plot of the original synthetic infected data along with the estimated SIRS model with PowerLawBessel transmission process. The data set and case number are A and 4, respectively.	272
A.13	A plot of the original synthetic infected data along with the estimated SIRS model with OU transmission process. The data set and case number are B and 1, respectively.	272
A.14	A plot of the original synthetic infected data along with the estimated SIRS model with OU transmission process. The data set and case number are B and 2, respectively.	273
A.15	A plot of the original synthetic infected data along with the estimated SIRS model with OU transmission process. The data set and case number are B and 3, respectively.	273
A.16	A plot of the original synthetic infected data along with the estimated SIRS model with OU transmission process. The data set and case number are B and 4, respectively.	274

A.17 A plot of the original synthetic infected data along with the estimated SIRS model with CKLS transmission process. The data set and case number are B and 1, respectively.	274
A.18 A plot of the original synthetic infected data along with the estimated SIRS model with CKLS transmission process. The data set and case number are B and 2, respectively.	275
A.19 A plot of the original synthetic infected data along with the estimated SIRS model with CKLS transmission process. The data set and case number are B and 3, respectively.	275
A.20 A plot of the original synthetic infected data along with the estimated SIRS model with CKLS transmission process. The data set and case number are B and 4, respectively.	276
A.21 A plot of the original synthetic infected data along with the estimated SIRS model with LogisticGBM transmission process. The data set and case number are B and 1, respectively.	276
A.22 A plot of the original synthetic infected data along with the estimated SIRS model with LogisticGBM transmission process. The data set and case number are B and 2, respectively.	277
A.23 A plot of the original synthetic infected data along with the estimated SIRS model with LogisticGBM transmission process. The data set and case number are B and 3, respectively.	277
A.24 A plot of the original synthetic infected data along with the estimated SIRS model with LogisticGBM transmission process. The data set and case number are B and 4, respectively.	278
A.25 A plot of the original synthetic infected data along with the estimated SIRS model with Jacobi transmission process. The data set and case number are B and 1, respectively.	278
A.26 A plot of the original synthetic infected data along with the estimated SIRS model with Jacobi transmission process. The data set and case number are B and 2, respectively.	279
A.27 A plot of the original synthetic infected data along with the estimated SIRS model with Jacobi transmission process. The data set and case number are B and 3, respectively.	279
A.28 A plot of the original synthetic infected data along with the estimated SIRS model with Jacobi transmission process. The data set and case number are B and 4, respectively.	280

A.29	A plot of the original synthetic infected data along with the estimated SIRS model with PowerLawBessel transmission process. The data set and case number are B and 1, respectively.	280
A.30	A plot of the original synthetic infected data along with the estimated SIRS model with PowerLawBessel transmission process. The data set and case number are B and 2, respectively.	281
A.31	A plot of the original synthetic infected data along with the estimated SIRS model with PowerLawBessel transmission process. The data set and case number are B and 3, respectively.	281
A.32	A plot of the original synthetic infected data along with the estimated SIRS model with PowerLawBessel transmission process. The data set and case number are B and 4, respectively.	282
A.33	A plot of the deterministic solution along with the simulated SIRS model with the OU transmission process. The data set is B and the case number is 1. . .	282
A.34	A plot of the deterministic solution along with the simulated SIRS model with the OU transmission process. The data set is B and the case number is 2. . .	283
A.35	A plot of the deterministic solution along with the simulated SIRS model with the OU transmission process. The data set is B and the case number is 3. . .	283
A.36	A plot of the deterministic solution along with the simulated SIRS model with the OU transmission process. The data set is B and the case number is 4. . .	284
A.37	A plot of the deterministic solution along with the simulated SIRS model with the CKLS transmission process. The data set is B and the case number is 1.	284
A.38	A plot of the deterministic solution along with the simulated SIRS model with the CKLS transmission process. The data set is B and the case number is 2.	285
A.39	A plot of the deterministic solution along with the simulated SIRS model with the CKLS transmission process. The data set is B and the case number is 3.	285
A.40	A plot of the deterministic solution along with the simulated SIRS model with the CKLS transmission process. The data set is B and the case number is 4.	286
A.41	A plot of the deterministic solution along with the simulated SIRS model with the Bessel transmission process. The data set is A and the case number is 1.	286
A.42	A plot of the deterministic solution along with the simulated SIRS model with the Bessel transmission process. The data set is A and the case number is 2.	287
A.43	A plot of the deterministic solution along with the simulated SIRS model with the Bessel transmission process. The data set is A and the case number is 3.	287
A.44	A plot of the deterministic solution along with the simulated SIRS model with the Bessel transmission process. The data set is A and the case number is 4.	288

A.45 A plot of the deterministic solution along with the simulated SIRS model with the Bessel transmission process. The data set is B and the case number is 1.	288
A.46 A plot of the deterministic solution along with the simulated SIRS model with the Bessel transmission process. The data set is B and the case number is 2.	289
A.47 A plot of the deterministic solution along with the simulated SIRS model with the Bessel transmission process. The data set is B and the case number is 3.	289
A.48 A plot of the deterministic solution along with the simulated SIRS model with the Bessel transmission process. The data set is B and the case number is 4.	290
A.49 A plot of the deterministic solution along with the simulated SIRS model with the PowerLawBessel transmission process. The data set is A and the case number is 1.	290
A.50 A plot of the deterministic solution along with the simulated SIRS model with the PowerLawBessel transmission process. The data set is A and the case number is 2.	291
A.51 A plot of the deterministic solution along with the simulated SIRS model with the PowerLawBessel transmission process. The data set is A and the case number is 3.	291
A.52 A plot of the deterministic solution along with the simulated SIRS model with the PowerLawBessel transmission process. The data set is A and the case number is 4.	292
A.53 A plot of the deterministic solution along with the simulated SIRS model with the PowerLawBessel transmission process. The data set is B and the case number is 1.	292
A.54 A plot of the deterministic solution along with the simulated SIRS model with the PowerLawBessel transmission process. The data set is B and the case number is 2.	293
A.55 A plot of the deterministic solution along with the simulated SIRS model with the PowerLawBessel transmission process. The data set is B and the case number is 3.	293
A.56 A plot of the deterministic solution along with the simulated SIRS model with the PowerLawBessel transmission process. The data set is B and the case number is 4.	294
A.57 A plot of the deterministic solution along with the simulated SIRS model with the LogisticGBM transmission process. The data set is A and the case number is 1.	294

A.58 A plot of the deterministic solution along with the simulated SIRS model with the LogisticGBM transmission process. The data set is A and the case number is 2.	295
A.59 A plot of the deterministic solution along with the simulated SIRS model with the LogisticGBM transmission process. The data set is A and the case number is 3.	295
A.60 A plot of the deterministic solution along with the simulated SIRS model with the LogisticGBM transmission process. The data set is A and the case number is 4.	296
A.61 A plot of the deterministic solution along with the simulated SIRS model with the LogisticGBM transmission process. The data set is B and the case number is 1.	296
A.62 A plot of the deterministic solution along with the simulated SIRS model with the LogisticGBM transmission process. The data set is B and the case number is 2.	297
A.63 A plot of the deterministic solution along with the simulated SIRS model with the LogisticGBM transmission process. The data set is B and the case number is 3.	297
A.64 A plot of the deterministic solution along with the simulated SIRS model with the LogisticGBM transmission process. The data set is B and the case number is 4.	298
A.65 A plot of the deterministic solution along with the simulated SIRS model with the Jacobi transmission process. The data set is A and the case number is 1.	298
A.66 A plot of the deterministic solution along with the simulated SIRS model with the Jacobi transmission process. The data set is A and the case number is 2.	299
A.67 A plot of the deterministic solution along with the simulated SIRS model with the Jacobi transmission process. The data set is A and the case number is 3.	299
A.68 A plot of the deterministic solution along with the simulated SIRS model with the Jacobi transmission process. The data set is A and the case number is 4.	300
A.69 A plot of the deterministic solution along with the simulated SIRS model with the Jacobi transmission process. The data set is B and the case number is 1.	300
A.70 A plot of the deterministic solution along with the simulated SIRS model with the Jacobi transmission process. The data set is B and the case number is 2.	301
A.71 A plot of the deterministic solution along with the simulated SIRS model with the Jacobi transmission process. The data set is B and the case number is 3.	301

A.72 A plot of the deterministic solution along with the simulated SIRS model with the Jacobi transmission process. The data set is B and the case number is 4.	302
A.73 Plots of the transmission processes for Set B, Case 2. Top row, from left to right: mean-reverting Bessel process, mean-reverting power law Bessel process, OU process; Bottom row, from left to right: CKLS process, Logistic GBM process, Jacobi process.	303
A.74 Plots of the transmission processes for Set B, Case 3. Top row, from left to right: mean-reverting Bessel process, mean-reverting power law Bessel process, OU process; Bottom row, from left to right: CKLS process, Logistic GBM process, Jacobi process.	304
A.75 Plots of the transmission processes for Set A. From left to right for Set B, Case 4. Top row, from left to right: mean-reverting Bessel process, mean-reverting power law Bessel process, OU process; Bottom row, from left to right: CKLS process, Logistic GBM process, Jacobi process.	305
A.76 Feasible region plot for the OU process in Set A, Case 1.	306
A.77 Feasible region plot for the CIR process in Set A, Case 1.	306
A.78 Feasible region plot for the OU process in Set B, Case 1.	307
A.79 Feasible region plot for the CIR process in Set B, Case 1.	307
A.80 Feasible region plot for the OU process in Set B, Case 2.	308
A.81 Feasible region plot for the CIR process in Set B, Case 2.	308
A.82 Feasible region plot for the OU process in Set B, Case 4.	309
A.83 Feasible region plot for the CIR process in Set B, Case 4.	309
A.84 Prediction for 2018-2019 Season by Linear Regression without Southern Hemisphere Covariate.	314
A.85 Prediction for 2018-2019 Season by Linear Regression including Southern Hemisphere Covariate.	315
A.86 2018-2019 Season Prediction Focused on Last 10% of Training Data without Southern Hemisphere Covariate.	316
A.87 2018-2019 Season Prediction Focused on Last 10% of Training Data including Southern Hemisphere Covariate.	317
A.88 2018-2019 Season Prediction on Test Set without Southern Hemisphere Covariate.	318
A.89 2018-2019 Season Prediction on Test Set including Southern Hemisphere Covariate.	318
A.90 Prediction for 2017-2018 Season by Linear Regression without Southern Hemisphere Covariate.	319

A.91 Prediction for 2017-2018 Season by Linear Regression including Southern Hemisphere Covariate.	319
A.92 2017-2018 Season Prediction Focused on Last 10% of Training Data without Southern Hemisphere Covariate.	320
A.93 2017-2018 Season Prediction Focused on Last 10% of Training Data including Southern Hemisphere Covariate.	321
A.94 2017-2018 Season Prediction on Test Set without Southern Hemisphere Covariate.	322
A.95 2017-2018 Season Prediction on Test Set including Southern Hemisphere Covariate.	322
A.96 Prediction for 2016-2017 Season by Linear Regression without Southern Hemisphere Covariate.	323
A.97 Prediction for 2016-2017 Season by Linear Regression including Southern Hemisphere Covariate.	323
A.98 2016-2017 Season Prediction Focused on Last 10% of Training Data without Southern Hemisphere Covariate.	324
A.99 2016-2017 Season Prediction Focused on Last 10% of Training Data including Southern Hemisphere Covariate.	325
A.100 2016-2017 Season Prediction on Test Set without Southern Hemisphere Covariate.	326
A.101 2016-2017 Season Prediction on Test Set including Southern Hemisphere Covariate.	326
A.102 Prediction for 2015-2016 Season by Linear Regression without Southern Hemisphere Covariate.	327
A.103 Prediction for 2015-2016 Season by Linear Regression including Southern Hemisphere Covariate.	327
A.104 2015-2016 Season Prediction Focused on Last 10% of Training Data without Southern Hemisphere Covariate.	328
A.105 2015-2016 Season Prediction Focused on Last 10% of Training Data including Southern Hemisphere Covariate.	329
A.106 2015-2016 Season Prediction on Test Set without Southern Hemisphere Covariate.	330
A.107 2015-2016 Season Prediction on Test Set including Southern Hemisphere Covariate.	330
A.108 Prediction for 2014-2015 Season by Linear Regression without Southern Hemisphere Covariate.	331

A.10	Prediction for 2014-2015 Season by Linear Regression including Southern Hemisphere Covariate.	331
A.11	2014-2015 Season Prediction Focused on Last 10% of Training Data without Southern Hemisphere Covariate.	332
A.11	2014-2015 Season Prediction Focused on Last 10% of Training Data including Southern Hemisphere Covariate.	333
A.11	2014-2015 Season Prediction on Test Set without Southern Hemisphere Covariate.	334
A.11	2014-2015 Season Prediction on Test Set including Southern Hemisphere Covariate.	334
A.11	4 Prediction for 2013-2014 Season by Linear Regression without Southern Hemisphere Covariate.	335
A.11	5 Prediction for 2013-2014 Season by Linear Regression including Southern Hemisphere Covariate.	335
A.11	6 2013-2014 Season Prediction Focused on Last 10% of Training Data without Southern Hemisphere Covariate.	336
A.11	7 2013-2014 Season Prediction Focused on Last 10% of Training Data including Southern Hemisphere Covariate.	337
A.11	8 2013-2014 Season Prediction on Test Set without Southern Hemisphere Covariate.	338
A.11	9 2013-2014 Season Prediction on Test Set including Southern Hemisphere Covariate.	338

Chapter 1

Introduction

1.1 Background

Predictive models are fundamental tools for understanding the uncertainty associated with unknown events, thereby playing a crucial role in decision-making processes. For example, individuals often use traffic reports to determine their commuting plans. If heavy congestion is expected, they might choose public transportation, an alternative route or leave earlier to avoid delays. This dependence on predictions is equally critical across various application domains.

In clinical settings, predictive models help in understanding the likely progression of diseases and in tailoring personalized treatments based on the effects of medical interventions. In industrial applications, these models are used to assess product reliability, guiding decisions in the production process such as determining optimal release times. Social-demographic studies also benefit from predictive models by explaining life events and their trajectories through time-varying covariates, such as educational levels.

Statistical models, which are the most commonly used predictive models, represent the probability distribution of a response variable as a function of explanatory input variables. These explanatory variables, also known as covariates, regressors, or features, provide essential background information for describing the response variable. Dynamic regression models, in particular, are employed where the predictive density function of the response variable evolves over time. These dynamics can arise from time-varying covariates, parameters, or the functional form of the response's probability distribution.

A Bayesian methodology for inference in dynamic predictive models offers a flexible approach by treating parameters as random variables and incorporating uncertainty through prior distributions. This approach allows for integrating expert knowledge, information from other data sources, and subjective beliefs. Upon observing new data, the prior distribution is

updated to a posterior distribution, forming the basis for predictive inference. This sequential inference facilitates online (real-time) predictions and is particularly useful for applications involving massive streaming data.

1.2 Objectives and main contributions

The main objective of this work is to develop frameworks for efficient model estimation and predictions for univariate and multivariate temporal data. Several novel enhancements of leading mathematical models are proposed, which include: (i) a closed-form covariance estimation algorithm that processes data online in real-time; (ii) a change detection algorithm for structural breaks; (iii) a comprehensive and extensible deterministic-probabilistic framework for modelling infectious disease data; (iv) a thorough study of mean-reverting transmission processes in a mechanistic stochastic model; and (v) a methodology for analyzing and forecasting peak time and height using innovative data and feature engineering.

1.3 Outline of the thesis

The thesis is organized into six chapters:

- **Chapter 1:** Includes background material and terminology related to concepts in time series analysis and Bayesian learning.
- **Chapter 2:** Introduces the state-space modeling framework and estimation methods, first by emphasizing linearity and normality in models. It includes several case studies illustrating the utility of dynamic regression and concludes with an online real-time multivariate covariance estimation method.
- **Chapter 3:** Advances to stochastic simulation methods for general state-space models, introducing Markov chain Monte Carlo methods and their application in non-linear regression case studies. It also contains online real-time structural break change detection using sequential Monte Carlo method.
- **Chapter 4:** Formulates and develops, both mathematically and data analytically, a compartmental model linked to a mobility enhanced probabilistic data-driven model for epidemic disease transmission and then applies the methodology to real data with comparison to established results.

- **Chapter 5:** Extends a classical SIRS model by incorporating stochastic, mean-reverting processes to reflect real-world epidemic variability that deterministic models often overlook. We integrate processes such as the CKLS, modified Bessel process and logistic GBM, deriving extinction thresholds to predict disease persistence. Simulations confirm that these models more accurately capture epidemic dynamics, offering valuable insights for public health forecasting and interventions.
- **Chapter 6:** Introduces a framework for analyzing peak characteristics of epidemic data series using various traditional machine learning and deep learning methods, employing innovative feature engineering and data sourcing by leveraging geographical delays in epidemic arrival to improve the forecasting capability of future epidemic waves.

1.4 Exploratory methods

A time series is sequentially sampled data [1], [2]. Analysis of time series requires consideration of the ordering and relationship within the data. The main goals are to devise methods to accurately estimate or learn the past, present, and future values of particularly interesting events and then to derive insight into the structure and influence of the events.

Estimation can be broadly grouped into three categories: smoothing, filtering, and prediction or forecasting for, respectively, past, present, and future time points [3]. Smoothing can further be categorized as fixed-interval smoothing by considering all values up to the present time point, fixed-lag smoothing by considering lagged steps ahead values to the estimate point, and fixed-point smoothing by considering a specific time point.

Time series analysis explicitly considers the relation between data at each time point. This is different from some other types of statistical analysis where the data are assumed to be independent and identically distributed. In general, time series can possess both cycles over a certain unit of time and also a trend. Each of these phenomena possess data correlation in time which needs to be considered for accurate estimation.

A stochastic or random process [4] is a series of random variables ordered in time. The correlation function of two stochastic processes X_t and Y_t is $R(t, k) = E[X_t Y_{t-k}]$, where the time offset k is the lag. As the correlation function is a type of covariance, the correlation function vanishes when the stochastic processes X_t and Y_t are independent.

The autocovariance of a stochastic process X_t is the covariance between the data and the data shifted in time and is a function of time and lag:

$$\text{Cov}[X_t, X_{t-k}] = E[(X_t - E[X_t])(X_{t-k} - E[X_{t-k}])].$$

Normalized autocovariance is the autocorrelation coefficient. It is the correlation coefficient between the data and the data shifted in time and is a function of time and lag:

$$\rho_{t,k} = \frac{\text{Cov}[X_t, X_{t-k}]}{\sqrt{\text{Var}[X_t]}\sqrt{\text{Var}[X_{t-k}]}.$$

The autocorrelation coefficient ranges from -1 to $+1$.

Stationarity of a stochastic process refers to a stochastic characteristic that does not change along with time. When the probability distribution for a stochastic process is independent of time, its stochastic process is strictly stationary. When the mean, variance, autocovariance, and autocorrelation coefficient of a stochastic process are invariant independently of time, the process is weakly stationary. In most applications, time series analysis is most amenable to weak stationarity, which is sometimes simply referred to as stationarity. A stationary stochastic process with zero mean and zero autocovariance and autocorrelation coefficient, except for lag 0, is an instance of white noise.

The autoregressive (AR) model is a linear relation between current and past data. The AR(1) model is defined by

$$Y_t = \phi Y_{t-1} + \epsilon_t,$$

where ϵ_t is white noise. The value of ϕ determines the stationarity of the AR(1) model. When $|\phi| \leq 1$, then there is stationarity, and otherwise, there is non-stationarity. Given a time series, the Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests can be used to assess stationarity. More sophisticated state-space models can deal with nonstationary time series directly, and determination of stationarity is not needed for analysis.

Some pieces of practical advice are in order. Preliminary examination of the data includes basic time plots, generation of histograms and data summarization, computation of autocorrelation coefficients, and frequency spectrum. The autocorrelation coefficient may be large for particular lags, implying periodicity in the data, which can further be investigated with a frequency spectrum. Once the data is examined, a model is defined to help express characteristics of the data, usually in terms of multiple combined simple components or structures. This approach is termed time series decomposition [5], [6] and typically begins with considering structures related to level, trend, and seasonality, also known as the basic structural model or standard seasonal adjustment model. The remainder of unexplained variation from the estimated components are the residuals.

1.5 Bayesian learning

This section lays the groundwork for Bayesian learning by introducing essential concepts such as prior distribution, likelihood function, and posterior distribution. Initially, covariates are excluded for simplicity, with models incorporating covariates discussed in later chapters. The data consists of n observations $y_1, \dots, y_n$, modeled as realizations from an unknown data-generating process. Extensive discussion of these concepts and more can be found in (Gelman, 2013) [7].

The observed data set is used to model the process by constructing a conditional probabilistic model $f(y_1, \dots, y_n \mid \theta)$, which describes the uncertainty about observations given an unknown parameter θ . The model assumes the observations are independent and identically distributed conditional on θ :

$$f(y_1, \dots, y_n \mid \theta) = \prod_{i=1}^n f(y_i \mid \theta),$$

where $f(y \mid \theta)$ is any parametric density function parameterized by θ . The likelihood function, which is central to statistical inference, is given by:

$$L(\theta; y_1, \dots, y_n) = \prod_{i=1}^n f(y_i \mid \theta).$$

The maximum likelihood estimator (MLE) is obtained by maximizing the likelihood function:

$$\hat{\theta}_{\text{MLE}} = \arg \max_{\theta} L(\theta; y_1, \dots, y_n).$$

In the frequentist approach, the MLE is used for inference, focusing on the repeated sampling variability of the estimator. For prediction, the frequentist approach conditions on the MLE to obtain the predictive density:

$$f_{\text{MLE}}(y_{n+1}, \dots, y_{n+m} \mid y_1, \dots, y_n) = f(y_{n+1}, \dots, y_{n+m} \mid \hat{\theta}_{\text{MLE}}).$$

In contrast, the Bayesian paradigm treats the parameter θ as a random variable. Prior beliefs about θ are updated with observed data using Bayes' theorem:

$$p(\theta \mid y_1, \dots, y_n) = \frac{f(y_1, \dots, y_n \mid \theta)p(\theta)}{\int_{\Omega} f(y_1, \dots, y_n \mid \theta)p(\theta)d\theta},$$

where Ω is the parameter space for θ . The posterior distribution $p(\theta \mid y_1, \dots, y_n)$ is a

compromise between the prior distribution $p(\theta)$ and the likelihood function $f(y_1, \dots, y_n \mid \theta)$.

As the sample size increases, the posterior distribution converges to a normal distribution with mean $\hat{\theta}_{\text{MLE}}$ and variance $I_{\hat{\theta}_{\text{MLE}}}^{-1}$, where $I_{\hat{\theta}_{\text{MLE}}}$ is the Fisher information:

$$p(\theta \mid y_1, \dots, y_n) \xrightarrow{d} \mathcal{N}(\hat{\theta}_{\text{MLE}}, I_{\hat{\theta}_{\text{MLE}}}^{-1}),$$

under regularity conditions. This result allows for practical Bayesian inference even with finite sample sizes, often using simulation algorithms like Markov Chain Monte Carlo (MCMC) and Sequential Monte Carlo (SMC).

In models with multiple parameters, the marginal posterior density is obtained by integrating out the other parameters:

$$p(\theta_k \mid y_1, \dots, y_n) = \int_{\Omega_{-k}} p(\theta_k, \theta_{-k} \mid y_1, \dots, y_n) d\theta_{-k},$$

where θ_{-k} denotes all parameters excluding θ_k and Ω_{-k} is the parameter space for θ_{-k} . Point estimates, such as the maximum a posteriori (MAP) estimator, and credible intervals are used to summarize the posterior distribution. The MAP estimator is defined as:

$$\hat{\theta}_{\text{MAP}} = \arg \max_{\theta} p(\theta \mid y_1, \dots, y_n).$$

A $100 \times (1 - \alpha)\%$ posterior credible interval which communicates the uncertainty about the parameter may be computed with respect to the posterior or the predictive distribution as:

$$\int_{\theta_l}^{\theta_u} p(\theta \mid y_1, \dots, y_n) d\theta = 1 - \alpha.$$

Prior elicitation involves selecting a suitable prior density function, often from a family of distributions parameterized by hyperparameters. Conjugate priors, which belong to the same probability distribution family as the likelihood, are computationally convenient as they allow the posterior to be computed in closed form. Such relationships are heavily relied upon in the covariance estimation technique in a subsequent chapter. Informative priors can be constructed from expert knowledge or historical data, while non-informative priors are used when prior information is limited.

Chapter 2

State-Space Modelling for Infectious Disease Surveillance Data: Dynamic Regression and Covariance Analysis

2.1 Introduction

State-space modelling is a powerful statistical method used for estimating the state of a dynamic system as it evolves over time. This technique leverages the principles of recursive Bayesian state estimation to update the probability distribution of the system's state based on new observations, making it particularly effective for applications involving noisy or incomplete data.

The primary objective of this chapter is to review, explore, apply and enhance the recursive Bayesian state estimation framework, with a focus on how to improve its accuracy and computational efficiency. We aim at providing a comprehensive overview of the fundamental concepts, methodologies, and applications of recursive Bayesian state estimation. Additionally, the frequentist approach to fitting dynamic models to epidemic outbreaks [8], has been proposed to address parameter uncertainty and identifiability in the context of infectious disease modelling.

More specifically, we introduce the theoretical underpinnings of Bayesian inference and its relevance to state estimation problems, and introduce the concept of the state-space model, which serves as the foundation for recursive Bayesian state estimation. We develop and provide here also various algorithms used for implementing Bayesian state estimation, including the Kalman filter and other advanced techniques.

We then present several case studies and simulations are presented to illustrate the prac-

tical applications and effectiveness of these methodologies in real-world scenarios. As shall be seen, the findings from these case studies not only highlight the strengths of recursive Bayesian state estimation but also reveal areas where further improvements can be made. For example, recent advancements in dynamic modelling tools [9] can generate real-time short-term forecasts with quantified uncertainty of diverse growth processes, including disease outbreaks for fitting and forecasting time-series trajectories.

Overall, our objective is to provide a detailed exploration of state-space modelling and recursive Bayesian state estimation, contributing to the broader understanding and advancement of this critical area in statistical modelling and dynamic system analysis for infectious diseases [10].

2.2 Methodology

2.2.1 State-space model

A state-space model [5], [6], [11] is a type of flexible stochastic model that considers observations as samples from a probability distribution. The estimation problem, is then of a full distribution and not a point value. More formally, a state-space model attributes time series data within a stochastic relationship. A crucial introduction is that of latent or hidden random variables, called a state. A state cannot be directly observed, unlike the observations or time series data and can be chosen such that interpretation of the analysis is possible.

Development of state-space models has been carried out by R. E. Bellman and R. E. Kalman in the control theory community during 1950s. Hidden Markov model [12], [13] widely applied in speech recognition is a state-space model where the random variable under consideration is discrete in nature.

The graphical model, more specifically, directed acyclic graph (DAG) of a state-space model is depicted as follows.

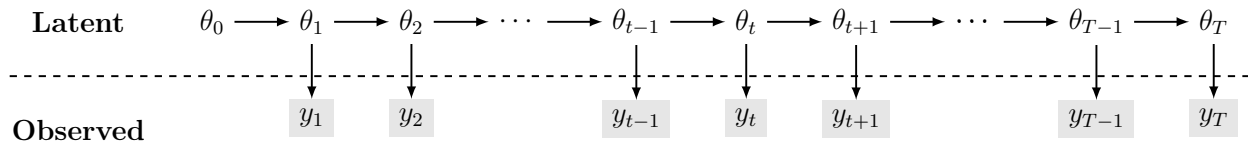


Figure 2.1: DAG of state-space model.

A major model assumption is the Markov property, which says that a state is related

only to the state at the previous time point:

$$p(\theta_t \mid \theta_{0:t-1}, y_{1:t-1}) = p(\theta_t \mid \theta_{t-1}). \quad (2.2.1)$$

Conditional independence on states are depicted as follows.

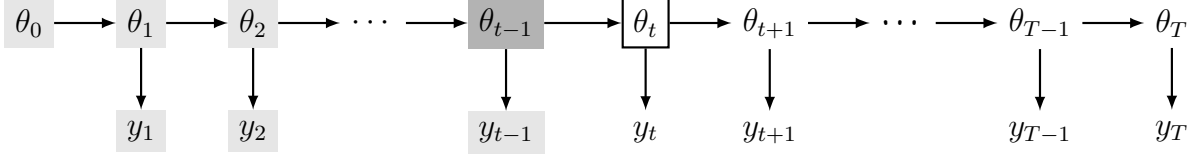


Figure 2.2: Conditional independence of model states.

A further assumption pertains to an observation at a certain time point depending only on the state at the same time point by conditional independence:

$$p(y_t \mid \theta_{0:t}, y_{1:t-1}) = p(y_t \mid \theta_t). \quad (2.2.2)$$

Conditional independence on observations is depicted in the following diagram:

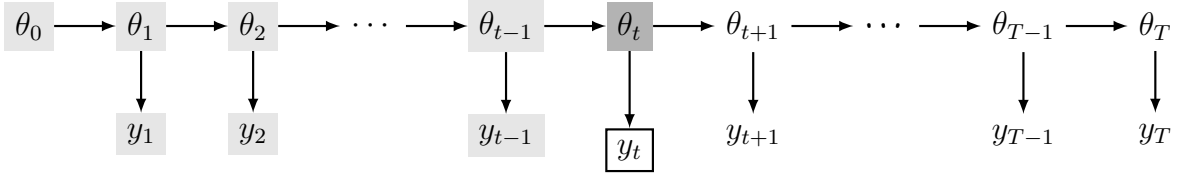


Figure 2.3: Conditional independence of observations.

Similarly, conditional independence also holds such that:

$$p(\theta_t \mid \theta_{t+1}, y_{1:T}) = p(\theta_t \mid \theta_{t+1}, y_{1:t}). \quad (2.2.3)$$

This can be seen as follows:

$$\begin{aligned}
p(\theta_t \mid \theta_{t+1}, y_{1:T}) &= p(\theta_t \mid \theta_{t+1}, y_{1:t}, y_{t+1:T}) \\
&= p(\theta_t \mid y_{t+1:T}, (\theta_{t+1}, y_{1:t})) \\
&= \frac{p(\theta_t, y_{t+1:T} \mid \theta_{t+1}, y_{1:t})}{p(y_{t+1:T} \mid \theta_{t+1}, y_{1:t})} \\
&= \frac{p(y_{t+1:T}, \theta_t \mid \theta_{t+1}, y_{1:t})}{p(y_{t+1:T} \mid \theta_{t+1}, y_{1:t})} \\
&= \frac{p(y_{t+1:T} \mid \theta_t, \theta_{t+1}, y_{1:t})p(\theta_t \mid \theta_{t+1}, y_{1:t})}{p(y_{t+1:T} \mid \theta_{t+1}, y_{1:t})} \\
&= \frac{p(y_{t+1:T} \mid \theta_{t+1}, y_{1:t})p(\theta_t \mid \theta_{t+1}, y_{1:t})}{p(y_{t+1:T} \mid \theta_{t+1}, y_{1:t})} \\
&= p(\theta_t \mid \theta_{t+1}, y_{1:t}).
\end{aligned}$$

The joint distribution of the state space model is $p(\text{all random variables}) = p(\theta_{0:T}, y_{1:T})$,
or,

$$\begin{aligned}
p(\theta_{0:T}, y_{1:T}) &= p(y_{1:T}, \theta_{0:T}) \\
&= p(y_T, y_{1:T-1}, \theta_{0:T}) \\
&= p(y_T \mid y_{1:T-1}, \theta_{0:T})p(y_{1:T-1}, \theta_{0:T}) \\
&= p(y_T \mid y_{1:T-1}, \theta_{0:T})p(\theta_T, \theta_{0:T-1}, y_{1:T-1}) \\
&= p(y_T \mid y_{1:T-1}, \theta_{0:T})p(\theta_T \mid \theta_{0:T-1}, y_{1:T-1})p(\theta_{0:T-1}, y_{1:T-1}) \\
&= \left(\prod_{t=2}^T p(y_t \mid y_{1:t-1}, \theta_{0:t})p(\theta_t \mid \theta_{0:t-1}, y_{1:t-1}) \right) p(\theta_{0:1}, y_1) \\
&= \left(\prod_{t=2}^T p(y_t \mid y_{1:t-1}, \theta_{0:t})p(\theta_t \mid \theta_{0:t-1}, y_{1:t-1}) \right) p(y_1 \mid \theta_{0:1})p(\theta_1 \mid \theta_0)p(\theta_0) \\
&= \left(\prod_{t=2}^T p(y_t \mid \theta_0)p(\theta_t \mid \theta_{t-1}) \right) p(y_1 \mid \theta_{0:1})p(\theta_1 \mid \theta_0)p(\theta_0) \\
&= p(\theta_0) \prod_{t=1}^T p(y_t \mid \theta_t)p(\theta_t \mid \theta_{t-1}). \tag{2.2.4}
\end{aligned}$$

Thus, the state-space model is a product of prior distribution $p(\theta_0)$, observation equation $p(y_t \mid \theta_t)$ and state equation $p(\theta_t \mid \theta_{t-1})$.

The state-space model can also be represented by functional equations:

$$\theta_t = f(\theta_{t-1}, w_t), \quad (2.2.5)$$

$$y_t = h(\theta_t, v_t), \quad (2.2.6)$$

where f and h are arbitrary functions, w_t and v_t are mutually independent white noise state noise (system, or process noise) and observation noise (or, measurement noise), respectively. w_t is p -dimensional column vector. Together, these two equations are termed the state (system) equation and observation (measurement) equations. They are equivalent statements of the probability distribution definition of the model. The state equation (2.2.5) is a stochastic difference equation for the state and the state noise is the tolerance for the time-dependent change of state. In the observation equation (2.2.6), the observation noise is interpreted as typical noise separate from the state noise with more accurate observations resulting from less observation noise.

Classification of state-space models is based on the linearity or non-linearity of the functions f or h and of normality assumption of the noise processes w_t or v_t .

A special instance of a state-space model in which both f and h are linear functions and both w_t and v_t have Gaussian distribution is referred to as linear Gaussian state-space model. The state and observation equations are then given by:

$$\theta_t = G_t \theta_{t-1} + w_t \quad w_t \sim \mathcal{N}(0, W_t), \quad (2.2.7)$$

$$y_t = F_t \theta_t + v_t \quad v_t \sim \mathcal{N}(0, V_t), \quad (2.2.8)$$

where G_t is a $p \times p$ state transition matrix (or, state evolution matrix/system matrix), F_t is a $1 \times p$ observation matrix (or, measurements matrix), W_t is a $p \times p$ covariance matrix for the state noise and V_t is a variance for the observation noise. The prior distribution is $\theta_0 \sim \mathcal{N}(m_0, C_0)$ where m_0 is the p -dimensional mean vector and C_0 is a $p \times p$ covariance matrix. Note that linear Gaussian models possess the attractive analytical property of reproducibility whereby linear operations on normal variates yield normal variates.

In terms of probability distributions, the linear Gaussian state-space model is given by:

$$p(\theta_t \mid \theta_{t-1}) = \mathcal{N}(G_t \theta_{t-1}, W_t), \quad (2.2.9)$$

$$p(y_t \mid \theta_t) = \mathcal{N}(F_t \theta_t, V_t), \quad (2.2.10)$$

where $p(\theta_0) = \mathcal{N}(m_0, C_0)$. The parameters of the linear Gaussian state-space model are denoted by $\psi = \{G_t, F_t, W_t, V_t, m_0, C_0\}$. Linear Gaussian state-space model is also called

a dynamic linear model (DLM) which has been extensively studied [14], [15] due to its widespread applicability on its own and by conditional general state-space models.

With the introduction of a state component to the model, more sophisticated and flexible models can be constructed by combining multiple states such that the interpretation of the model is clear. Adding an extra layer of the state equation allows modelling to occur between observations and states as opposed to between observations only. In other words, the observations are not connected to one another, but rather, through the states. This can be seen from the graphical model for the state-space model. This is in opposition to the classic autoregressive moving average model (ARMA) which directly models the relation between observations [14]. The ARMA(p, q) model is defined as $y_t = \sum_{j=1}^p \phi_j y_{t-j} + \sum_{j=1}^q \psi_j \epsilon_{t-j} + \epsilon_t$. Here p and q are non-negative integers denoting the order of the AR and MA terms, respectively, ϕ_j and ψ_j are real numbers denoting the AR coefficient and MA coefficient, respectively, and ϵ_t is Gaussian white noise with variance σ_ϵ^2 . The d -th differenced time series yields the extended autoregressive integrated moving average model, or ARIMA(p, d, q). ARIMA and state-space models can be shown to be equivalent (Harvey 1989). Thus the two competing stochastic modelling frameworks can both capture temporal relations in data equally well. The state-space approach possesses advantages of increased flexibility in handling non-stationary processes, missing information and external covariates [11]. The ARMA model can be written as a state-space model as follows. Consider the ARMA($r, r-1$) model where $r = \max\{p, q+1\}$, $\phi_j = 0$ for $j > p$ and $\psi_j = 0$ for $j > q$ defined by

$$y_t = \sum_{j=1}^r \phi_j y_{t-j} + \sum_{j=1}^{r-1} \psi_j \epsilon_{t-j} + \epsilon_t. \quad (2.2.11)$$

Let $F_t = [1, 0, \dots, 0, 0]$, $R_t = [1, \psi_1, \dots, \psi_{r-2}, \psi_{r-1}]^\top$ and

$$G_t = \begin{bmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \\ \phi_{r-1} & 0 & \cdots & 0 & 1 \\ \phi_r & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

For an r -dimensional state vector $\theta_t = [\theta_t^1, \theta_t^2, \dots, \theta_t^{r-1}, \theta_t^r]^\top$ and state noise covariance $W_t = RR^\top \sigma_\epsilon^2$ such that $w_t = R\epsilon_t$, and observation noise covariance $V_t = 0$. Such formulation

is equivalent to the ARMA(p, q) model in (2.2.11). To see this, consider the state equation

$$\begin{aligned}\theta_t^1 &= \phi_1 \theta_{t-1}^1 + \theta_{t-1}^2 + \epsilon_t, \\ \theta_t^2 &= \phi_2 \theta_{t-1}^1 + \theta_{t-1}^3 + \psi_1 \epsilon_t, \\ &\vdots \\ \theta_t^{r-1} &= \phi_{r-1} \theta_{t-1}^1 + \theta_{t-1}^r + \psi_{r-2} \epsilon_t, \\ \theta_t^r &= \phi_r \theta_{t-1}^1 + \psi_{r-1} \epsilon_t.\end{aligned}$$

Substituting the equation for x_t^2 into the first equation gives

$$\theta_t^1 = \phi_1 \theta_{t-1}^1 + \phi_2 \theta_{t-2}^1 + \theta_{t-2}^3 + \psi_1 \epsilon_{t-1} + \epsilon_t.$$

Continuing with substitution in a similar fashion gives

$$\theta_t^1 = \phi_1 \theta_{t-1}^1 + \cdots + \phi_r \theta_{t-r}^1 + \psi_1 \epsilon_{t-1} + \cdots + \psi_{r-1} \epsilon_{t-(r-1)} + \epsilon_t.$$

Setting $y_t = x_t^1$ in the observation equation produces the result as shown in (2.2.11) with $r = \max\{p, q + 1\}$:

$$y_t = \phi_1 y_{t-1}^1 + \cdots + \phi_r y_{t-r}^1 + \psi_1 \epsilon_{t-1} + \cdots + \psi_{r-1} \epsilon_{t-(r-1)} + \epsilon_t.$$

The equivalence of the ARIMA(p, d, q) model follows by a similar argument [14].

2.2.2 State estimation

The posterior distribution of the state θ_s is a conditional distribution and is denoted by

$$p(\theta_s \mid y_{1:t}). \tag{2.2.12}$$

The joint posterior distribution of the state considers all time points and is denoted by:

$$p(\theta_0, \theta_1, \dots, \theta_{t'}, \dots \mid y_{1:t}). \tag{2.2.13}$$

It is the estimation target in the solution of the state space model using MCMC and the particle filter. Some works also refer to the joint posterior distribution simply as the smoothing distribution.

In the joint posterior distribution (2.2.13), the time point of interest t' determines the

type of state estimation:

smoothing, $t' < t$,
filtering, $t' = t$,
prediction, $t' > t$.

For a particular state of interest $\theta_{t'}$, its marginal posterior distribution can be found by:

$$p(\theta_{t'} \mid y_{1:t}) = \int p(\theta_0, \theta_1, \dots, \theta_{t'}, \dots \mid y_{1:t}) d\theta_{\text{except } t'}. \quad (2.2.14)$$

The marginal posterior distribution is the central target distribution in Bayesian inference of state space models. It can be derived using MCMC by marginalization of the joint posterior distribution. Additionally, direct methods such as Weiner filter (solution in the frequency domain) and Kalman filter (solution in the time domain) can be used to obtain the marginal posterior distribution [3], [16] as well as approximate methods that utilize particle filters [17].

We now provide formulation of filtering, predictive and smoothing distributions. Recursion relations for the marginal posterior distribution of that state are derived first for the filtering distribution, which is central to prediction and smoothing. The filtering distribution can be obtained from the marginal posterior distribution of the state by setting $t' = t$ in (2.2.14):

$$\begin{aligned}
p(\theta_t \mid y_{1:t}) &= p(\theta_t \mid y_t, y_{1:t-1}) \\
&= \frac{p(\theta_t, y_t \mid y_{1:t-1})}{p(y_t \mid y_{1:t-1})} \\
&= \frac{p(y_t, \theta_t \mid y_{1:t-1})}{p(y_t \mid y_{1:t-1})} \\
&= \frac{p(y_t \mid \theta_t, y_{1:t-1})p(\theta_t \mid y_{1:t-1})}{p(y_t \mid y_{1:t-1})} \\
&= \frac{p(y_t \mid \theta_t)p(\theta_t \mid y_{1:t-1})}{p(y_t \mid y_{1:t-1})} \tag{2.2.15}
\end{aligned}$$

$$\begin{aligned}
&= \frac{p(y_t \mid \theta_t) \int p(\theta_t, \theta_{t-1} \mid y_{1:t-1})d\theta_{t-1}}{\int p(y_t, \theta_t \mid y_{1:t-1})d\theta_t} \\
&= \frac{p(y_t \mid \theta_t) \int p(\theta_t \mid \theta_{t-1}, y_{1:t-1})p(\theta_{t-1} \mid y_{1:t-1})d\theta_{t-1}}{\int p(y_t \mid \theta_t, y_{1:t-1})p(\theta_t \mid y_{1:t-1})d\theta_t} \\
&= \frac{p(y_t \mid \theta_t) \int p(\theta_t \mid \theta_{t-1})p(\theta_{t-1} \mid y_{1:t-1})d\theta_{t-1}}{\int p(y_t \mid \theta_t)p(\theta_t \mid y_{1:t-1})d\theta_t}, \tag{2.2.16}
\end{aligned}$$

where (2.2.15) is the fundamental form of filtering distribution, which follows by conditional independence. In the numerator $p(\theta_t \mid y_{1:t-1}) = \int p(\theta_t \mid \theta_{t-1})p(\theta_{t-1} \mid y_{1:t-1})d\theta_{t-1}$ is the one-step-ahead predictive distribution multiplied by the likelihood $p(y_t \mid \theta_t)$ and the denominator $p(y_t \mid y_{1:t-1}) = \int p(y_t \mid \theta_t)p(\theta_t \mid y_{1:t-1})d\theta_t$ is the one-step-ahead predictive likelihood. Here, we can think of the one-step-ahead predictive distribution as playing the role of a prior distribution and it represents the filtering distribution one time step before it moves forward based on the state equation. The one-step-ahead predictive likelihood is a normalization factor for the filtering distribution and it transforms the one-step-ahead predictive distribution from the observation equations.

Conceptually, filtering is achieved by first obtaining the one-step-ahead predictive distribution through state transition of the filtering distribution one time step prior and then using the observations to correct the obtained distribution [3]. As an illustration of this process, recursion diagrams are depicted later.

The k -steps-ahead predictive distribution can be obtained from the marginal posterior

distribution of the state by setting $t' = t + k$ in (2.2.14):

$$\begin{aligned}
p(\theta_{t+k} \mid y_{1:t}) &= \int p(\theta_{t+k}, \theta_{t+k-1} \mid y_{1:t}) d\theta_{t+k-1} \\
&= \int p(\theta_{t+k} \mid \theta_{t+k-1}, y_{1:t}) p(\theta_{t+k-1} \mid y_{1:t}) d\theta_{t+k-1} \\
&= \int p(\theta_{t+k} \mid \theta_{t+k-1}) p(\theta_{t+k-1} \mid y_{1:t}) d\theta_{t+k-1}.
\end{aligned} \tag{2.2.17}$$

The k -steps-ahead predictive distribution (2.2.17) is the $k - 1$ -steps-ahead predictive distribution propagated forward in time by the state equation. In other words, applying the state equation to the one-step-ahead predictive distribution $k - 1$ times gives the k -steps-ahead predictive distribution. Also, since the filtering distribution for future time points are calculated by applying the state equation, the one-step-ahead predictive distribution can be obtained. Therefore, by applying the state equation in the forward time direction k times to the filtering distribution $p(\theta_t \mid y_{1:t})$, the k -steps-ahead distribution can be obtained.

Conceptually, k -steps-ahead prediction is obtained by repeated k application of the state equation to the filtering distribution. Prediction can also be contrasted with filtering. In filtering, the state transition is corrected based on the observations. This is not true of prediction, where rather the accuracy of prediction decreases as k increases, but the most likely distributions appear through the filtering point. This also means that missing observations in the model are treated as prediction values. Hence, the state-space model framework readily handles missing observations.

The fixed-interval smoothing distribution can be obtained from the marginal posterior

distribution of the state by setting $T = t$ in (2.2.14):

$$\begin{aligned}
p(\theta_t \mid y_{1:T}) &= \int p(\theta_t, \theta_{t+1} \mid y_{1:T}) d\theta_{t+1} \\
&= \int p(\theta_t \mid \theta_{t+1}, y_{1:T}) p(\theta_{t+1} \mid y_{1:T}) d\theta_{t+1} \\
&= \int p(\theta_{t+1} \mid y_{1:T}) p(\theta_t \mid \theta_{t+1}, y_{1:T}) d\theta_{t+1} \\
&= \int p(\theta_{t+1} \mid y_{1:T}) p(\theta_t \mid \theta_{t+1}, y_{1:t}) d\theta_{t+1} \\
&= \int p(\theta_{t+1} \mid y_{1:T}) \frac{p(\theta_t, \theta_{t+1} \mid y_{1:t})}{p(\theta_{t+1} \mid y_{1:t})} d\theta_{t+1} \\
&= \int p(\theta_{t+1} \mid y_{1:T}) \frac{p(\theta_{t+1}, \theta_t \mid y_{1:t})}{p(\theta_{t+1} \mid y_{1:t})} d\theta_{t+1} \\
&= \int p(\theta_{t+1} \mid y_{1:T}) \frac{p(\theta_{t+1} \mid \theta_t, y_{1:t}) p(\theta_t \mid y_{1:t})}{p(\theta_{t+1} \mid y_{1:t})} d\theta_{t+1} \\
&= p(\theta_t \mid y_{1:t}) \int p(\theta_{t+1} \mid y_{1:T}) \frac{p(\theta_{t+1} \mid \theta_t, y_{1:t})}{p(\theta_{t+1} \mid y_{1:t})} d\theta_{t+1} \\
&= p(\theta_t \mid y_{1:t}) \int p(\theta_{t+1} \mid y_{1:T}) \frac{p(\theta_{t+1} \mid \theta_t)}{p(\theta_{t+1} \mid y_{1:t})} d\theta_{t+1}. \tag{2.2.18}
\end{aligned}$$

In the formulation of the smoothing distribution (2.2.18), the filtering distribution is updated based on the smoothing distribution one time step ahead. Conceptually, smoothing at each time point is accomplished by repeatedly applying the smoothing recursion in the backward time direction beginning with the filtering distribution at time point T .

Recursions for the marginal posterior distribution are summarized below:

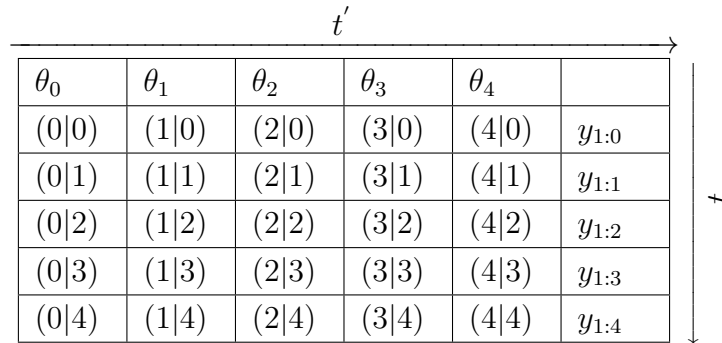


Figure 2.4: Recursion diagram for posterior distribution $p(\theta_{t'} \mid y_{1:t}) = (t' \mid t)$. The calculation flow for the recursion is reached by moving forward in time denoted by right and down arrows.

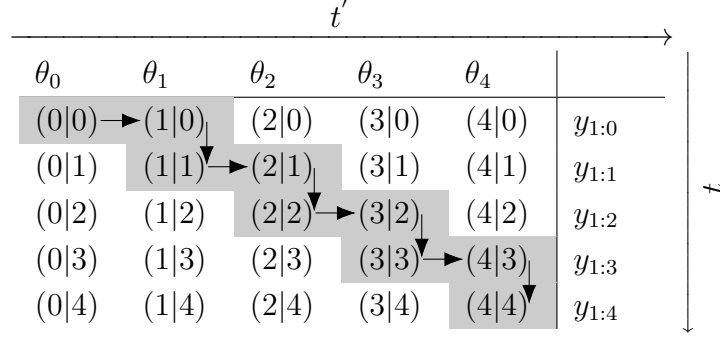


Figure 2.5: Recursion processing for filtering distribution $p(\theta_t | y_{1:t})$. Filtering is completed by first obtaining the one-step-ahead predictive distribution through the state transition from the filtering distribution one time before and then secondly correcting the obtained distribution based on the observations. The conceptual diagram shows the calculation follows a stepwise pattern.

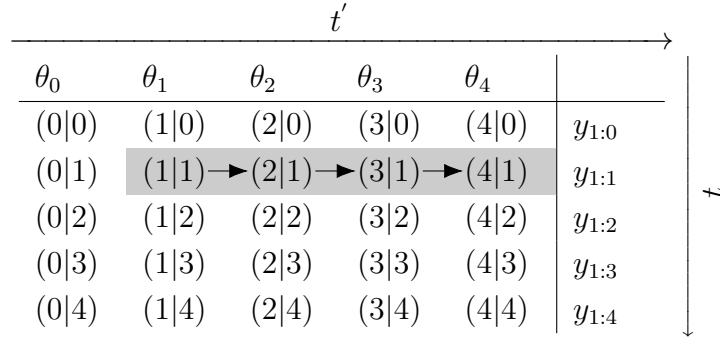


Figure 2.6: Recursion processing for predictive distribution $p(\theta_{t+k} | y_{1:t})$. The k -steps-ahead prediction can be achieved by repeating the state transition k times for the filtering distribution. The processing to obtain three-steps-ahead prediction from the current time point $t = 1$ is represented by repeating the right arrow thrice.

For k -steps-ahead prediction, the state transition (right arrow) is repeated without correction based on the observations (down arrow) in contrast to filtering. The accuracy of prediction gradually decreases as k increases, but the most likely distributions continue to be generated based on the information through the filtering point. The state-space model applies this concept directly to missing observations as they are compensated for with prediction values.

t'					
θ_0	θ_1	θ_2	θ_3	θ_4	
(0 0)	(1 0)	(2 0)	(3 0)	(4 0)	$y_{1:0}$
(0 1)	(1 1)	(2 1)	(3 1)	(4 1)	$y_{1:1}$
(0 2)	(1 2)	(2 2)	(3 2)	(4 2)	$y_{1:2}$
(0 3) ←	(1 3) ←	(2 3) ←	(3 3) ←	(4 3)	$y_{1:3}$
(0 4)	(1 4)	(2 4)	(3 4)	(4 4)	$y_{1:4}$

Figure 2.7: Recursion processing for smoothing distribution $p(\theta_t \mid y_{1:T})$. At $T = 3$, the processing to obtain the smoothing distribution at $t = 0$ is represented by repeating the left arrow operation four times.

In the state-space model, the likelihood for a time series $\{y_t : t = 1, 2, \dots, T\}$ with parameters ψ is the probability distribution

$$\begin{aligned}
p(y_1, \dots, y_T; \psi) &= p(y_T, y_{1:T-1}; \psi) \\
&= p(y_T \mid y_{1:T-1}; \psi) p(y_{1:T-1}; \psi) \\
&= \prod_{t=1}^T p(y_t \mid y_{1:t-1}; \psi).
\end{aligned} \tag{2.2.19}$$

Taking the log of the likelihood yields

$$\begin{aligned}
\ell(\psi) &= \log p(y_1, \dots, y_T; \psi) \\
&= \sum_{t=1}^T \log p(y_t \mid y_{1:t-1}; \psi).
\end{aligned} \tag{2.2.20}$$

Therefore, the log-likelihood for the time series is equivalent to the sum of the logarithms of one-step-ahead predictive likelihoods, which can be efficiently calculated by filtering recursion. It is a numerical measure of how well the time ordered observations fit the model. It also indicates the predictive ability of future values based on past values. It is therefore useful as the basis for a model selection criteria of such metrics as the information theoretic extensions Akaike information criterion (AIC) and widely applicable information criterion (WAIC).

Furthermore, it is useful to note that the likelihood does not depend on the state θ as it has been marginalized out and so it is apt to call it the marginal likelihood. In terms of implementation, the importance of extreme values depending on separate states has been

lessened as stable values are more attainable.

Estimation of the unknown parameters of a state-space model can be accomplished in various ways. In the frequentist approach, parameters are treated as non-random quantities. The main objective is to determine the maximum likelihood estimate (MLE)

$$\begin{aligned}\hat{\psi} &= \operatorname{argmax}_{\psi} \ell(\psi) \\ &= \operatorname{argmax}_{\psi} \log p(y_{1:T}; \psi) \\ &= \operatorname{argmax}_{\psi} \sum_{t=1}^T \log p(y_t \mid y_{1:t-1}; \psi).\end{aligned}\tag{2.2.21}$$

which can be found by numerical methods such as Newton-type methods, or expectation maximization (EM) algorithm [18].

In the Bayesian paradigm, parameters are treated as random variables and this makes uncertainty quantification possible for both the state and parameters of the model. From the previous result obtained as (2.2.4): $p(\theta_{0:T}, y_{1:T}) = p(\theta_0) \prod_{t=1}^T p(y_t \mid \theta_t) p(\theta_t \mid \theta_{t-1})$, it can be observed that adding the parameters to the probability statements would yield state-parameter augmented state $\{\theta, \psi\}$. So, in a similar line of reasoning, the joint distribution of this extended model is

$$\begin{aligned}p(\theta_{0:T}, y_{1:T}, \psi) &= p(\theta_0, \psi) \prod_{t=1}^T p(y_t \mid \theta_t, \psi) p(\theta_t \mid \theta_{t-1}, \psi) \\ &= p(\theta_0 \mid \psi) p(\psi) \prod_{t=1}^T p(y_t \mid \theta_t, \psi) p(\theta_t \mid \theta_{t-1}, \psi).\end{aligned}\tag{2.2.22}$$

The joint posterior distribution of the augmented state follows from (2.2.13):

$$p(\theta_0, \theta_1, \dots, \theta_{t'}, \dots, \psi \mid y_{1:t}) = p(\theta_0, \theta_1, \dots, \theta_{t'}, \dots \mid y_{1:t}, \psi) p(\psi \mid y_{1:t}),\tag{2.2.23}$$

and the equivalent statement of (2.2.14) is given by

$$p(\theta_{t'}, \psi \mid y_{1:t}) = p(\theta_{t'} \mid \psi, y_{1:t}) p(\psi \mid y_{1:t}).\tag{2.2.24}$$

The marginal posterior distribution $p(\psi \mid y_{1:t})$ can then be obtained by integrating out the state. Maximizing the marginal posterior distribution gives the maximum a posteriori

(MAP) solution:

$$\begin{aligned}\operatorname{argmax}_{\psi} p(\psi \mid y_{1:t}) &= \operatorname{argmax}_{\psi} \frac{p(y_{1:t} \mid \psi) p(\psi)}{p(y_{1:t})} \\ &\propto \operatorname{argmax}_{\psi} p(y_{1:t} \mid \psi) p(\psi).\end{aligned}\tag{2.2.25}$$

Taking logarithm gives

$$\operatorname{argmax}_{\psi} \log p(\psi \mid y_{1:t}) \propto \operatorname{argmax}_{\psi} \{\log p(y_{1:t} \mid \psi) + \log p(\psi)\}.\tag{2.2.26}$$

The MAP is essentially the MLE taking into account the parameter prior distribution $p(\psi)$. Parameter estimation can also be carried out by determining point estimates such as mean, median or mode of the marginal posterior distribution $p(\psi \mid y_{1:t})$. In terms of classification, the state-space model with unknown random variable parameters in addition to the original state thus makes the model general, regardless of whether the original model is linear Gaussian. Solutions based on classical filtering solutions such as the Weiner filter or Kalman filter for linear Gaussian state-space model are thus not applicable and numerical methods such as MCMC or particle filter are required.

2.2.3 Sequential solution of the linear Gaussian state-space model

From the state and observation equations (2.2.7) and (2.2.8) defining the linear Gaussian state-space model and the filtering result (2.2.16), Kalman filtering can be described. Each of the filtering distribution, one-step-ahead predictive distribution, and one-step-ahead predictive likelihood are normal distributions in the Kalman filter. They are defined as follows. Let the filtering distribution be distributed as $\mathcal{N}(m_t, C_t)$, the one-step-ahead predictive distribution be distributed as $\mathcal{N}(a_t, R_t)$, and the one-step-ahead likelihood be distributed as $\mathcal{N}(f_t, Q_t)$. The procedure to obtain the filtering distribution at time t from time $t - 1$ is shown below.

Algorithm 1 Kalman filtering

Input. filtering distribution at time point $t - 1$: m_{t-1}, C_{t-1}

Time t update rule:

one-step-ahead predictive distribution:

(Mean) $a_t \leftarrow G_t m_{t-1}$

(Covariance) $R_t \leftarrow G_t C_{t-1} G_t^\top + W_t$

one-step-ahead predictive likelihood:

(Mean) $f_t \leftarrow F_t a_t$

(Covariance) $Q_t \leftarrow F_t R_t F_t^\top + V_t$

Kalman gain:

$K_t \leftarrow R_t F_t^\top Q_t^{-1}$

State update:

(Mean) $m_t \leftarrow a_t + K_t[y_t - f_t]$

(Covariance) $C_t \leftarrow [I - K_t F_t] R_t$

Output. filtering distribution at time t : m_t, C_t

The Kalman filtering algorithm sequentially generates the filtering distribution at each time point beginning from $t = 1$, with the initial time corresponding to the prior distribution. The one-step-ahead predictive distribution is the filtering distribution one time before is propagated forward based on the state equation. The one-step-ahead predictive likelihood is created by taking the one-step-ahead predictive distribution and incorporating the observation equation. Lastly, in the filtering distribution, the one-step-ahead predictive distribution is corrected using the likelihood model. The correction of the one-step-ahead predictive distribution is made by a matrix that is known as Kalman gain, which is a weight of the extent of the prediction error impacts the correction. Here, the prediction error (also referred to as residuals or innovations), is simply the difference between the observations y_t and the mean one-step-ahead predictive likelihood f_t . When the model is good, the residuals are independent identically distributed normal variates. The signal-to-noise ratio W_t/V_t impacts the Kalman gain. In the case of the local level model, which is perhaps the simplest dynamic linear model, the state dimension is $p = 1$ and the model matrices are scalar with $G_t = F_t = [1]$. When $V_t = 0$, the observations dominate and the filtering distribution is completely determined by the observations only. In such case, the Kalman gain is 1 and $m_t = y_t$. When $V_t = \infty$, the observations become useless and the filtering distribution is determined by the one-step-ahead prediction and now the Kalman gain is 0 with $m_t = a_t$. When V_t lies between 0 and ∞ , the Kalman gain ranges between 1 and 0 adjusting the influence of observation equation and state equation.

The log-likelihood for the linear Gaussian state-space model can be derived from the one-step-ahead predictive likelihood.

$$\begin{aligned}
\ell(\psi) &= \log p(y_1, \dots, y_T; \psi) \\
&= \sum_{t=1}^T \log p(y_t \mid y_{1:t-1}; \psi) \\
&= -\frac{1}{2} \sum_{t=1}^T \log |Q_t| - \frac{1}{2} \sum_{t=1}^T \frac{(y_t - f_t)^2}{Q_t}.
\end{aligned} \tag{2.2.27}$$

Kalman prediction can be explained using the general framework developed in (2.2.17). The k -steps-ahead predictive distribution is normal with mean $a_t(k)$ and covariance $R_t(k)$. Obtaining the k -steps-ahead predictive distribution starting from time $t+k-1$ to time $t+k$ is outlined below.

Algorithm 2 Kalman prediction

Input. the $k-1$ -steps-ahead predictive distribution at time $t+(k-1)$: $a_t(k-1), R_t(k-1)$
Time $t+k$ update rule:
the k -steps-ahead predictive distribution
(Mean) $a_t(k) \leftarrow G_{t+k} a_t(k-1)$
(Covariance) $R_t(k) \leftarrow G_{t+k} R_t(k-1) G_{t+k}^\top + W_{t+k}$
Output. the k -steps-ahead predictive distribution at time $t+k$: $a_t(k), R_t(k)$

Kalman prediction from time $k=1$ gives the predictive distribution at time $t+1$ with the 0-steps-ahead predictive distribution equal to the filtering distribution at time t .

Fixed-interval smoothing requires Kalman filtering on the entire series. At time T , the smoothing and filtering distributions agree. The filtering distribution is then corrected based on the smoothing distribution one time point forward. The following is a Rauch-Tung-Striebel (RTS) algorithm [19] for Kalman smoothing.

Algorithm 3 Kalman smoothing

Input. smoothing distribution at time $t+1$: s_{t+1}, S_{t+1}
Time t update rule:
Smoothing gain
 $A_t \leftarrow C_t G_{t+1}^\top R_{t+1}^{-1}$
State update
(Mean) $s_t \leftarrow m_t + A_t[s_{t+1} - a_{t+1}]$
(Covariance) $S_t \leftarrow C_t + A_t[S_{t+1} - R_{t+1}]A_t^\top$
Output. smoothing distribution at time t : s_t, S_t

The Kalman filter can be nicely illustrated through a local-level model for analysis as:

$$\theta_t = \theta_{t-1} + w_t \quad w_t \sim \mathcal{N}(0, W), \quad (2.2.28)$$

$$y_t = \theta_t + v_t \quad v_t \sim \mathcal{N}(0, V), \quad (2.2.29)$$

from setting $p = 1, G_t = [1], F_t = [1]$ in the linear Gaussian state-space formulation (2.2.7) and (2.2.8) with time-invariance. The state θ_t fluctuates in random fashion over time. The local-level model is identical to AR(1) model with coefficient ϕ_1 set to one which gives non-stationary time series for the state. This is known as a random walk. Another name for the local-level model is a random walk plus noise model.

The log-likelihood for the linear Gaussian state-space model can be derived from the one-step-ahead predictive likelihood (2.2.20).

$$\begin{aligned} \ell(\psi) &= \log p(y_1, \dots, y_T; \psi) \\ &= \sum_{t=1}^T \log p(y_t \mid y_{1:t-1}; \psi) \\ &= -\frac{1}{2} \sum_{t=1}^T \log |Q_t| - \frac{1}{2} \sum_{t=1}^T \frac{(y_t - f_t)^2}{Q_t}. \end{aligned}$$

Prior distribution $p(\theta_0) = \mathcal{N}(m_0, C_0)$ along with state and observation equations define the joint model (2.2.4). Some common empirical priors include setting to one or the observation mean, with variance set to an arbitrarily large value according to the data. Parameter specification can be accomplished by maximum likelihood estimate as in (2.2.21). Standard error in the maximum likelihood estimation is the average fluctuation of the estimates under the hypothesis that all observations have identical statistical properties. It can be obtained from the Hessian matrix $\mathcal{H}$ of log-likelihood function:

$$\mathcal{H} = - \begin{bmatrix} \frac{\partial^2 \ell(W, V)}{\partial W^2} & \frac{\partial^2 \ell(W, V)}{\partial W \partial V} \\ \frac{\partial^2 \ell(W, V)}{\partial V \partial W} & \frac{\partial^2 \ell(W, V)}{\partial V^2} \end{bmatrix}. \quad (2.2.30)$$

Maximum likelihood asymptotic theory states that diagonal elements of $\mathcal{H}^{-1}(\ell)$ are the asymptotic variance for the corresponding parameter. Typically, the delta method is used to obtain standard errors of the parameter after an exponential transformation. Essentially,

this is a Taylor expansion about $\hat{\theta}$ truncated at first order,

$$\text{Var}[\exp(\theta)] \approx \text{Var} \left[\exp(\hat{\theta}) + \frac{\exp(\hat{\theta})}{1!}(\theta - \hat{\theta}) \right] \quad (2.2.31)$$

$$= \exp^2(\hat{\theta}) \text{Var}[\theta - \hat{\theta}] \quad (2.2.32)$$

$$= \exp^2(\hat{\theta}) \text{Var}[\theta]. \quad (2.2.33)$$

Hence, $\text{diag}(\mathcal{H}^{-1})$ are related to $\text{Var}[\theta]$ and standard errors for θ are obtained after exponential transformation,

$$\exp(\hat{\theta}) \times \sqrt{\text{diag}(\mathcal{H}^{-1})_{ij}}. \quad (2.2.34)$$

With respect to model diagnostics and model checking, the main methods of assessing model fit are through likelihood and prediction error. Larger likelihood is preferred in the linear Gaussian state space model, as given by (2.2.27). The difference between observations and the mean one-step-ahead predictive likelihood is also known as innovations, prediction error, or residuals and is defined as $e_t = y_t - \hat{f}_t$. Under the assumption of adequate model fit, the innovations follow identical and independent normal variates. The normalized innovations, $\frac{e_t}{\sqrt{Q_t}}$, follow identical and independent normal standard distribution. Thus, correlation and normality of the standard innovations provide a means of model assessment. Typically, this is done by plotting standardized residuals against time, plotting autocorrelation of residuals against lag variable and checking p -value for Ljung-Box statistic which posits that the data sequence is independent. In addition, quantile-quantile plot allows checking normality by taking the quantile of the theoretical normal distribution and plotting it against the quantile of the data. With regard to model diagnosis, a poor model is suggested by standardized innovations that have high residual autocorrelation coefficient. An unreasonably large standardized innovation may hint at a possible outlier or structural change in the series. Consideration of such a feature in the data through the appropriate model specification can lead to improved estimation accuracy.

2.2.4 Component linear Gaussian state-space model

It is possible to express a variety of individual component models in order to capture the different properties of the time series data. State-space models easily facilitate such decomposition to combine with the additive linear Gaussian state-space model.

The different K types of individual models in the Gaussian state-space models can be denoted with superscript k assigned to the variables:

$$\theta_t^{(k)} = G_t^{(k)} \theta_{t-1}^{(k)} + w_t^{(k)} \quad w_t^{(k)} \sim \mathcal{N}(0, W^{(k)}), \quad (2.2.35)$$

$$y_t^{(k)} = F_t^{(k)} \theta_t^{(k)} + v_t^{(k)} \quad v_t^{(k)} \sim \mathcal{N}(0, V^{(k)}), \quad (2.2.36)$$

with prior distribution, $\theta_0^{(k)} \sim \mathcal{N}(m_0^{(k)}, C_0^{(k)})$. The dimension of each state vector is allowed to be different for each of the K individual components.

In (2.2.7) and (2.2.8), the general specifications are as follows:

$$\theta_t = \begin{bmatrix} \theta_t^{(1)} \\ \vdots \\ \theta_t^{(K)} \end{bmatrix}, \quad G_t = \begin{bmatrix} G_t^{(1)} & & & \\ & G_t^{(2)} & & \\ & & \ddots & \\ & & & G_t^{(K)} \end{bmatrix}, \quad W_t = \begin{bmatrix} W_t^{(1)} & & & \\ & W_t^{(2)} & & \\ & & \ddots & \\ & & & W_t^{(K)} \end{bmatrix}, \quad (2.2.37)$$

$$y_t = \sum_{k=1}^K y_t^{(k)}, \quad F_t = [F_t^{(1)} \dots F_t^{(K)}], \quad V_t = \sum_{k=1}^K V_t^{(k)}, \quad (2.2.38)$$

with prior distribution,

$$\theta_0 = \begin{bmatrix} \theta_0^{(1)} \\ \vdots \\ \theta_0^{(K)} \end{bmatrix}, \quad m_0 = \begin{bmatrix} m_0^{(1)} \\ \vdots \\ m_0^{(K)} \end{bmatrix}, \quad C_0 = \begin{bmatrix} C_0^{(1)} & & & \\ & C_0^{(2)} & & \\ & & \ddots & \\ & & & C_0^{(K)} \end{bmatrix}.$$

Several types of model components are possible, including local-level, local-trend, seasonal, autoregressive moving average (ARMA), and regression models. Some details of ARMA have been provided. The following contains details of the other types of models.

Local-level, or random walk plus noise model, is useful for representing series which do not have a clear pattern and have small short term fluctuations:

$$\theta_t = \theta_{t-1} + w_t \quad w_t \sim \mathcal{N}(0, W), \quad (2.2.39)$$

$$y_t = \theta_t + v_t \quad v_t \sim \mathcal{N}(0, V). \quad (2.2.40)$$

Local-trend, or linear growth model can be used to capture linearity of the series such as

inclining or declining trends:

$$\theta_t^{(2)} = \theta_{t-1}^{(2)} + \theta_{t-1}^{(1)} + w_t^{(2)} \quad w_t^{(2)} \sim \mathcal{N}(0, W^{(2)}), \quad (2.2.41)$$

$$\theta_t^{(1)} = \theta_{t-1}^{(1)} + w_t^{(1)} \quad w_t^{(1)} \sim \mathcal{N}(0, W^{(1)}), \quad (2.2.42)$$

$$y_t = \theta_t^{(2)} + v_t \quad v_t \sim \mathcal{N}(0, V). \quad (2.2.43)$$

Local-level and local-trend models are polynomial models of order one and two, respectively, and are most commonly used since increase in the order gives rise to overfit of noise and poor interpretability. An equivalence exists between integrated random walk whereby $W^{(1)}, W^{(2)} = 0$ and the preprocessing difference operation in the autoregressive integrated moving average (ARIMA) model [14].

Seasonal or periodical model is useful for expressing periodic or cyclic pattern in the series. ARMA model can complement the polynomial and seasonal models which are applied to non-stationary components. It captures residual correlation by taking into account irregular fluctuations that remain after considering polynomial or seasonal components, such as unexpected movements in the series.

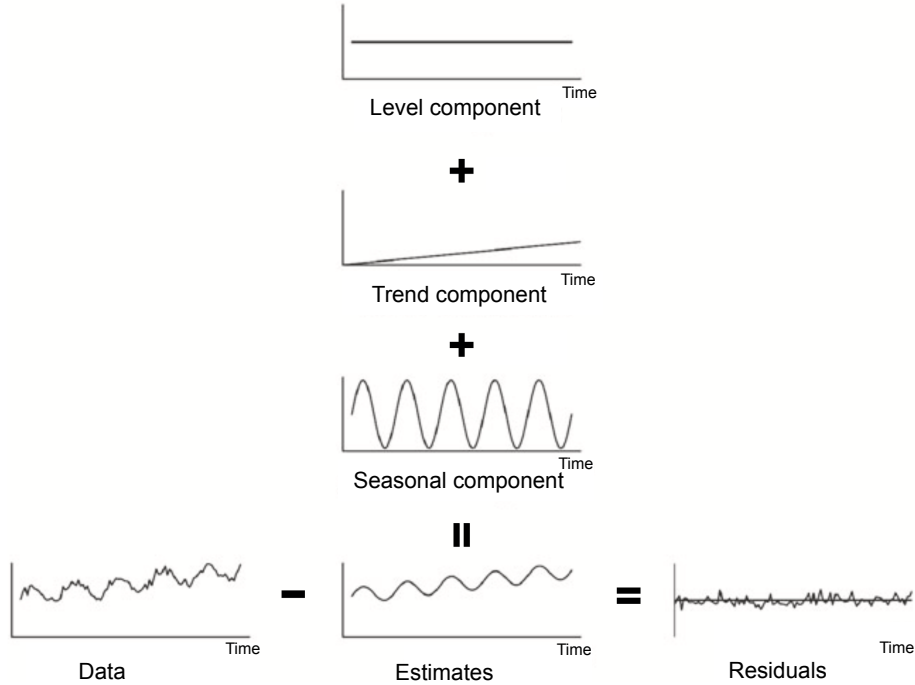


Figure 2.8: Decomposition of time series data.

2.2.5 Dynamic regression models

When two or more time series are available and each correlates in some way with the other, a regression component model can be used to enhance the explanatory and predictive capability of the time series under study. The time series data which correlates with the target time series is known as the explanatory variable. State-space models may be created by regarding a regression component as an individual model [20]. Take for example, a linear regression model:

$$y_t = \alpha_t + \beta_t^{(1)}x_t^{(1)} + \cdots + \beta_t^{(n)}x_t^{(n)} + \cdots + \beta_t^{(N)}x_t^{(N)} + \epsilon_t. \quad (2.2.44)$$

For the regression model specification, denote the regression intercept α_t and the regression coefficients $\beta_t^{(n)}$ and let them be uncorrelated stochastic time-varying parameters that evolve according to independent random walk. The $x_t^{(n)}$ are deterministic explanatory variables.

In this setting, the equations corresponding to (2.2.7) and (2.2.8) are:

$$\theta_t = \begin{bmatrix} \alpha_t \\ \beta_t^{(1)} \\ \vdots \\ \beta_t^{(N)} \end{bmatrix}, \quad G_t = I_{N+1}, \quad W_t = \begin{bmatrix} W^{(\alpha)} & & & \\ & W^{(1)} & & \\ & & \ddots & \\ & & & W^{(N)} \end{bmatrix}, \quad (2.2.45)$$

$$F_t = \begin{bmatrix} 1 & x_t^{(1)} & \cdots & x_t^{(N)} \end{bmatrix}, \quad V_t = V. \quad (2.2.46)$$

Standard static linear regression model is recovered when $W_t = 0$.

2.2.6 Covariance estimation in multivariate state-space models

In the analysis of multiple time series, the estimation of covariance matrices from the innovation terms in a multivariate state-space model is crucial for determining the cross-correlation among these series at each observation point.

Bayesian and maximum likelihood estimation for the time-varying covariance matrix of observable data is studied. Multivariate and matrix-variate distributions are employed to precisely define the covariance matrix. The derivation of explicit updates for both the one-step-ahead forecast distribution of the observed data vector and the posterior distribution of the state vector, unconditional of the observation covariance matrix, are provided. Furthermore, under the model assumptions, the posterior distribution of the observation covariance matrix is shown to follow an inverse Wishart process [1].

Unlike retrospective smoothing estimation via expectation-maximization algorithm, the methodology enables real-time learning and forecasting of cross-series covariances, aligning with the principles of the Kalman filter.

The methodology is applied to multivariate time series data in epidemiology [21]. The results are in agreement with those obtained from static estimation methods. Additional insights into the complex interplay between disease dynamics and surveillance data are gained by incorporating external climatological and wastewater data [22] as covariates in the state-space model framework.

Some basic results from ordinary least squares theory are provided. Suppose observations $y_1, \dots, y_n$ become available over time $t = 1, \dots, n$ and let p variables x_{it} form a column vector $x_t = [x_{1t}, \dots, x_{pt}]^\top$. Then the linear regression model is

$$y_t = x_{1t}\theta_1 + \dots + x_{pt}\theta_p + \epsilon_t = x_t^\top \theta + \epsilon_t. \quad (2.2.47)$$

By Gauss-Markov conditions, $\{\epsilon_t\}$ is a sequence of zero mean, common variance σ^2 independent random variables (white noise).

The regression estimator is of interest for statistical inference. Given realized observations $y_{1:n} = \{y_1, \dots, y_n\}$, estimation of θ is obtained by minimizing the sum of squares

$$S(\theta) = \sum_{t=1}^n (y_t - x_t^\top \theta)^2 \quad (2.2.48)$$

$$= (y - X\theta)^\top (y - X\theta). \quad (2.2.49)$$

Assuming X is full rank, the solution is

$$\hat{\theta} = (X^\top X)^{-1} X^\top y. \quad (2.2.50)$$

Regarding the distributional specification of innovations, let $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$, then

$$y_t \mid \theta, \sigma^2 \sim \mathcal{N}(x_t^\top \theta, \sigma^2). \quad (2.2.51)$$

The following are MLEs of θ and σ^2 :

$$\hat{\theta} = (X^\top X)^{-1} X^\top y \quad (2.2.52)$$

$$\hat{\sigma}^2 = \frac{1}{n} (y - X\hat{\theta})^\top (y - X\hat{\theta}) \quad (2.2.53)$$

$$= \frac{1}{n} y^\top (I - X(X^\top X)^{-1} X^\top) y. \quad (2.2.54)$$

The above estimation of θ is known as ordinary least squares (OLS), because all observations $y_1, \dots, y_n$ have on average the same contribution to $\hat{\theta}$.

For large n , distant observations (such as y_1 or y_2) have the same weight on $\hat{\theta}$ as more recent observations (such as y_n or y_{n-1}). This is usually not desirable when the data exhibit a time-dependence structure or when a localized estimation of θ is required where more weight is put on more recent observations in the estimation of $\hat{\theta}$ which motivates local regression.

In light of this, consider weighted (discounted) sum of squares [23]:

$$S(\theta) = \sum_{j=1}^{n-1} \delta^j (y_{n-j} - x_{n-j}^\top \theta)^2 \quad (2.2.55)$$

$$= (y_n - x_n^\top \theta)^2 + \delta (y_{n-1} - x_{n-1}^\top \theta)^2 + \dots + \delta^{n-1} (y_1 - x_1^\top \theta)^2, \quad (2.2.56)$$

where $0 < \delta \leq 1$ and $\delta^j \approx 0$ for sufficiently large j . At each time j , the data is forgotten at a rate of δ^j .

The memory of the model is the geometric series

$$\sum_{j=0}^{n-1} \delta^j = \frac{1 - \delta^n}{1 - \delta} \quad (2.2.57)$$

$$= \frac{1}{1 - \delta}, \quad \text{as } n \rightarrow \infty. \quad (2.2.58)$$

For illustration regarding the discounting or forgetting factor, consider the following:

- when $\delta = 1$ (OLS), the memory is equal to infinity; do not forget any data in the estimation of θ ,
- when $\delta = 0.5$, then the memory is equal to 2; use only the two most recent observations in the calculation of the new $\hat{\theta}$.

The Gauss-Plackett parameter recursions are then given by:

$$\hat{\theta} = (X^\top X)^{-1} X^\top y \quad (2.2.59)$$

$$= \left(\sum_{j=0}^{n-1} \delta^j x_{n-j} x_{n-j}^\top \right)^{-1} \sum_{j=0}^{n-1} \delta^j x_{n-j} y_{n-j} \quad (2.2.60)$$

$$= H_n^{-1} h_n, \quad (2.2.61)$$

where

$$H_t = \sum_{j=0}^{t-1} \delta^j x_{t-j} x_{t-j}^\top = x_t x_t^\top + \delta H_{t-1}, \quad (2.2.62)$$

$$h_t = \sum_{j=0}^{t-1} \delta^j x_{t-j} y_{t-j} = x_t y_t + \delta h_{t-1}. \quad (2.2.63)$$

The definitions for the recursions are as follows:

$$e_t = y_t - x_t^\top \hat{\theta}_{t-1} \quad (\text{one-step-ahead forecast error}), \quad (2.2.64)$$

$$r_t = y_t - x_t^\top \hat{\theta}_t \quad (\text{posterior residual error}), \quad (2.2.65)$$

$$K_t = H_t^{-1} x_t, \quad (2.2.66)$$

$$P_t = H_t^{-1}. \quad (2.2.67)$$

The full recursive least squares algorithm is provided below.

Algorithm 4 Recursive Least Squares

- 1: Input. $\hat{\theta}_0, P_0, n_0 = 0, \hat{\sigma}_0^2, \delta$
 - 2: Time t update rule:
 - 3: • *recursion update for inverse Gram matrix*
 - 4: $P_t \leftarrow \frac{1}{\delta} \left(I - \frac{P_{t-1} x_t x_t^\top}{\delta + x_t^\top P_{t-1} x_t} \right) P_{t-1}$
 - 5: • *state update*
 - 6: $\hat{\theta}_t \leftarrow \hat{\theta}_{t-1} + K_t e_t$
 - 7: • *variance update*
 - 8: $n_t \hat{\sigma}^2 \leftarrow n_{t-1} \hat{\sigma}_{t-1}^2 + r_t e_t, \quad n_t = n_{t-1} + 1$
 - 9: Output. parameter estimates at time t : $\hat{\theta}_t, \hat{\sigma}_t^2$
-

For the representation of multivariate state-space models important for correlation analysis, consider the following d -variate linear Gaussian state-space model:

$$\theta_t = G_t \theta_{t-1} + w_t \quad w_t \sim \mathcal{N}(0, W_t), \quad (2.2.68)$$

$$y_t = F_t \theta_t + v_t \quad v_t \sim \mathcal{N}(0, V_t). \quad (2.2.69)$$

where:

- G_t is a $p \times p$ state transition matrix
- F_t is a $d \times p$ design matrix

- W_t is a $p \times p$ covariance matrix for the state noise and
- V_t is a $d \times d$ covariance for the observation noise.

The prior distribution is $\theta_0 \sim \mathcal{N}(m_0, C_0)$ where m_0 is the p -dimensional mean vector and C_0 is a $p \times p$ covariance matrix.

The design and transition matrices characterize the forecast function, $y_{t+k} \mid y_t$, therefore their specification is critical for filtering, forecasting and smoothing. Each row of F_t describes the univariate component time series in the multivariate series y_t . As a simplifying assumption for the multivariate model, the structure of the component time series will belong to the same type of structural state-space model; i.e., the state vector θ_t is driven by the same transition matrix for each of the univariate component series.

Under the condition of time-invariance, noise covariances can be estimated by EM algorithm for parameters $\psi = \{m_0, C_0, W, V\}$.

$$\begin{aligned} \ell(\psi \mid \theta, y) = & -\frac{n}{2} \log |V| - \frac{1}{2} \sum_{t=1}^n (y_t - F_t \theta_t)^\top V^{-1} (y_t - F_t \theta_t) \\ & - \frac{n}{2} \log |W| - \frac{1}{2} \sum_{t=1}^n (\theta_t - G_t \theta_{t-1})^\top W^{-1} (\theta_t - G_t \theta_{t-1}) \\ & - \frac{1}{2} \log |C_0| - \frac{1}{2} \sum_{t=1}^n (\theta_0 - m_0)^\top C_0^{-1} (\theta_0 - m_0). \end{aligned} \quad (2.2.70)$$

That is, static parameter estimation is obtained as follows [24].

Algorithm 5 EM algorithm

- 1: Input:
 - 2: • *initial parameter estimate* $i = 0$:
 - 3: $\psi^{(0)} = [\text{vech}(V^{(0)})^\top, \text{vech}(W^{(0)})^\top, (m_0^{(0)})^\top, \text{vech}(C_0^{(0)})^\top]^\top$;
 - 4: For iteration $i = 0, 1, 2, \dots, N - 1$:
 - 5: For each $t = 0, 1, \dots, n$:
 - 6: • *compute* m_t, C_t *conditioned on* $\psi^{(i)}$ *via smoothing*
 - 7: • *compute covariance matrices*
 - 8: $V^{(i+1)} = \frac{1}{n} \sum_{t=1}^n [(y_t - F_t \theta_t)(y_t - F_t \theta_t)^\top + F_t C_t F_t^\top]$
 - 9: $W^{(i+1)} = \frac{1}{n} \sum_{t=1}^n [(\theta_t - G_t \theta_{t-1})(\theta_t - G_t \theta_{t-1})^\top +$
 $+ C_t - C_{t,t-1} G_t^\top - G_t C_{t,t-1}^\top + G_t C_{t-1} G_t^\top]$
 - 10: $C_0^{(i+1)} = \frac{1}{n} [(\theta_0 - m_0)(\theta_0 - m_0)^\top + C_0]$
 - 11: Output:
 - 12: • *parameter estimate at iteration* $i + 1$:
 - 13: $\psi^{(i+1)} = [\text{vech}(V^{(i+1)})^\top, \text{vech}(W^{(i+1)})^\top, (m_0^{(i+1)})^\top, \text{vech}(C_0^{(i+1)})^\top]^\top$;
-

Covariance structure is determined by a complex interplay between some matrix and multivariate statistical distributions. First, consider the multivariate normal distribution [25]. Let $X = [X_1, \dots, X_p]^\top$ denote a random vector. Then $X \sim \mathcal{N}(\mu, \Sigma)$ with p.d.f. given by

$$p(x) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left[-\frac{1}{2} (x - \mu)^\top \Sigma^{-1} (x - \mu) \right]. \quad (2.2.71)$$

The matrix-variate normal distribution can be defined as follows. Let $X_1, X_2, \dots, X_d$ be p -dimensional random column vectors and let $p \times d$ random matrix $X = [X_1, X_2, \dots, X_d]$. Then X follows a matrix-variate normal distribution if its density is

$$p(X) = c_1 \exp \left\{ -\frac{1}{2} \text{trace}[(X - M)^\top P^{-1} (X - M) \Sigma^{-1}] \right\}, \quad (2.2.72)$$

where $c_1 = (2\pi)^{-\frac{dp}{2}} |P|^{-\frac{d}{2}} |\Sigma|^{-\frac{p}{2}}$, M is a $p \times d$ mean matrix and P and Σ are $p \times p$ (left) and $d \times d$ (right) covariance matrices, respectively. That is,

$$X \sim \mathcal{N}(M, P, \Sigma).$$

Introducing the necessary notation of matrix operators, let $\text{vec}(\cdot)$ denote the column stacking operator of an unrestricted matrix. For $m \times n$ matrix A , $\text{vec}(A)$ is an $mn \times 1$ vector and $\text{vech}(\cdot)$ denotes the column stacking operator with distinct elements only (repeated

entries are omitted). For $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix}$, $\text{vech}(A) = \begin{bmatrix} a_{11} \\ a_{12} \\ a_{22} \end{bmatrix}$ where $\otimes$ denotes the Kronecker

product of two matrices: For $m \times n$ matrix A and $p \times q$ matrix B , $A \otimes B = [a_{ij} B]$ is an $mp \times nq$ matrix.

As an important property of the matrix-variate normal, note that if $X \sim \mathcal{N}(M, P, \Sigma)$, then the random vector

$$Y = \text{vec}(X) \sim \mathcal{N}[\text{vec}(M), \Sigma \otimes P]. \quad (2.2.73)$$

Then M is mean matrix of X and $\Sigma \otimes P$ is covariance matrix of Y .

The inverse Wishart distribution is defined next. Let Σ be a $d \times d$ random symmetric, positive definite covariance matrix. Then $\Sigma \sim IW(n, S)$ with p.d.f. given by

$$p(\Sigma) = c_2 |\Sigma|^{-\frac{(n+d+1)}{2}} \exp \left[-\frac{1}{2} \text{trace}(S \Sigma^{-1}) \right], \quad (2.2.74)$$

where $c_2 = 2^{-\frac{nd}{2}} \Gamma_d(\frac{n}{2})^{-1} |S|^{\frac{n}{2}}$, $n > d - 1$ degrees of freedom and $\Gamma_d(\cdot)$ is a multivariate gamma function. The inverse Wishart is a generalization of the inverse gamma distribution.

Matrix-variate Student t distribution is defined as follows. Let X be a $p \times d$ random matrix. Then $X \sim t(n, M, P, S)$ with p.d.f. given by

$$p(X) = c_3 |I + (X - M)P^{-1}(X - M)^T S^{-1}|^{-\frac{(n+p+d-1)}{2}}, \quad (2.2.75)$$

where $c_3 = \Gamma_d[\frac{(n+p+d-1)}{2}] \pi^{-\frac{dp}{2}} \Gamma_d[\frac{(n+d-1)}{2}]^{-1} |P|^{-\frac{d}{2}} |S|^{-\frac{p}{2}}$. See (Gupta and Nagar, 1999) [26] which contains further details relevant to the above discussion.

For model definition and inference, consider the following. This model is essentially a matrix version of the multivariate state-space model [27].

$$\theta_t = G_t \theta_{t-1} + w_t \quad w_t \sim \mathcal{N}(0, W_t, V_t), \quad (2.2.76)$$

$$y_t^T = F_t \theta_t + v_t^T \quad v_t \sim \mathcal{N}(0, V_t). \quad (2.2.77)$$

where G_t is a $p \times p$ state transition matrix, F_t is a p -dimensional vector, w_t is a $p \times d$ state innovation and v_t is a d -dimensional observation innovation with $\text{vec}(w_t) \sim \mathcal{N}(0, V_t \otimes W_t)$.

The prior specification takes the form,

$$\theta_0 \mid V_t \sim \mathcal{N}(m_0, C_0, V_t), \quad (2.2.78)$$

$$V_t \sim IW(n_0, n_0 S_0), \quad (2.2.79)$$

where m_0 is a $p \times d$ mean matrix, C_0 is a $p \times p$ covariance matrix, and $n_0 S_0$ is a $d \times d$ scale covariance matrix with $n_0 > d - 1$ degrees of freedom.

Conditional distributions are as follows:

$$y_t = \text{vec}(y_t^T) = (I \otimes F_t) \text{vec}(\theta_t) + w_t \quad w_t \sim \mathcal{N}(0, V_t), \quad (2.2.80)$$

$$\text{vec}(\theta_t) = (I \otimes G_t) \text{vec}(\theta_{t-1}) + \text{vec}(v_t) \quad \text{vec}(v_t) \sim \mathcal{N}(0, V_t \otimes W_t). \quad (2.2.81)$$

By Kalman filter,

$$\text{vec}(\theta_t) \mid V_t, y_{1:t} \sim \mathcal{N}[\text{vec}(m_t), V_t \otimes C_t], \quad (2.2.82)$$

$$\theta_t \mid y_{1:t} \sim \mathcal{N}(m_t, C_t, V_t). \quad (2.2.83)$$

The posterior of observation covariances is

$$p(V_t \mid y_{1:t}) \propto p(y_t \mid V_t, y_{1:t-1}) \times p(V_t \mid y_{1:t-1}) \quad (2.2.84)$$

$$\propto |V_t|^{-\frac{1}{2}} \exp \left(-\frac{1}{2Q_t} [(y_t - f_t)^\top V_t^{-1} (y_t - f_t)] \right) \quad (2.2.85)$$

$$\times |V_t|^{-\frac{(n_{t-1}+d+1)}{2}} \exp \left[-\frac{1}{2} \text{trace}[n_{t-1} S_{t-1} V_t^{-1}] \right] \quad (2.2.86)$$

$$\propto |V_t|^{-\frac{(n_{t-1}+d+1)}{2}} \exp \left[-\frac{1}{2} \text{trace}[(Q_t^{-1} (y_t - f_t)(y_t - f_t)^\top + n_{t-1} S_{t-1}) V_t^{-1}] \right] \quad (2.2.87)$$

$$= |V_t|^{-\frac{(n_{t-1}+d+1)}{2}} \exp \left[-\frac{1}{2} \text{trace}[(y_t - F_t m_t)(y_t - f_t)^\top + n_{t-1} S_{t-1}) V_t^{-1}] \right], \quad (2.2.88)$$

$$V_t \mid y_{1:t} \sim IW(n_t, n_t S_t), \quad n_t = n_{t-1} + 1, n_t S_t = n_{t-1} S_{t-1} + (y_t - F_t m_t)(y_t - f_t). \quad (2.2.89)$$

Unconditional distributions yield the model inference as follows:

$$p(\theta_t \mid y_{1:t}) = \int_{V_t > 0} p(\theta_t \mid V_t, y_{1:t-1}) dV_t \quad (2.2.90)$$

$$= \int p(\theta_t \mid V_t, y_{1:t}) \times p(V_t \mid y_{1:t}) dV_t \quad (2.2.91)$$

$$\propto \int |V_t|^{-\frac{p}{2}} \exp \left(-\frac{1}{2} \text{trace}[(\theta_t - m_t) C_t^{-1} (\theta_t - m_t)^\top V_t^{-1}] \right) \quad (2.2.92)$$

$$\times |V_t|^{-\frac{(n_{t-1}+d+1)}{2}} \exp \left[-\frac{1}{2} \text{trace}[n_{t-1} S_{t-1} V_t^{-1}] \right] dV_t \quad (2.2.93)$$

$$= \int |V_t|^{-\frac{(n_{t-1}+d+1)}{2}} \quad (2.2.94)$$

$$\times \exp \left[-\frac{1}{2} \text{trace}[(\theta_t - m_t) C_t^{-1} (\theta_t - m_t)^\top + n_{t-1} S_{t-1}) V_t^{-1}] \right] dV_t \quad (2.2.95)$$

$$\propto |(\theta_t - m_t) C_t^{-1} (\theta_t - m_t)^\top + n_t S_t|^{-\frac{(n_t+p)}{2}} \quad (2.2.96)$$

$$\propto |I + n_t^{-1} (\theta_t - m_t) C_t^{-1} (\theta_t - m_t)^\top S_t|^{-\frac{(n_t-d+1+p+d-1)}{2}}, \quad (2.2.97)$$

$$\theta_t \mid y_{1:t} \sim t(n_t - d + 1, m_t, C_t, n_t S_t). \quad (2.2.98)$$

A maximum likelihood estimator is obtained by maximizing the log likelihood of the

data:

$$\ell(V_t; y_{1:t}) = \log(y_1, y_2, \dots, y_n, | V_t) \quad (2.2.99)$$

$$= \log \prod_{t=1}^n p(y_t | V_t, y_{1:t-1}) \quad (2.2.100)$$

$$= -\frac{dn}{2} \log(2\pi) - \frac{d}{2} \sum_{t=1}^n \log |Q_t| - \frac{n}{2} \log |V_t| \quad (2.2.101)$$

$$- \frac{1}{2} Q_t^{-1} \text{trace}[(y_t - f_t)(y_t - f_t)^\top V_t^{-1}], \quad (2.2.102)$$

$$\frac{\partial \ell(V_t; y_{1:t})}{\partial V_t} = -nV_t^{-1} + \frac{n}{2} \text{diag}(V_t^{-1}) \quad (2.2.103)$$

$$- \frac{1}{2} \sum_{t=1}^n Q_t^{-1} \left[-2V_t^{-1}(y_t - f_t)(y_t - f_t)^\top V_t^{-1} + \text{diag}(V_t^{-1}(y_t - f_t)(y_t - f_t)^\top V_t^{-1}) \right], \quad (2.2.104)$$

$$nV_t^{-1} = \sum_{t=1}^n Q_t^{-1} V_t^{-1} (y_t - f_t)(y_t - f_t)^\top V_t^{-1} \quad (2.2.105)$$

$$V_t = \frac{1}{n} \sum_{t=1}^n Q_t^{-1} (y_t - f_t)(y_t - f_t)^\top = \frac{1}{n} \sum_{t=1}^n (y_t - F_t m_t)(y_t - f_t)^\top. \quad (2.2.106)$$

Due to inverse Wishart posterior distribution of V_t , the estimator has convenient properties:

$$\text{mode}(V_t | y_{1:t}) = \frac{n_t S_t}{n_t + d + 1}, \quad (2.2.107)$$

$$\text{E}(V_t | y_{1:t}) = \frac{n_t S_t}{n_t - d - 1} \quad (2.2.108)$$

$$\approx \frac{1}{t} \sum_{i=1}^t (y_i - F_i m_i)(y_i - f_i)^\top \text{ as } t \rightarrow \infty. \quad (2.2.109)$$

The result for dynamic covariance estimation [5], [15], [28] is shown below.

Algorithm 6 Dynamic Covariance Estimation

- 1: Input:
 - 2: • *prior distributions at time $t = 0$:*
 - 3: $\theta_0 \sim t(n_0 - d + 1, m_0, C_0, n_0 S_0)$;
 - 4: $V_t \sim IW(n_0, n_0 S_0)$.
 - 5: Time t update rule:
 - 6: • *one-step-ahead predictive distribution*
 - 7: $\theta_t \mid y_{1:t} \sim t(n_t - d + 1, m_t, C_t, n_t S_t)$
 - 8: • *one-step-ahead predictive likelihood*
 - 9: $y_t \mid y_{1:t-1} \sim t(n_{t-1} - d + 1, f_t, n_{t-1} Q_t S_{t-1})$
 - 10: Output:
 - 11: • *posterior distribution of V_t at time t :*
 - 12: $V_t \mid y_{1:t} \sim IW(n_t, n_t S_t)$
 - 13: $n_t = n_{t-1} + 1$
 - 14: $n_t S_t = n_{t-1} S_{t-1} + [y_t - F_t m_t][y_t - f_t]$
-

2.3 Illustrations using COVID-19 data

2.3.1 Application of component linear Gaussian state-space model to case surveillance data

We consider COVID-19 case data from the province of Ontario, Canada [21] as shown below.

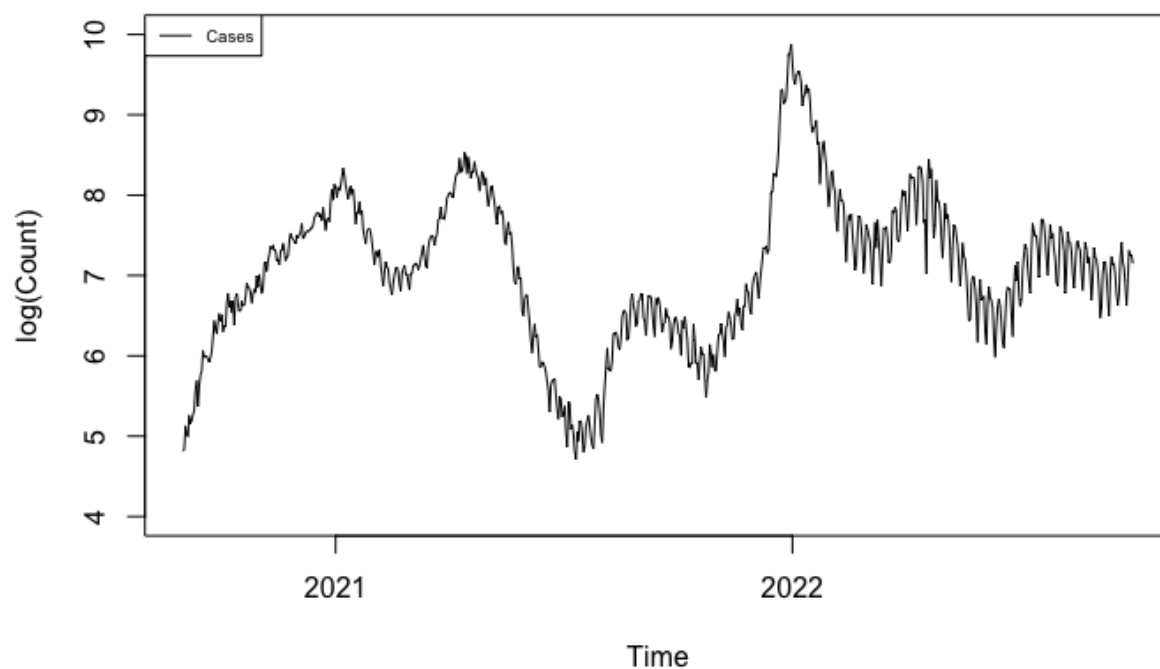


Figure 2.9: COVID-19 cases of Ontario, Canada. The data are sampled on a daily basis and shown on a $\log_e$ scale. [21]

We can perform a component model analysis of the COVID-19 case data with a linear Gaussian state-space model. The data show periodicity, segments of different trend and random fluctuation along short time intervals.

The following models can be implemented:

- (A) Local trend + periodic;
- (B) Local level + periodic;
- (C) Local trend + periodic (frequency domain);
- (D) Local level + periodic (frequency domain);
- (E) Local level + periodic + AR(1).

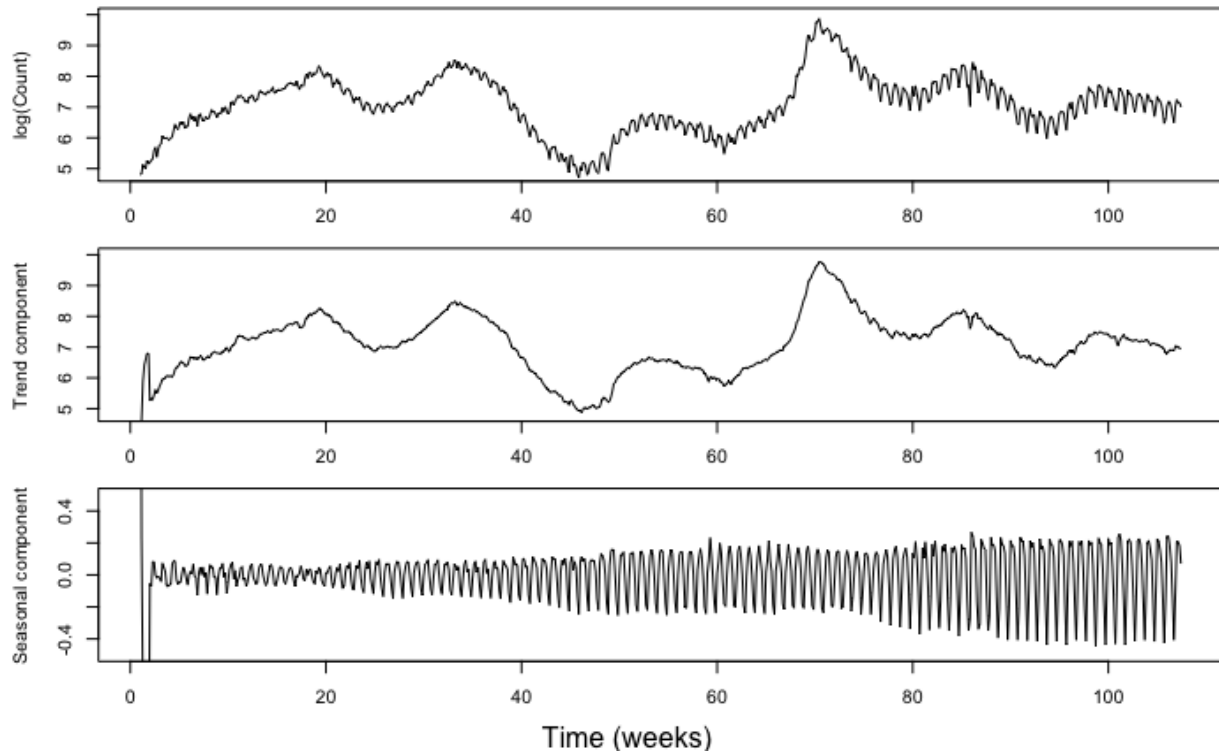


Figure 2.10: Local trend + periodic component plots.

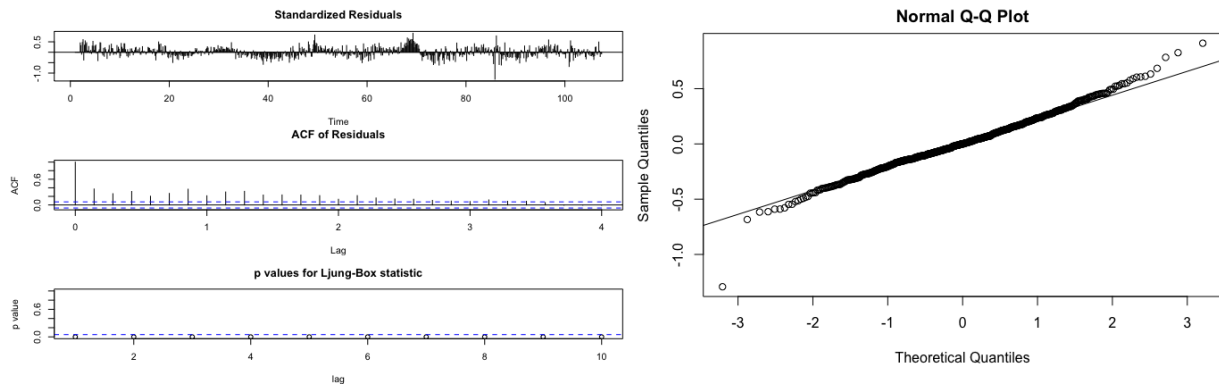


Figure 2.11: Local trend + periodic model validation and diagnostic plots.

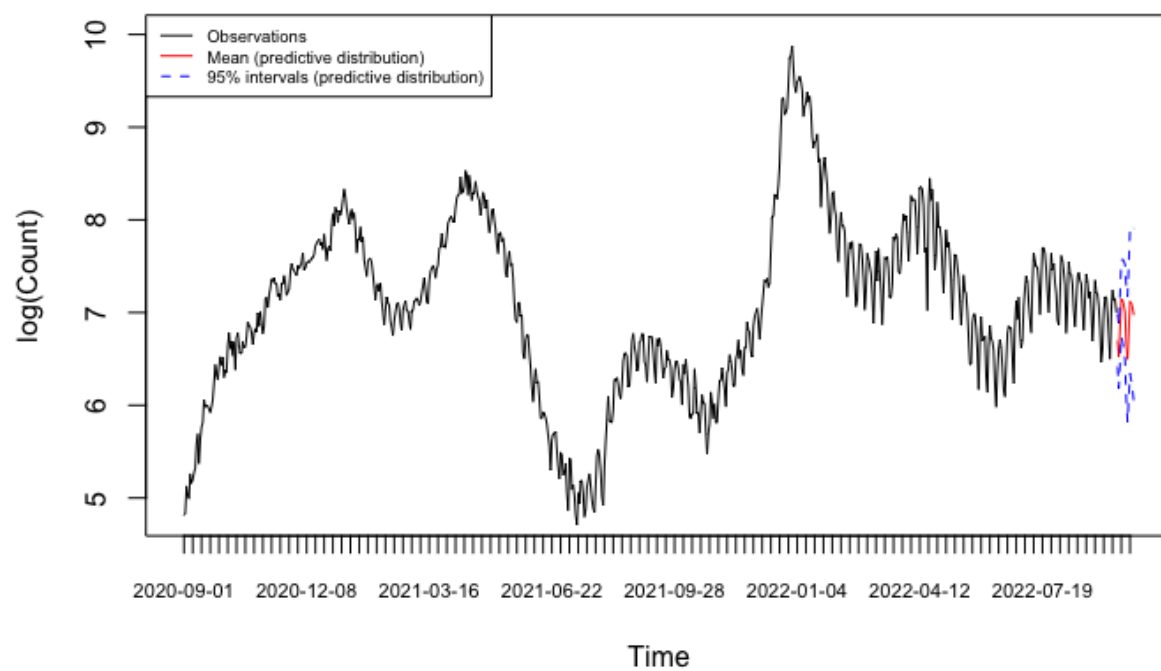


Figure 2.12: Local trend + periodic model Kalman prediction 14-day forecast.

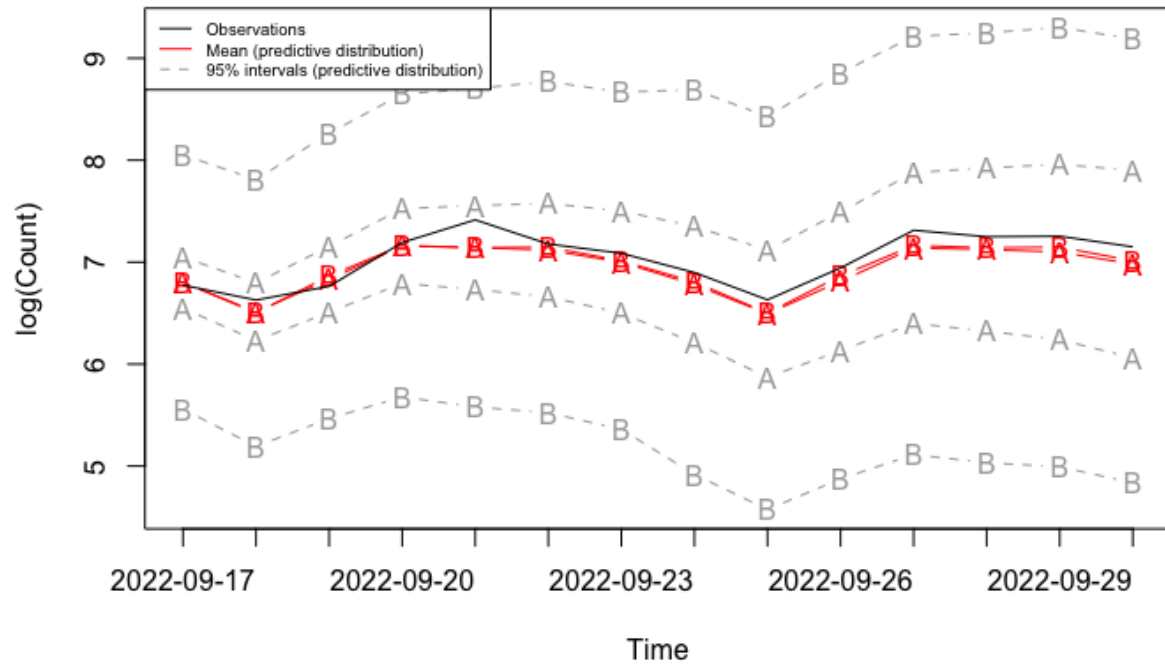


Figure 2.13: Kalman prediction 14-day forecast are compared for two competing models (A) and (B).

Model evaluation is by log-likelihood, mean absolute percent error and Theil's U .

Model	Log-likelihood	MAPE	Theil's U
(A) Trend + periodic	1077.19	0.01684	0.65349
(B) Level + periodic	253.97	0.01484	0.75959
(C) Trend + periodic (freq.)	1120.18	0.01755	0.63109
(D) Level + periodic (freq.)	1095.49	0.01709	0.66279
(E) Level + periodic + AR(1)	29.33	0.01399	0.81853

Table 2.1: Model comparison via log-likelihood, MAPE and Theil's U .

2.3.2 Dynamic regression models

Dynamic regression models may be used to study the following relationships:

1. COVID-19 hospitalization and case counts;
2. COVID-19 case counts and type of workday;
3. COVID-19 wastewater signal measurement and case counts.

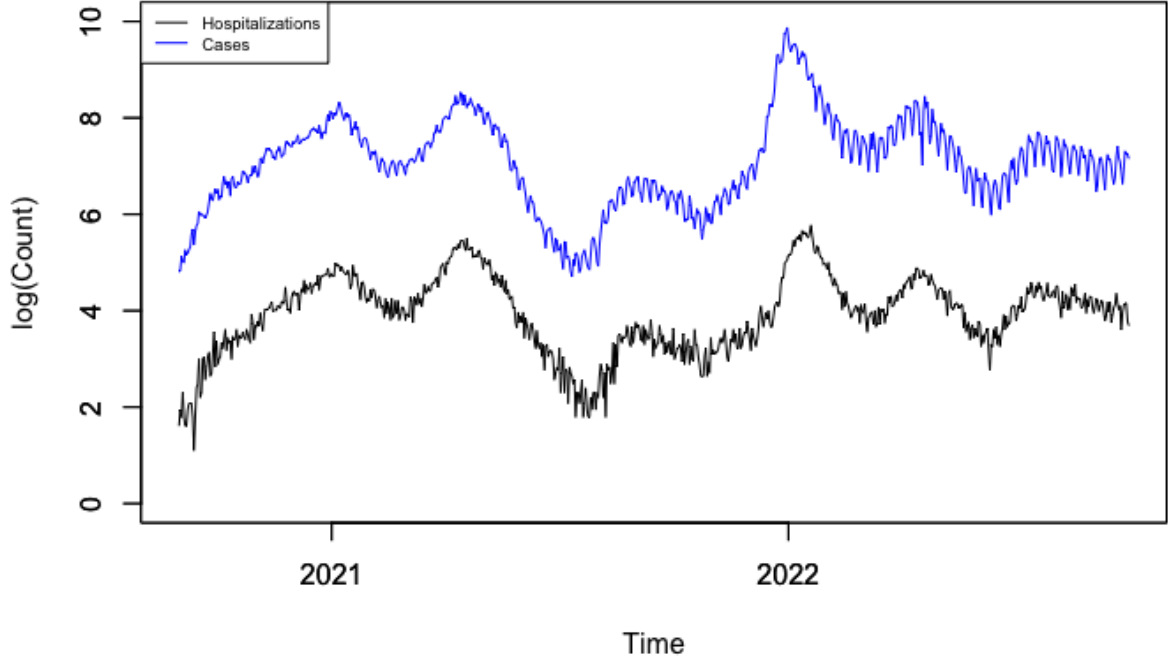


Figure 2.14: Ontario COVID-19 hospitalizations and cases.[21]

In the first application, the dynamic regression component is:

$$\theta_t^{reg} = \begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix}, \quad G_t^{reg} = I_2, \quad W_t^{reg} = \begin{bmatrix} W^{(\alpha)} & 0 \\ 0 & W^{(1)} \end{bmatrix}, \quad (2.3.1)$$

$$F_t^{reg} = \begin{bmatrix} 1 & x_t \end{bmatrix}, \quad V_t^{reg} = V, \quad (2.3.2)$$

where x_t represents the explanatory variable of case counts.

The periodical component is defined by:

$$\theta_t^{per} = \begin{bmatrix} \theta_t^{(1)} \\ \theta_t^{(2)} \\ \theta_t^{(3)} \\ \theta_t^{(4)} \\ \theta_t^{(5)} \\ \theta_t^{(6)} \end{bmatrix}, \quad G_t^{per} = \begin{bmatrix} -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \quad (2.3.3)$$

$$W_t^{per} = \begin{bmatrix} W & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, F_t^{per} = [1, 0, 0, 0, 0, 0], V_t^{per} = V'. \quad (2.3.4)$$

The autoregressive component is defined by:

$$\theta_t^{AR} = \begin{bmatrix} \theta_t^{(1)} \\ \theta_t^{(2)} \\ \vdots \\ \theta_t^{(r-1)} \\ \theta_t^{(r)} \end{bmatrix}, \quad G_t^{AR} = \begin{bmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \\ \phi_{r-1} & 0 & \cdots & 0 & 1 \\ \phi_r & 0 & \cdots & 0 & 0 \end{bmatrix}, \quad (2.3.5)$$

$$W_t^{AR} = RR^\top \sigma_\epsilon^2, \quad F_t^{AR} = [1, 0, \dots, 0, 0], \quad V_t^{AR} = 0. \quad (2.3.6)$$

Here $R = [1, \psi_1, \dots, \psi_{r-2}, \psi_{r-1}]^\top$, $w_t = R\epsilon_t$ and ϵ_t is white noise with variance σ_ϵ^2 . For analysis, the simple AR(1) model will be used: $y_t = \phi y_{t-1} + \epsilon_t$.

A block structured time-varying seasonal autoregressive model is obtained by combining the appropriate individual components. In (2.2.7) and (2.2.8), the general specifications are as follows:

$$\theta_t = \begin{bmatrix} \theta_t^{reg} \\ \theta_t^{per} \\ \theta_t^{AR} \end{bmatrix}, G_t = \begin{bmatrix} G_t^{reg} & & \\ & G_t^{per} & \\ & & G_t^{AR} \end{bmatrix}, W_t = \begin{bmatrix} W_t^{reg} & & \\ & W_t^{per} & \\ & & W_t^{AR} \end{bmatrix}, \quad (2.3.7)$$

$$y_t = y_t^{reg} + y_t^{per} + y_t^{AR}, \quad F_t = [F_t^{reg}, F_t^{per}, F_t^{AR}], \quad V_t = V + V', \quad (2.3.8)$$

with prior distribution,

$$\theta_0 = \begin{bmatrix} \theta_0^{reg} \\ \theta_0^{per} \\ \theta_0^{AR} \end{bmatrix}, m_0 = \begin{bmatrix} m_0^{reg} \\ m_0^{per} \\ m_0^{AR} \end{bmatrix}, C_0 = \begin{bmatrix} C_t^{reg} & & \\ & C_0^{per} & \\ & & C_0^{AR} \end{bmatrix}.$$

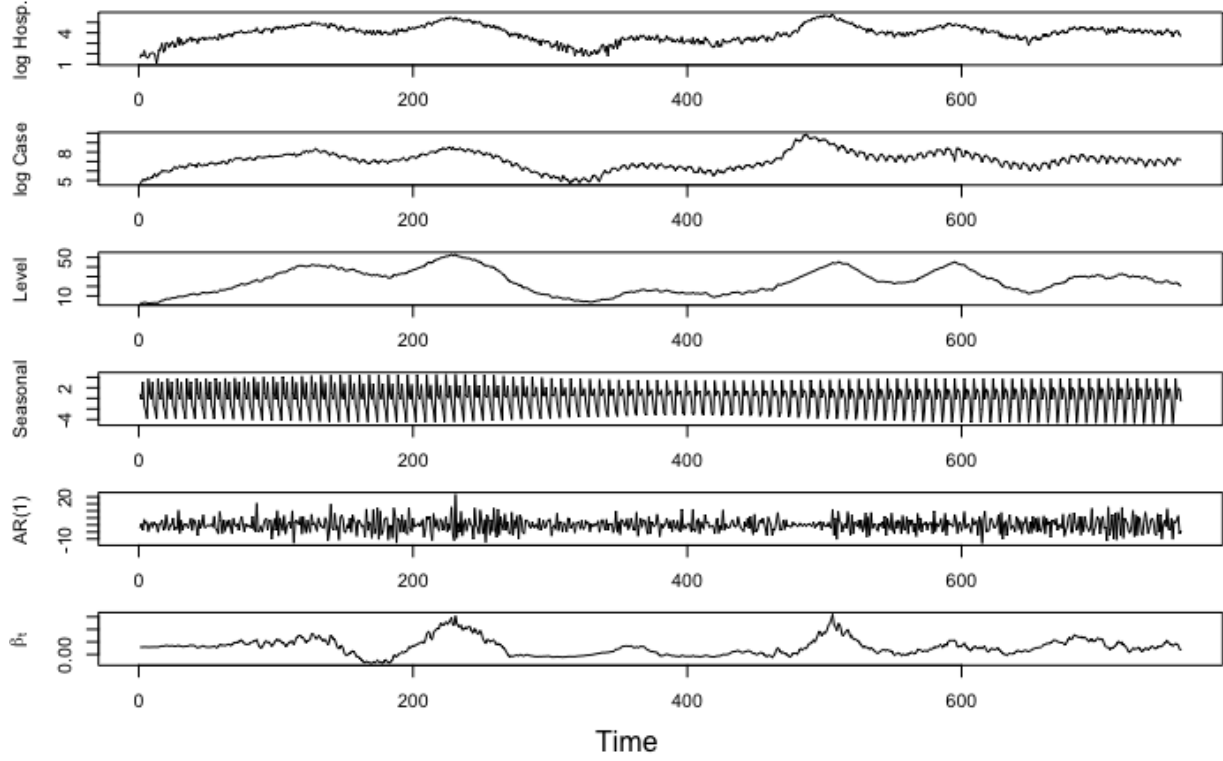


Figure 2.15: Ontario COVID-19 hospitalizations and cases filtering result. The explanatory and dependent variables are shown in the first two panels while model level, seasonal, autoregressive and regression components are shown in the subsequent panels.

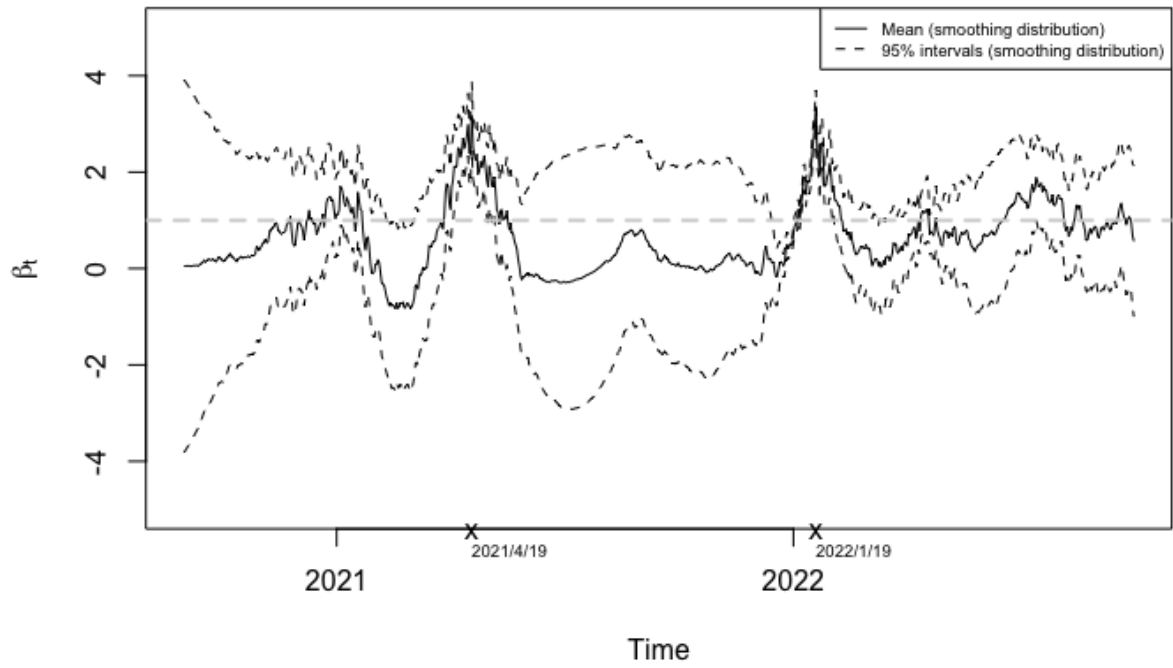


Figure 2.16: Time-varying regression coefficient β_t plot as defined in θ_t^{reg} .

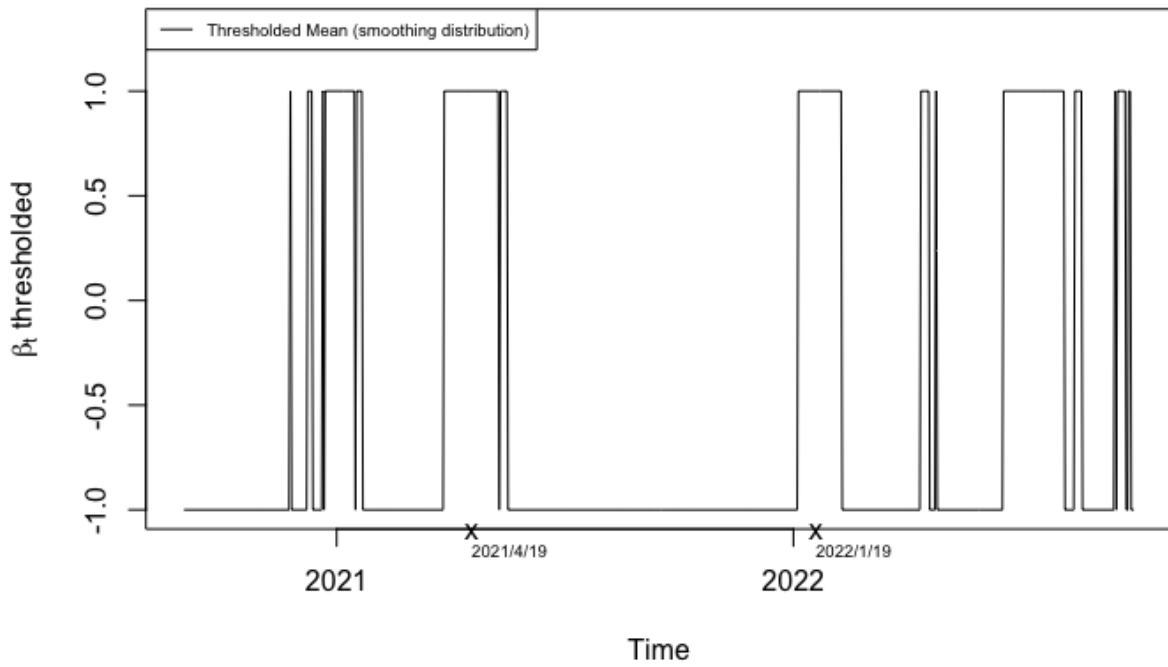


Figure 2.17: Thresholded β_t plot.

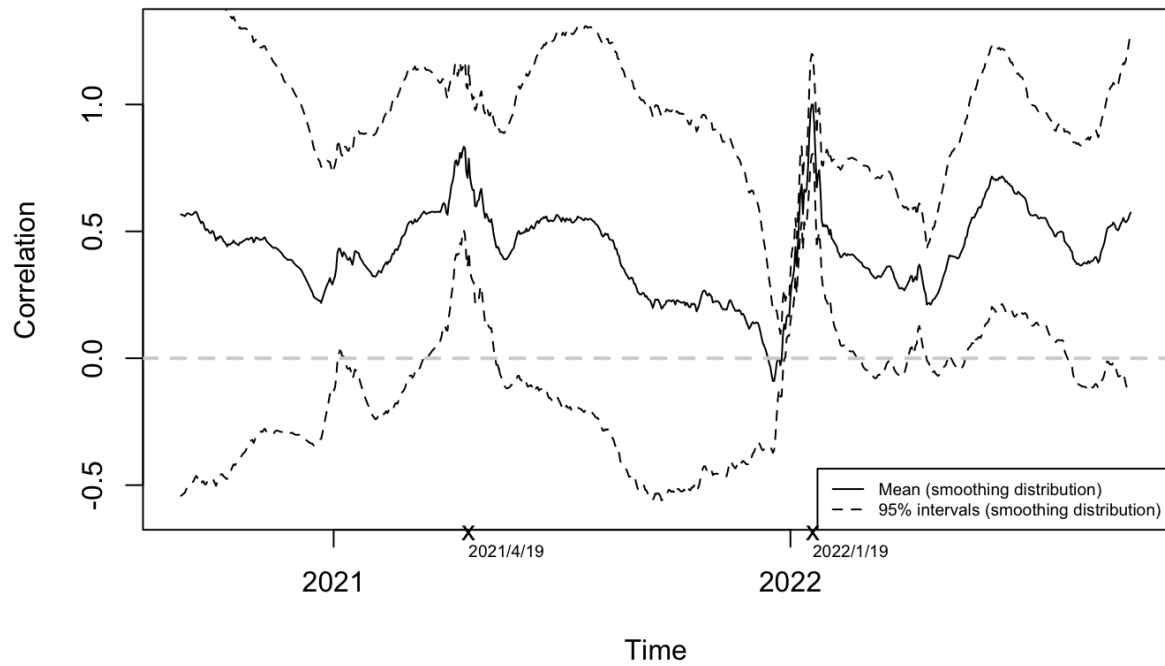


Figure 2.18: Time-varying correlation plot.

Cross-correlation and lagged scatter plots provide an alternative approach to modelling the data under consideration. A significant drawback is the subjective determination of the number of lagged case counts to include in the regression analysis.

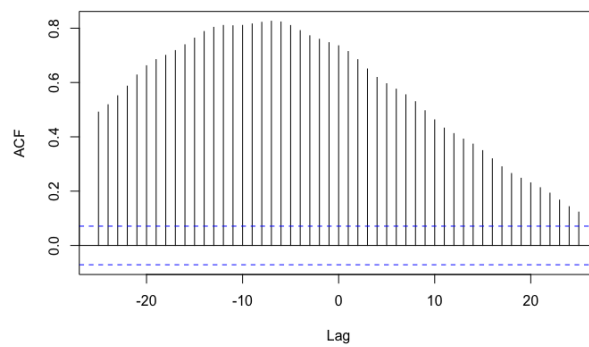


Figure 2.19: Cross-correlation function between cases and hospitalizations plot.

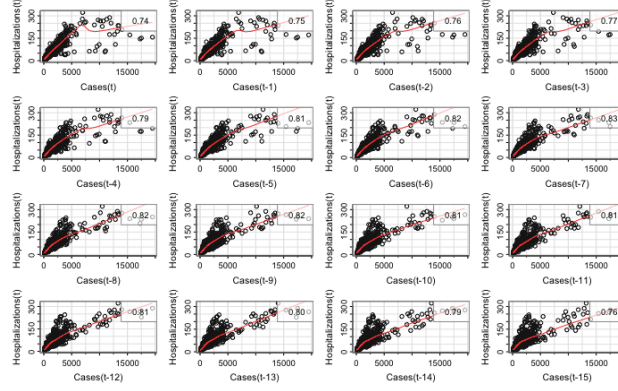


Figure 2.20: Scatterplots useful for determining lag number.

As a second application, consider Ontario COVID-19 cases by workday.

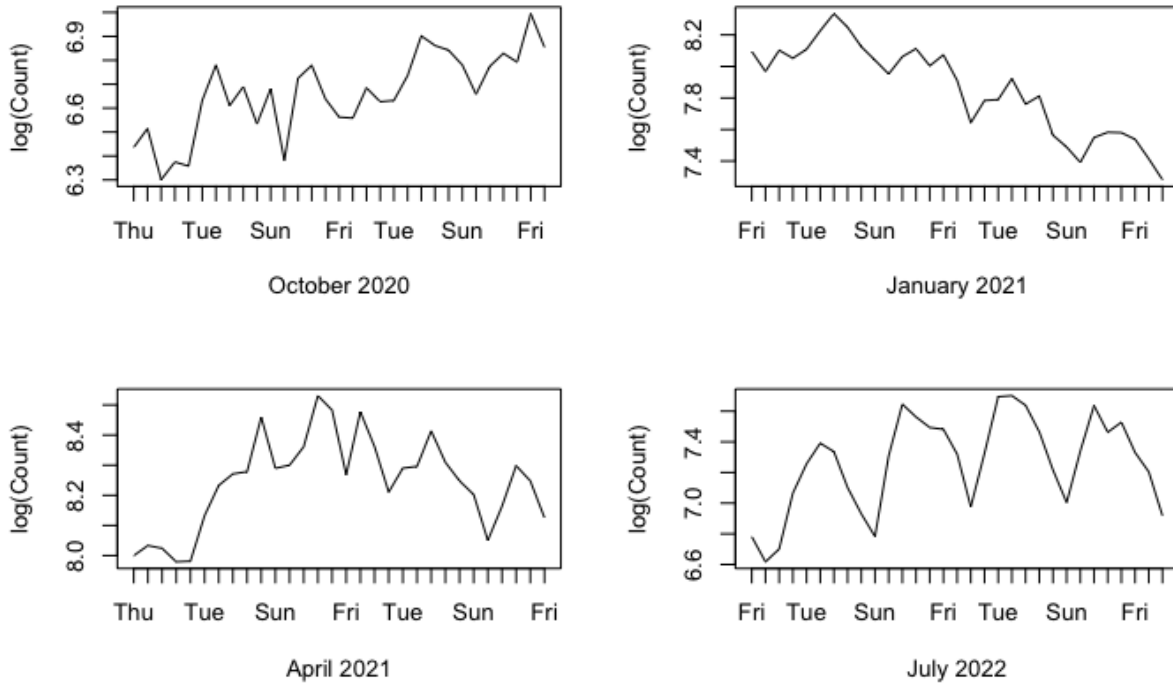


Figure 2.21: COVID-19 cases by workday.

An intervention variable for workday is formed and a dynamic regression model with explanatory variable

$$x_t = \begin{cases} 1 & \text{for non-workday} \\ 0 & \text{for workday} \end{cases}$$

is used to analyze the relationship between case counts and type of workday.

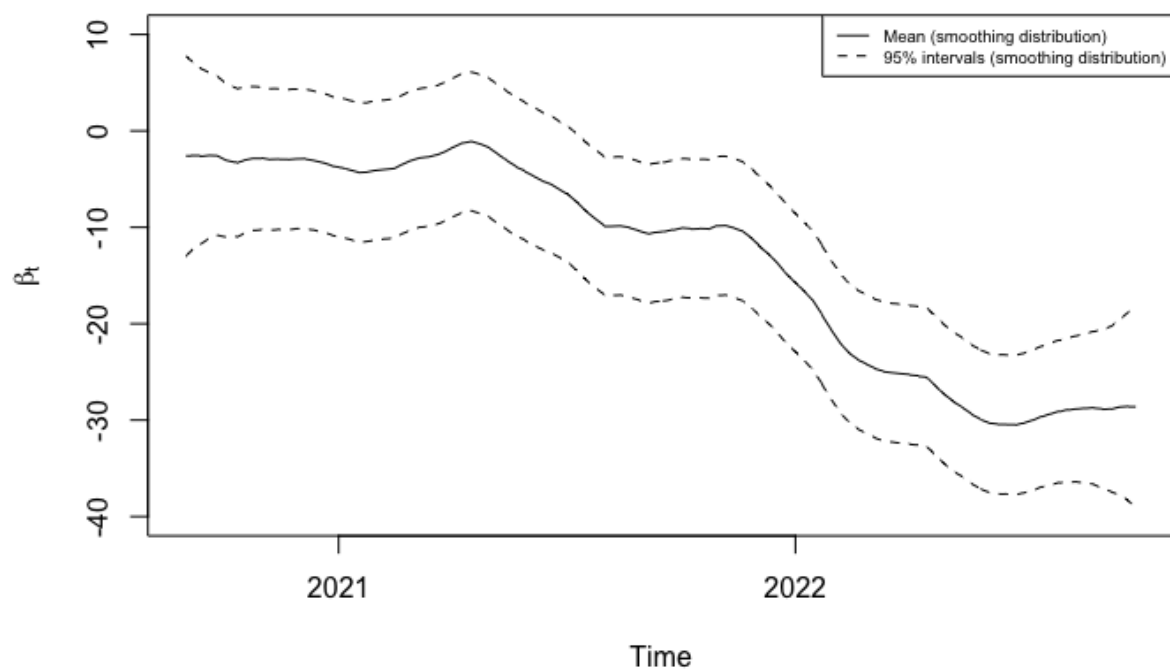


Figure 2.22: Percent change in cases associated with a weekend or holiday.

Finally, consider Ontario COVID-19 wastewater signal (standardized concentration of SARS-CoV-2 gene copies) and cases [22].

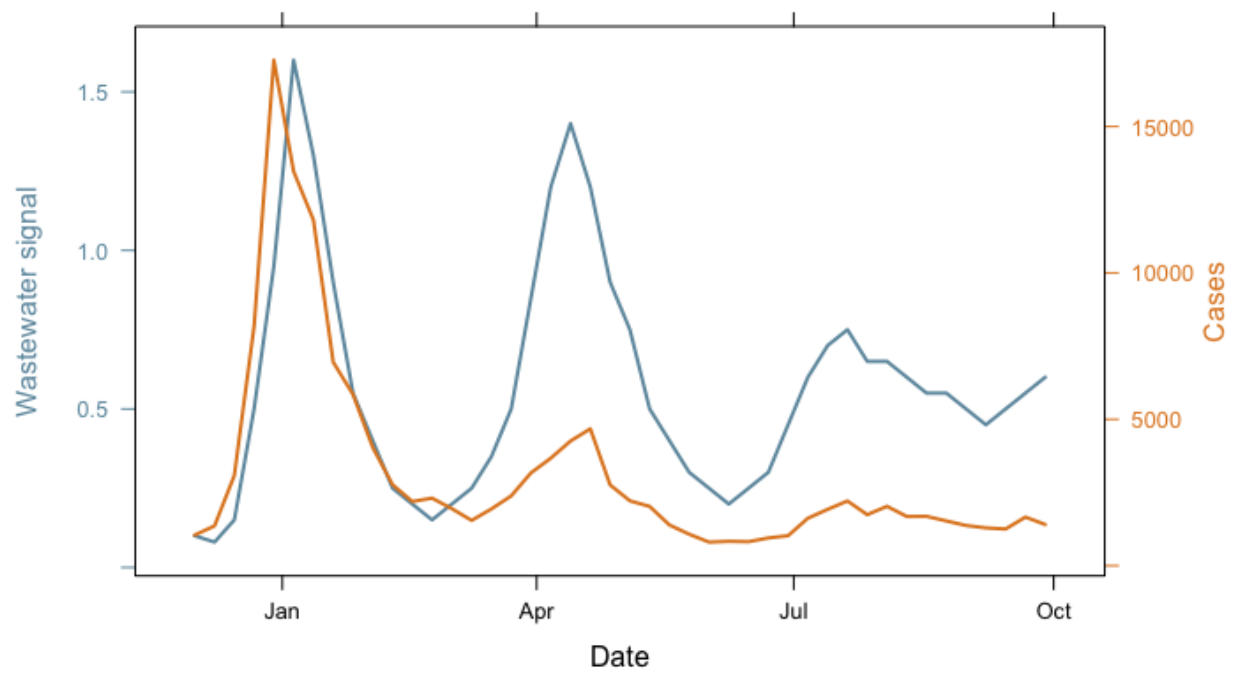


Figure 2.23: Standardized concentration of SARS-CoV-2 gene copies [22] and case count [21] plot.

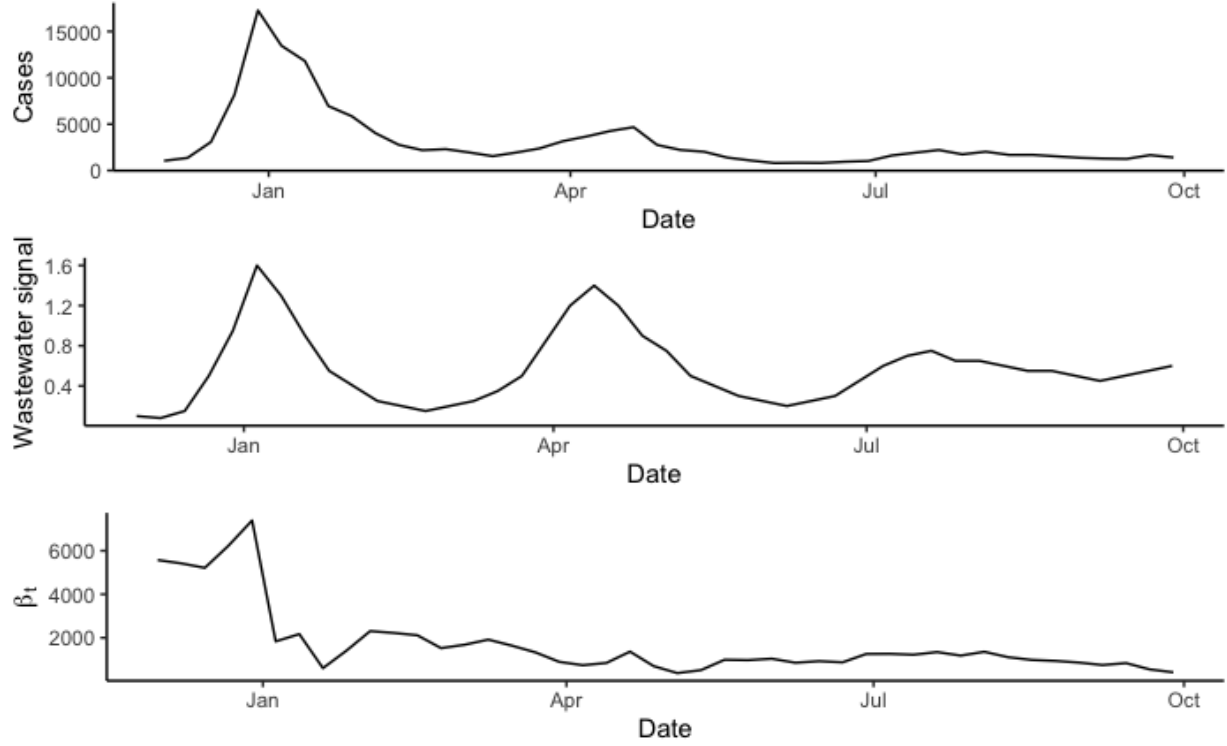


Figure 2.24: Smoothing results.

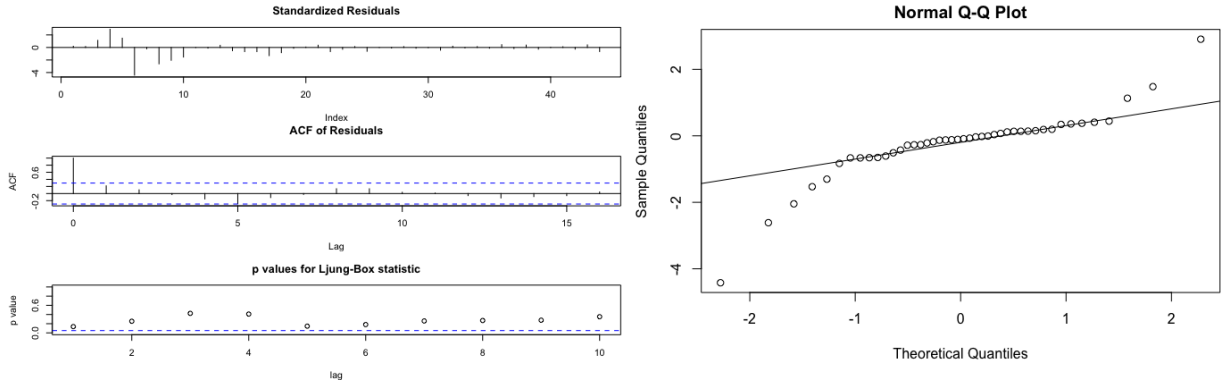


Figure 2.25: Model diagnostics.

Salient results are highlighted as follows. At the first peak in cases of 17266 (week of Dec. 28, 2021), a one standard deviation (0.36 unit) increase in wastewater signal measurement corresponds to a 7392 increase in cases. The model estimate agrees with the data quite well at the peak:

$$\text{cases}_{t_{\text{peak}}} - \text{cases}_{t_{\text{peak}}-1} \frac{\text{WW}_{t_{\text{peak}}} - \text{WW}_{t_{\text{peak}}-1}}{\sigma_{\text{ww}}} = 17266 - 8171 \frac{0.95 - 0.50}{0.3594325} = 7037.$$

At the second high point of 4258 to 4676 cases (weeks of Apr. 13 and 20, 2022), a one standard deviation increase in wastewater signal measurement corresponds to a 842 to 1363 increase in cases. The increase in cases due to a one standard deviation increase in wastewater signal is relatively stable for weeks 27 to 40 (June through August 2022) with a range of 852 to 1356 cases.

2.3.3 Application of the covariance estimation method to public health surveillance data

We now provide an application of the preceding multivariate covariance estimation algorithm. Data analysis focuses on COVID-19 surveillance targets, cases, hospitalization and deaths [21] and covariates pertaining to viral load, temperature and air quality [22].

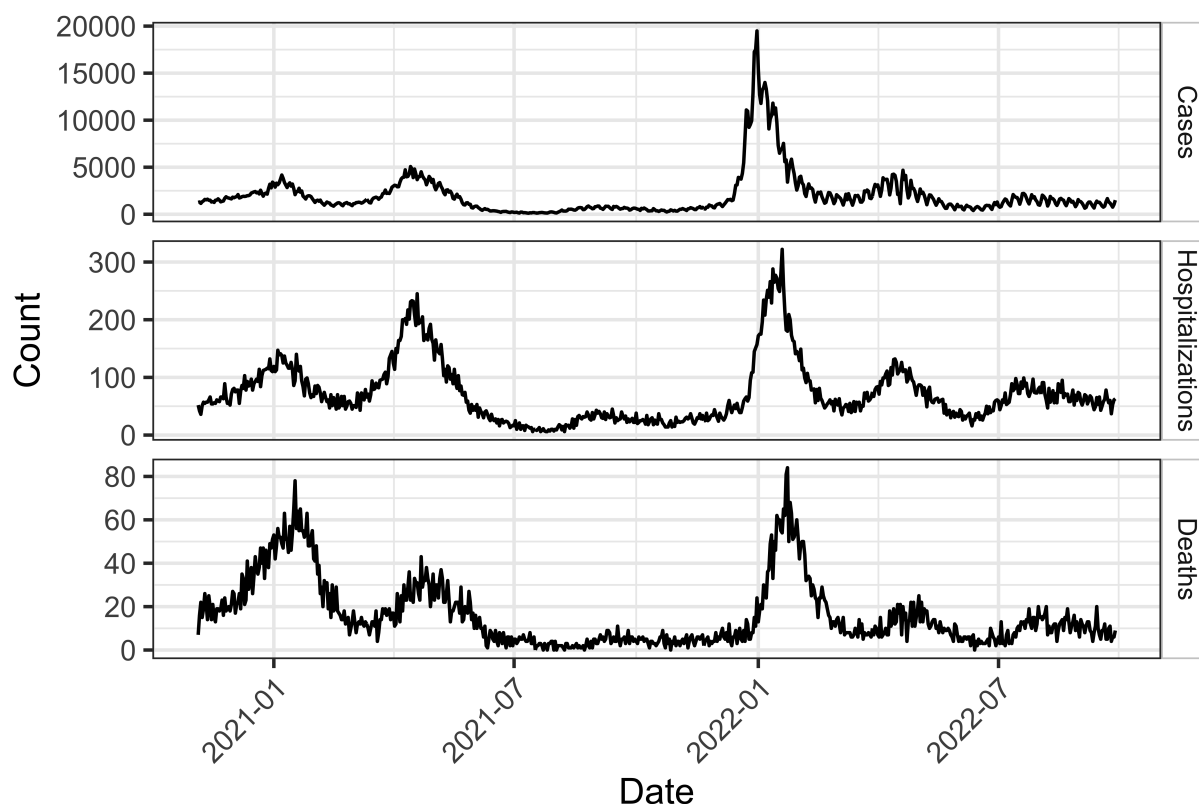


Figure 2.26: Cases, hospitalizations and deaths related to COVID-19.[21]

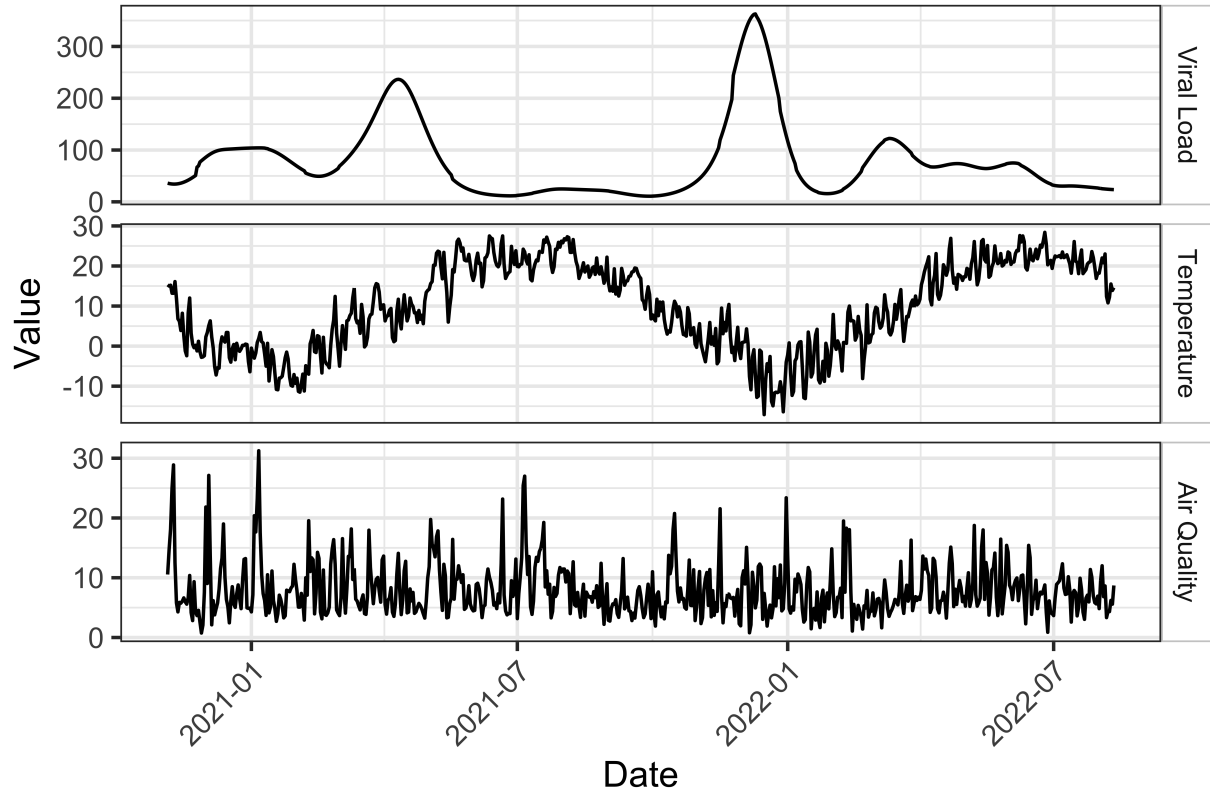


Figure 2.27: Viral load, temperature and air quality data.[\[22\]](#)

First, consider Model 1. This model considers only the surveillance targets. The prior initialization settings are as follows:

$$G_0 = V_0 = \theta_0 = I_3, \quad (2.3.9)$$

$$F_0 = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^\top, \quad (2.3.10)$$

$$C_0 = 1000I_3, \quad (2.3.11)$$

$$n_0 = 10, \quad (2.3.12)$$

$$S_0 = 0.1I_3, \quad (2.3.13)$$

$$\delta = 0.99, \quad (2.3.14)$$

$$W_t = \frac{1 - \delta}{\delta} G_t C_t G_t^\top. \quad (2.3.15)$$

Posterior modes of observation variances are depicted below.

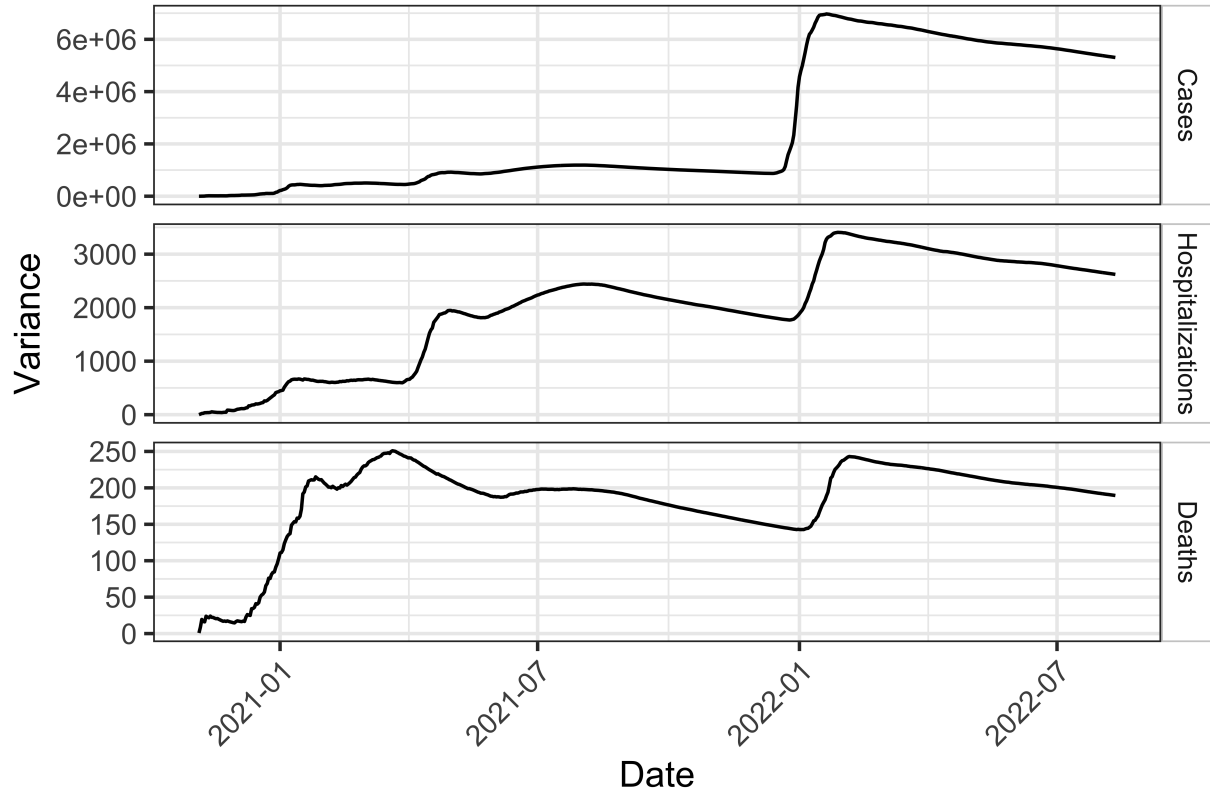


Figure 2.28: Estimated Model 1 time-varying variances.

Standardized residuals can be used to check overestimation or underestimations of the true observation variance. To do this, compare the residuals of each response series to credible limits ± 1.96 of the standard normal distribution: above 5% outliers suggests underestimation of the observation variance; below 5% suggests overestimation of the observation variance (expect to obtain 5% outliers for a perfect fit). This can be further explored by studying the mean squared error, mean squared standardized error, or mean absolute deviation.

Cross-correlations from the observation matrix are shown in the following plot.

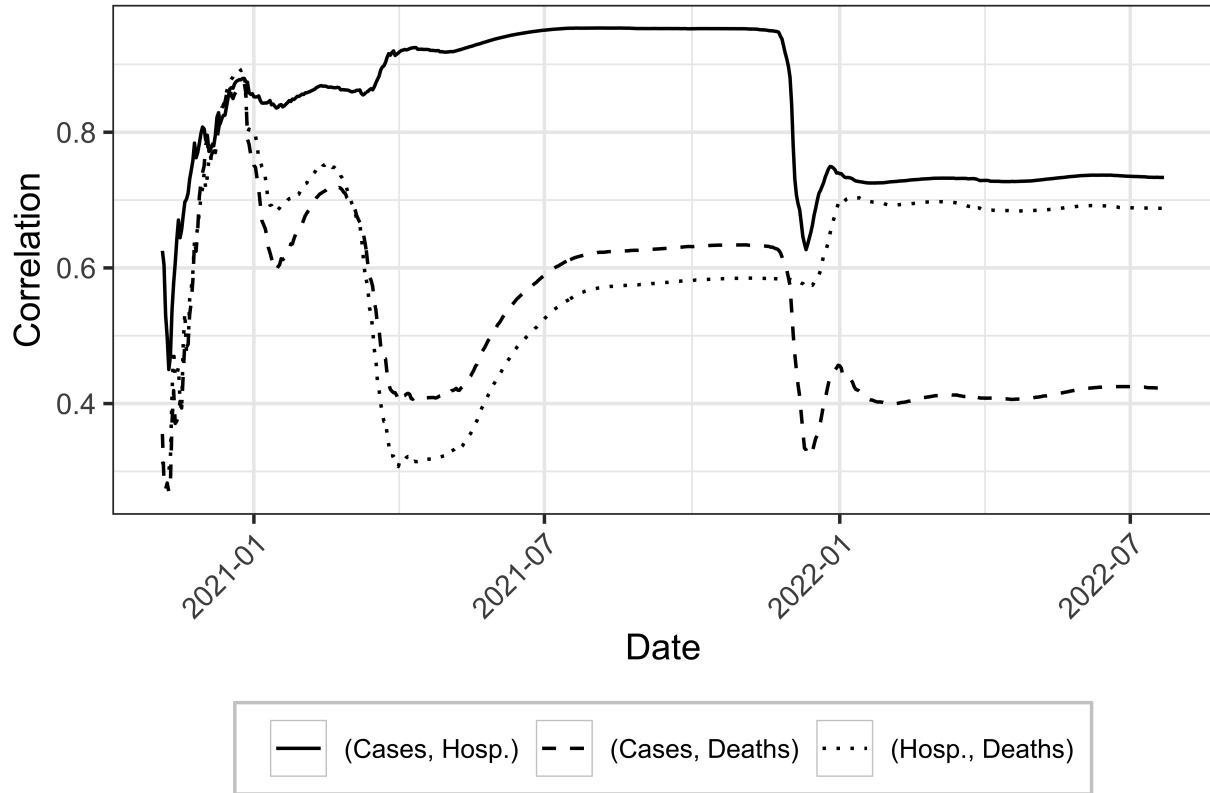


Figure 2.29: Estimated Model 1 time-varying cross-correlations.

As a baseline for comparison, consider the following static sample correlation matrix (%):

	Cases	Hospitalizations	Deaths
Cases	100	72.64	41.09
Hospitalizations	72.64	100	69.27
Deaths	41.09	69.27	100

Table 2.2: Static sample correlations.

To further summarize the model results, let the time period before the major epidemic peak be Period 1 and the time after the major epidemic peak be Period 2. Pairwise cross-correlations between cases (C), hospitalizations (H) and deaths (D) are shown in the following table.

	Period 1	Period 2
ρ_{CH}	95%	75%
ρ_{CD}	60%	40%
ρ_{HD}	55%	70%

Table 2.3: Model 1 summarized cross-correlations.

One might query as to why the correlations change after the peak or further or, why the ranking of the size of the correlations change after the peak.

Consider first the pre-peak period cross-correlations. Cases and hospitalizations (95%) which signify strong, almost linear relationship between the number of reported cases and hospitalizations, which is expected in the early stages of an epidemic when the healthcare system is not overwhelmed, and most cases that require hospitalization are likely to get it. Cases and deaths (60%) exhibit lower correlation due to lag between case diagnosis and death, which can cause discrepancies in the data alignment, and the fact that the fatality rate might not be proportional to the case count due to varying severity of the disease among different populations. Hospitalizations and Deaths (55%) show moderate correlation due to medical interventions that successfully treat many hospitalized patients, variations in the severity of cases, or differences in the time lag between hospitalization and death.

Now, the post-peak period cross-correlations between cases and hospitalizations (75%) show a change in the relationship between these two variables post-peak, possibly due to changes in healthcare policies (hospitalizing only severe cases), improvements in treatment reducing hospitalization rates, or a shift in the infected demographic (younger individuals who are less likely to require hospitalization).

The further reduced correlation indicates an even greater decoupling of case counts from death counts post-peak (40%). This could be due to improved treatments, increased testing leading to the identification of more mild/asymptomatic cases (less likely to result in death), a delay in occurrence of deaths following a reduction in cases. The increased correlation between hospitalizations and deaths (70%) is intriguing. This might suggest that post-peak, the severity of cases requiring hospitalization increased, delay in the impact of improved treatments on mortality rates, the healthcare system's focus shifted towards treating more severe cases more effectively, thus making hospitalization a more direct predictor of mortality.

Furthermore, additional considerations are as follows. Incorporating time lags into the analysis is crucial, as the effects of cases on hospitalizations and deaths are not instantaneous. The relationships between these variables are likely non-linear and might be better explained by models that account for such non-linear dynamics. Heterogeneity in data may arise due

to differences in testing rates, healthcare policy changes, and population demographics over time can significantly impact these correlations. Other unmeasured external factors (e.g., public health interventions, changes in public behaviour) might also have influenced these dynamics. In addition, especially in the post-peak period, the data might be subject to more statistical noise due to varying rates of reporting and changes in testing strategies. Finally, regarding data quality, the reliability and consistency of data collection over time can significantly affect correlation estimates.

The change in the ranking of the size of the correlations between case counts, hospitalizations, and deaths after the major epidemic peak can be attributed to several epidemiological, healthcare system, and behavioural factors. These include shifts in healthcare system dynamics, changes in testing and reporting, evolution of the disease and population immunity, public health interventions and behavioural changes, and demographic shifts in infected populations.

Shifts in healthcare system dynamics are also observed. Pre-peak and during the initial phase of an epidemic, healthcare systems are typically more reactive. The high correlation between cases and hospitalizations suggests that the healthcare system was closely tracking the epidemic's progression, hospitalizing a consistent proportion of cases. In contrast, post-peak, healthcare systems adapt and become more selective in hospitalizing patients, due to resource constraints or improved understanding of the disease leading to a lower correlation between cases and hospitalizations but a higher correlation between hospitalizations and deaths, as more severe cases are likely to be hospitalized.

Changes in testing and reporting also impact the results. Increased testing post-peak leads to the identification of more mild or asymptomatic cases, which would not require hospitalization or lead to death. This increased detection of less severe cases decreases the correlation between cases and hospitalizations/deaths. Furthermore, reporting delays in reporting deaths become more pronounced post-peak, due to overwhelmed systems, changing the correlation dynamics.

Evolution of the disease and population immunity impacts the results. Changes in virus characteristics may occur if the virus becomes less virulent post-peak, leading to fewer severe cases requiring hospitalization, and consequently fewer deaths. An increased level of population immunity (due to prior infection or vaccination) post-peak leads to fewer severe cases, altering the correlation structure.

Public health interventions and behavioural changes early in the epidemic occurs when there is less public awareness and fewer interventions, leading to a more straightforward progression from cases to hospitalizations and deaths. Post-peak, there are increased public health measures, improved treatment protocols, and changes in population behaviour (social

distancing and mask-wearing) which leads to a disconnect between the number of cases and the severity of those cases, thereby changing the correlation pattern.

Demographic shifts in infected populations account for some differences as different age groups are affected post-peak, the demographic profile of those infected shifts (e.g., younger individuals with lower risk of hospitalization and death), affecting the correlation structure.

Technically, statistical and methodological considerations such as sample variability/non-stationarity accounting for the variability in each of the data sets (cases, hospitalizations, deaths) changes over time, affecting correlations as well as increased model sensitivity to certain types of changes (absolute increases in case numbers) than others.

Now, we consider Model 2 where viral load [22] is included as a covariate in the dynamic regression.

$$y_t = R_{1t} + v_t = F_t \theta_t + v_t, \quad (2.3.16)$$

$$G_t = \begin{bmatrix} 1 \end{bmatrix}. \quad (2.3.17)$$

The inclusion of wastewater viral load as a covariate introduces an additional layer of complexity and aims to capture indirect pathways or environmental influences on disease spread and severity.

The innovations are $v_t \sim \mathcal{N}(0, V_t)$ and $w_t \sim \mathcal{N}(0, W_t, V_t)$ where V_t is a 3×3 observation covariance to be estimated and W_t is a 3×3 transition covariance matrix specified using a discount factor $\delta = 0.99$.

The prior initialization is given by distributions,

$$\theta_0 \sim \mathcal{N}(m_0 = 0, C_0 = 1000I_3, V_0), \quad (2.3.18)$$

$$V_0 \sim IW(n_0 = 10, n_0 S_0 = I_3). \quad (2.3.19)$$

Posterior modes of observation variances for Model 2 are depicted below.

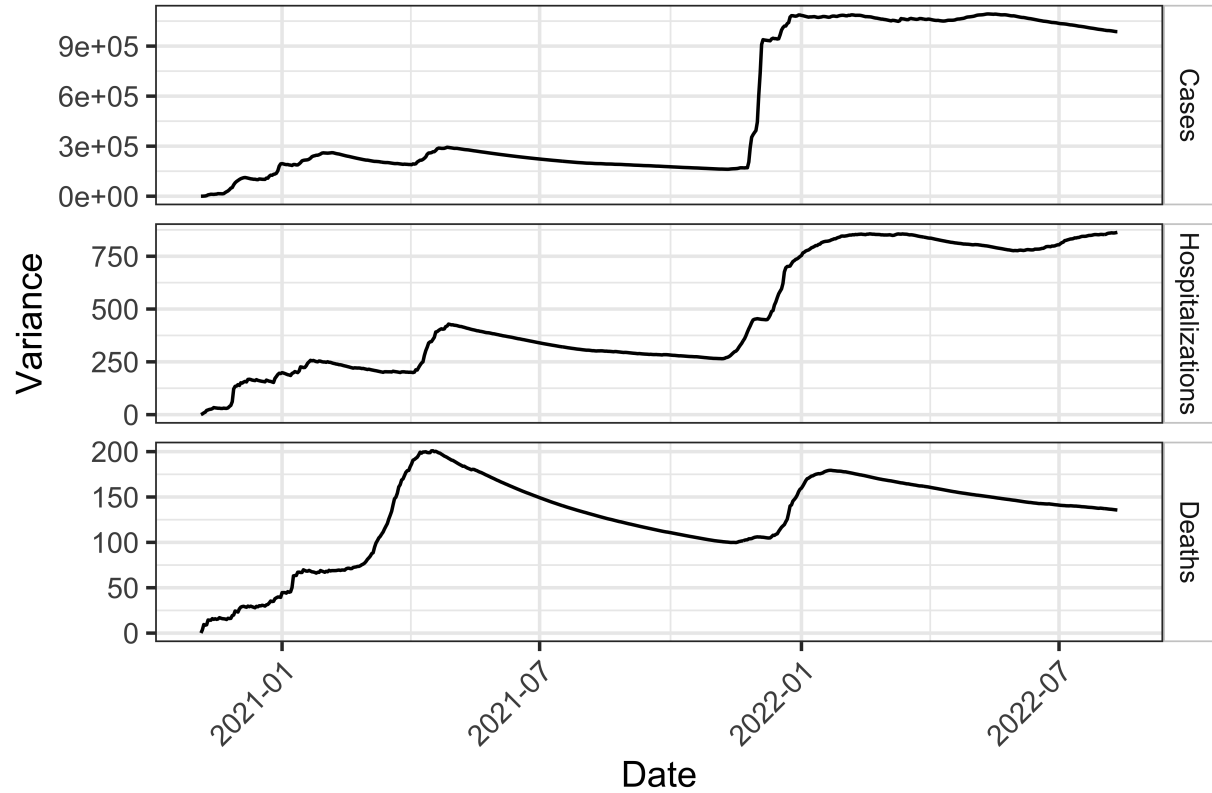


Figure 2.30: Estimated Model 2 time-varying variances.

Cross-correlations from the observation matrix for Model 2 are shown in the following plot:

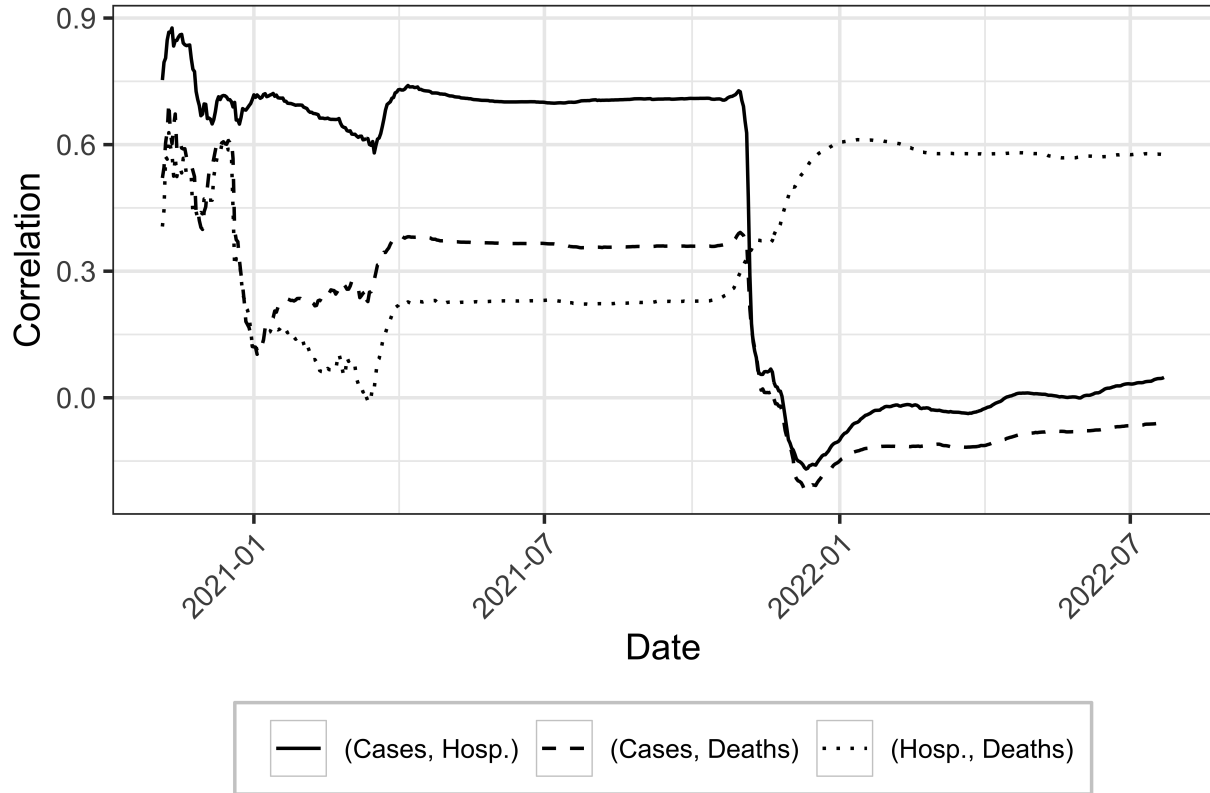


Figure 2.31: Estimated Model 2 time-varying cross-correlations.

Below is a summary of the results for the model with no covariate (Model 1) and the model with covariate included (Model 2) in both the pre-peak (Period 1) and post-peak (Period 2) periods.

	Model 1		Model 2	
	Period 1	Period 2	Period 1	Period 2
ρ_{CH}	95%	75%	70%	0%
ρ_{CD}	60%	40%	35%	-10%
ρ_{HD}	55%	70%	25%	60%

Table 2.4: Comparison of Models 1 and 2 by period.

The inclusion of wastewater viral load in Model 2 reduces the correlation between cases and hospitalizations in both periods, particularly eliminating it in the post-peak period. This suggests that environmental viral load is a strong determinant of hospitalization rates independent of the number of cases reported. The presence of the covariate not only reduces

the positive correlation seen in Model 1 but also introduces an inverse correlation in the post-peak period in Model 2. This indicates that factors captured by the wastewater viral load, such as unreported or asymptomatic cases, inversely influence the relationship between reported cases and deaths. The reduction in correlation coefficients in Model 2 suggests that viral load has a moderating effect on the relationship between hospitalizations and deaths, due to varying levels of viral load influencing disease severity or detection rates.

We now discuss differences in Model 1 and Model 2 cross-correlations across periods.

For the differences in ρ_{CH} between the models, note the following. Period 1 drop (25-30%): In the early stages of an outbreak (Period 1), the healthcare system might not be overwhelmed, and the relationship between cases and hospitalizations are more direct and less influenced by external factors. Period 2 drop (75%): As the outbreak progresses (Period 2), environmental viral loads become more indicative of community spread, overshadowing the direct relationship observed in the baseline model. This may occur due to factors such as increased asymptomatic cases, changes in testing practices, or healthcare system saturation, leading to a sharper drop in correlation.

Differences in ρ_{CD} between the models are described as follows. Period 1 drop (25-30%): Similar to ρ_{CH} , the initial stages of the outbreak might exhibit a more straightforward relationship between cases and deaths. Period 2 drop (50%): The reduction becomes more pronounced in Period 2, possibly due to improved medical treatments, changes in the demographics of affected individuals, or other public health interventions becoming more effective. The environmental viral load might reflect these changes more acutely than the direct case counts.

For the differences in ρ_{HD} between the models, note the following. Period 1 drop (25-30%): This correlation may be more resilient to changes due to the covariate, since hospitalizations and deaths are both late-stage outcomes of the disease. Period 2 drop (10%): The relatively small drop suggests that, even with high environmental viral loads, the relationship between hospitalizations and deaths remains more direct and less influenced by external factors, possibly due to the severity of cases leading to hospitalization generally following through to mortality.

Implication of the differences in results can be viewed from different angles. From an epidemiological viewpoint, the variations in correlation drops reflect the complex interplay between disease dynamics and environmental factors. Environmental viral load as a proxy for community spread may capture nuances in disease transmission and severity not reflected in direct health metrics alone. Mathematically, the varying degrees of correlation reduction highlight the nonlinear and interaction effects in disease modelling. The influence of the covariate is not uniform across all relationships and periods, suggesting a more complex

model than linear assumptions can capture.

Furthermore, correlation changes do not imply causation. The observed changes in correlations are indicative of associations that may require further investigation to understand causal mechanisms.

The analysis highlights environmental sensitivity, whereby differences in correlation coefficients with the inclusion of wastewater viral load as a covariate highlight the sensitivity of epidemiological data to environmental factors. Wastewater data can act as an early indicator of community spread by preceding clinical cases, thus significantly influencing the dynamics observed in health metrics.

The influence of wastewater signal is apparent from multiple aspects. Unreported cases and asymptomatic spread impacts the results as the negative correlation in ρ_{CD} during the post-peak period in Model 2 points towards the impact of unreported or asymptomatic cases that are likely captured by wastewater analysis but not in clinical case reports.

Furthermore, there are policy implications. Reliance solely on clinical data without considering environmental surveillance might lead to incomplete or misleading interpretations of the disease dynamics. Policies based on such incomplete models might fail to address hidden spread and early signs of outbreaks.

modelling robustness should be noted, as changes observed with the inclusion of the covariate underscore the importance of considering multifaceted data sources in epidemiological models. This improves the robustness of the models, making them more reliable for public health decision-making.

Finally, consider Model 3 which captures many different features of the data using regression, trend, and seasonality components. This model considers wastewater viral load signal, temperature, and air quality as a covariates in the dynamic regression on the three surveillance targets. In addition, trend-seasonal components, for the seasonal variation of which, a reduced form state space representation of the full Fourier expansion is used. For computational efficiency, the first 5 harmonics out of a total $182 = \frac{364}{2}$ are used.

For the model components, let θ_t be a time-varying 15×3 unobserved state matrix which

drives dynamics of y_t as follows:

$$y_t = R_{1t} + R_{2t} + R_{3t} + T_t + s_t + v_t \quad (2.3.20)$$

$$= F_t \theta_t + v_t, \quad (2.3.21)$$

$$R_{it} = F_{it} \theta_{it}, \quad i = 1, 2, 3 \quad (2.3.22)$$

$$T_t = [1, 0]^\top \theta_{4t}, \quad (2.3.23)$$

$$s_t = [1, 0, 1, 0, 1, 0, 1, 0, 1, 0]^\top \theta_{5t} \quad (2.3.24)$$

$$\theta_t = [\theta_{1t}, \theta_{2t}, \theta_{3t}, \theta_{4t}, \theta_{5t}], \quad (2.3.25)$$

$$F_t = [F_{1t}, F_{2t}, F_{3t}, 1, 0, 1, 0, 1, 0, 1, 0]^\top, \quad (2.3.26)$$

$$G_t = \begin{bmatrix} I_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & J & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & G_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & G_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & G_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & G_4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & G_5 \end{bmatrix}, \quad (2.3.27)$$

$$J = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad (2.3.28)$$

$$G_i = \begin{bmatrix} \cos(\frac{2\pi i}{365}) & \sin(\frac{2\pi i}{365}) \\ -\sin(\frac{2\pi i}{365}) & \cos(\frac{2\pi i}{365}) \end{bmatrix}, \quad i = 1, 2, 3, 4, 5. \quad (2.3.29)$$

For the prior initializations, the innovations are $v_t \sim \mathcal{N}(0, V_t)$ and $w_t \sim \mathcal{N}(0, W_t, V_t)$ where V_t is a 3×3 observation covariance matrix to be estimated and W_t is a 15×15 transition covariance matrix specified using a discount factor $\delta = 0.99$. The prior distributions are

$$\theta_0 \sim \mathcal{N}(m_0 = 0, C_0 = 1000I_3, V_t) \quad (2.3.30)$$

$$V_t \sim IW(n_0, n_0 S_0 = 10I_3). \quad (2.3.31)$$

Posterior modes of observation variances for Model 3 are depicted below.

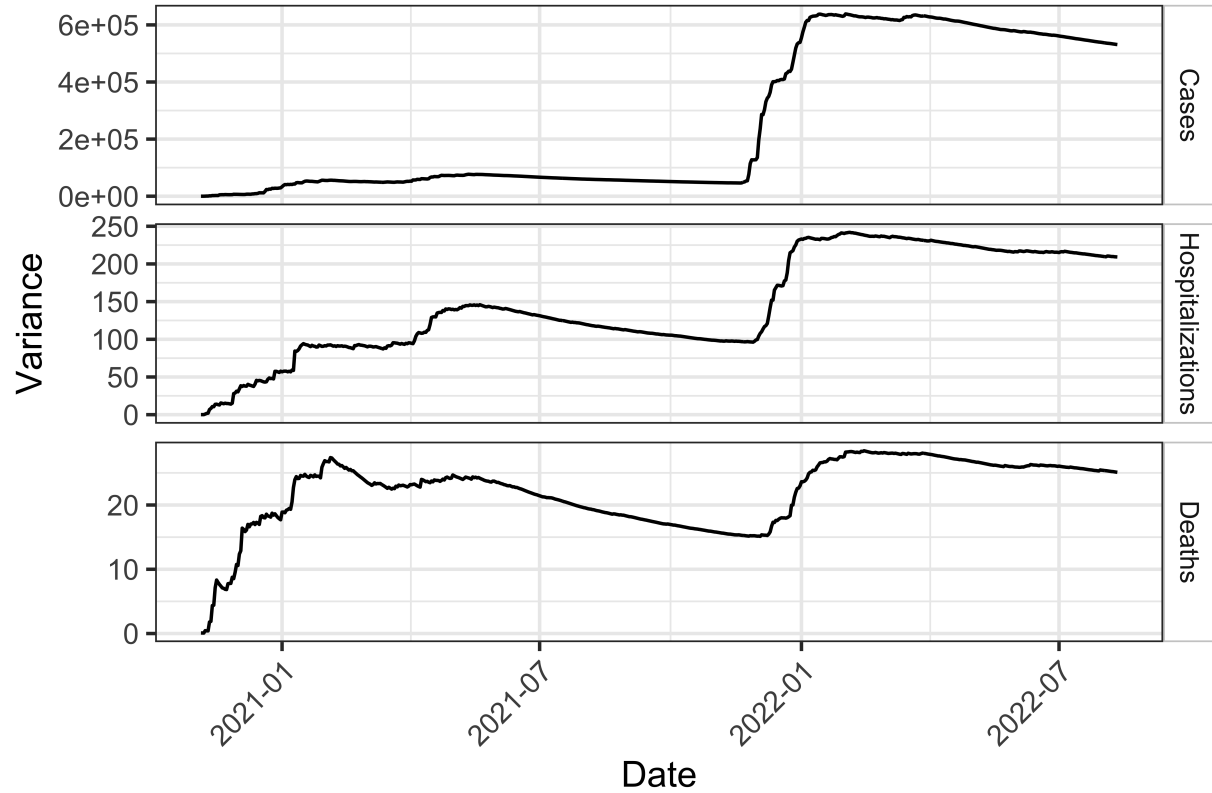


Figure 2.32: Estimated Model 3 time-varying model variances.

Cross-correlations from the observation matrix for Model 3 are depicted in the following plot.

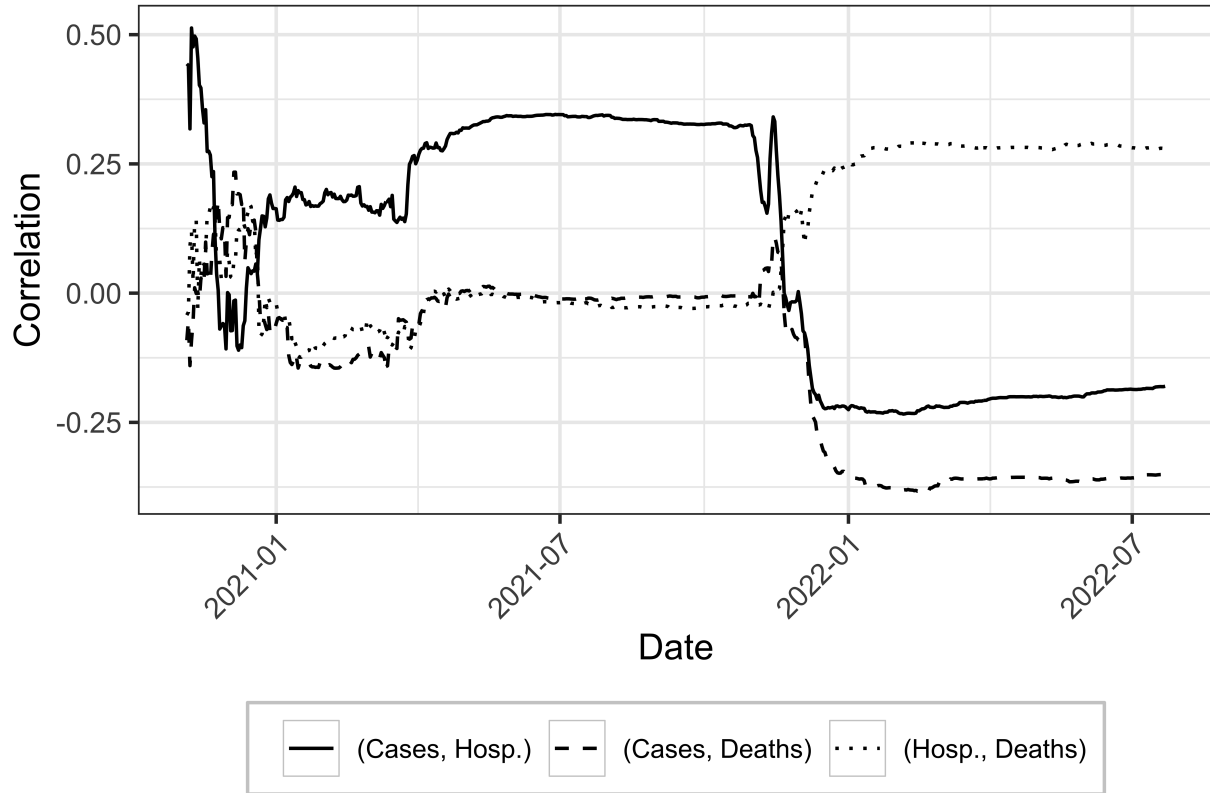


Figure 2.33: Estimated Model 3 time-varying cross-correlations.

Below is a summary of the results for the model with no covariate (Model 1) and the models with covariates included (Model 2 and Model 3) in both the pre-peak (Period 1) and post-peak (Period 2) periods.

	Model 1		Model 2		Model 3	
	Period 1	Period 2	Period 1	Period 2	Period 1	Period 2
ρ_{CH}	95%	75%	70%	0%	35%	-20%
ρ_{CD}	60%	40%	35%	-10%	0%	-35%
ρ_{HD}	55%	70%	25%	60%	0%	30%

Table 2.5: Comparison between the models by period.

2.4 Discussion

Interpretation of findings in case studies

Estimation of parameter β_t for COVID-19 hospitalization counts provides information on when the number of hospitalizations changes when compared to the average case count. $\beta_t = 1$ indicates the change in hospitalizations is the same as the average case count; $\beta_t < 1$ indicates the change in hospitalization counts is more conservative than the average case counts; $\beta_t > 1$ indicates the change in hospitalization counts is more aggressive than the average case counts which contributes to ‘system strain’. Over the analysis period, this occurs during April 19, 2021 and January 19, 2022 as depicted by the time-varying regression coefficient plot.

The analysis reveals that COVID-19 cases are lower on non-workdays as fewer patients get tested on the weekends due to fewer hours and sites available and doctor’s offices open, and labs process fewer tests on weekends and holidays. Health departments experience reporting lags and backlogs on non-workdays and weekend and holiday gatherings lead to more initial symptoms on weekdays. There exists a ‘widening gap’ phenomenon that arises: when compared to workdays, the percent decrease in cases associated with a non-workday has increased over time.

Remarks on the dynamic correlation analysis the results are as follows. Both static and dynamic approaches suggest that pairwise correlations are positive. However, there are differences in outcomes. The static correlation estimates are based on the sample correlations which fail to take into account the dynamics of the data and in particular the validity of sample statistics is based both of the following invalid assumptions: (i) the availability of large amounts of data, here 692 observations, and, (ii) the data is stationary, in disagreement with the apparent non-stationarity of the time series data. Thus, one may not rely only on sample statistics in order to estimate covariance and correlation matrices in this case. In addition, the inferred correlations of the dynamic algorithm offer the significant advantage of on-line correlation estimates. The results somehow agree with the sample cross-correlations, but they are more valid as they explicitly make use of the dynamic state-space model.

Interpretation of results include consideration of the magnitude of changes in cross-correlation coefficients from Model 1 to Model 3 which reveals the influence of including covariates. The more covariates added, the more the original correlations are reduced or even inverted. The direction of change implies a shift from positive to negative correlations in some cases (especially in Model 3 for Period 2) which is particularly noteworthy, suggesting that the factors represented by the covariates have an impact on the relationships

between the health metrics.

Model complexity and real-world representation of the problem shows that the increasing complexity from Model 1 to Model 3 seems to reveal more about the underlying dynamics of the disease and its interaction with environmental factors. It suggests that simple models (like Model 1) may not fully capture the nuanced relationships in real-world scenarios, especially in the presence of environmental influences.

Policy implications are practically impactful as these quantitative changes indicate that epidemiological modelling and decision-making should consider a broad range of factors, including environmental ones, to accurately understand and respond to health crises.

From an epidemiological standpoint, the introduction of environmental factors capture indirect pathways of disease spread and highlight aspects such as asymptomatic transmission. Biological and environmental mechanisms explain these changes, as high viral loads in wastewater precede clinical cases, thus influencing the correlations. The results highlight the importance of considering a range of factors in epidemiological models for more accurate predictions and public health strategies.

Limitations and Challenges

Despite its strengths, the recursive Bayesian state estimation approach requires significant computational resources, especially for high-dimensional systems. Another limitation is the need for precise initial parameter settings, which can affect convergence and accuracy.

The primary challenge encountered was the handling of extremely noisy data, which occasionally led to suboptimal performance. Additionally, the possibility of integration of this methodology with real-time systems poses practical design and implementation issues.

To address these challenges, future research could focus on developing more efficient algorithms that reduce computational overhead. Improving the robustness of the model to handle higher levels of noise and exploring adaptive parameter estimation techniques are also potential areas for further investigation.

2.5 Conclusion

This article demonstrates the utility of state-space modelling, recursive Bayesian state estimation, dynamic regression and correlation analysis for public health data that significantly enhances descriptive and predictive analysis of dynamic systems. We noted that the methodology’s adaptability to non-stationary and noisy environments. We emphasized that the case study clearly demonstrates the efficacy of using component linear Gaussian state-space

models and dynamic regression models in analyzing COVID-19 case surveillance data. The results demonstrate the models' ability to capture key epidemiological trends, such as the relationship between case counts and hospitalization, workdays, and wastewater signals. The model predictions, as assessed by log-likelihood, mean absolute percent error, and Theil's U, provide a robust framework for understanding the dynamics of the pandemic in Ontario, Canada, and can be instrumental in guiding public health interventions.

Future work could explore ways to optimize the algorithm for real-time or streaming data applications, focusing on reducing computational complexity. This may also be extended by investigating advanced noise reduction techniques and adaptive filtering methods to improve the model's robustness against high noise levels. Lastly, developing practical implementations and testing the integration of this methodology with real-time systems in various fields such as environmental monitoring or urban planning may be of benefit.

The recursive Bayesian state estimation approach represents a significant advancement in the application of state-space modelling. Its ability to adapt to changing conditions and maintain high accuracy makes it a valuable tool for researchers and practitioners alike. The insights gained from this research pave the way for further innovations in more complex systems and applications, promising substantial improvements in areas requiring precise state estimation under uncertainty.

Chapter 3

State-Space Modelling for Infectious Disease Surveillance Data: Stochastic Simulation Techniques and Structural Change Detection

3.1 Introduction

State-space modelling has become indispensable in the analysis of dynamic systems across various fields, including economics, engineering, and public health [6], [11], [29]. These models provide a framework for understanding how unobservable states evolve over time based on observable data, making them particularly useful for real-time monitoring and forecasting [5], [30], [31].

In our previous work, we applied state-space models to COVID-19 surveillance data from the province of Ontario, Canada, implementing dynamic regression models and novel covariance estimation methodologies to investigate the relationships between COVID-19 case counts, hospitalizations, and environmental factors such as wastewater signal measurements [21]. This integration of covariates into state-space models provided a nuanced understanding of the factors driving COVID-19 transmission and public health outcomes [32], [33].

Building on that foundation, the current article extends our exploration by focusing on the application of stochastic simulation methods, including Markov Chain Monte Carlo (MCMC) and particle filters, within the state-space modelling framework. These methods address the challenges of real-time updating and sequential estimation, which are crucial for timely decision-making in public health [34]–[36]. We apply a Kalman smoothing approach

with MCMC for state sampling in a local level and seasonal model to further enhance the accuracy and robustness of our models.

This work also investigates Bayesian inference for nonlinear dynamic regression models, focusing on multivariate Bayesian dynamic models and their applications to public health data [1], [21], [37]. By exploring these advanced techniques, we aim to provide deeper insights into the temporal dynamics of COVID-19 transmission, building on the foundations laid in our previous analysis.

Overall, this article demonstrates the efficacy of these advanced state-space modelling techniques, offering valuable insights into the complex relationship between different epidemiological factors and improving the robustness of predictions in public health modelling [17], [38].

3.2 Methodology

3.2.1 Markov chain Monte Carlo method

For the linear Gaussian state-space model, Wiener and Kalman filters provide an analytic solution wherein the number of observations are fixed and evolving, respectively. However, in the general state-space model, the derivation of an analytic solution proves to be elusive for all but a handful of the simplest cases. Historically, analytic derivation was possible through the method of conjugate prior distribution, which belongs to the same family as the final posterior distribution. More recently, computational methods have been employed to approximate numerical solutions of the targeted filtering, predictive and marginal smoothing distribution. Some simple methods that work at low state dimension include the deterministic grid filter and the stochastic importance sampling which draws a sample from a surrogate proposal distribution from which statistics are calculated by taking into account the likelihood of the target distribution. In essence, the proposal distribution is approximated by the linear Gaussian state-space model via the Laplace approximation and statistics are calculated with the aid of sampling from the proposal distribution through Kalman filtering and smoothing. Such calculations are made more amenable to analysis by assuming that the observations follow an exponential family distribution, provided that the states are known. Further, the random number based Markov chain Monte Carlo (MCMC) method [34] has exploded in widespread applicability owing to the advances in computational resource available. The main idea in MCMC is to explore high-density regions of the target distribution by employing the properties of a Markov chain. From this process, a random number is generated from the target distribution in the general state-space model. The MCMC approach works well

for the backward-looking smoothing operations when the number of observation is fixed, however suffers from the need to restart for sequential data when obtaining the filtering and predictive distribution.

For a linear Gaussian state-space model with unknown parameters, the target distribution for estimation is the joint posterior distribution $p(\theta_{0:T}, \psi \mid y_{1:T})$.

$$\begin{aligned}
p(\theta_{0:T}, \psi \mid y_{1:T}) &= \frac{p(\theta_{0:T}, \psi, y_{1:T})}{p(y_{1:T})} \\
&\propto p(\theta_{0:T}, \psi, y_{1:T}) \\
&= p(y_{1:T} \mid \theta_{0:T}, \psi) \cdot p(\theta_{0:T}, \psi) \\
&= p(\theta_0 \mid \psi) \cdot p(\psi) \cdot \prod_{t=1}^T p(y_t \mid \theta_t, \psi) \cdot p(\theta_t \mid \theta_{t-1}, \psi) \\
&= \prod_{t=1}^T p(y_t \mid \theta_t, \psi) \cdot \prod_{t=1}^T p(\theta_t \mid \theta_{t-1}, \psi) \cdot p(\theta_0) \cdot p(\psi). \tag{3.2.1}
\end{aligned}$$

MCMC can be used to generate samples from a distribution of interest, $p(\psi)$. There are several approaches, however the most basic is the Metropolis method as follows.

Algorithm 7 Metropolis method

Input. Initialize ψ^0 . Let N be the number of iterations in the calculation of the MC.

For $j = 1, \dots, N$:

Set $\psi = \psi^{j-1}$: Draw proposal $\tilde{\psi}$ from $q(\cdot \mid \psi)$

Calculate acceptance ratio $\alpha = \frac{p(\tilde{\psi})}{p(\psi)}$

If $\alpha \geq 1$ then accept proposal $\tilde{\psi}$ by setting $\psi^j = \tilde{\psi}$

If $\alpha < 1$ then accept proposal $\tilde{\psi}$ with probability equal to α :

by setting $\psi^j = \tilde{\psi}$ (acceptance), or,

by setting $\psi^j = \psi^{j-1}$ (rejection)

Output. $\{\psi^j\}_{j=1}^N$ a sample from $p(\psi)$

The proposal distribution $q(\cdot \mid \psi)$ is used to obtain the new proposal value $\tilde{\psi}$ given the present value ψ . Commonly, a normal distribution is selected with mean ψ . Sampling from $q(\cdot \mid \psi)$ follows random walk Markov chain. The method is randomized since it is based on the acceptance ratio, α which implies that it is a Monte Carlo method.

The Metropolis method that assumes a symmetric proposal distribution such that $q(\tilde{\psi} \mid \psi) = q(\psi \mid \tilde{\psi})$ as in a random walk, is known as Metropolis-Hastings method. The main difference is that the Metropolis-Hasting method includes appropriate factors in the acceptance ratio to account for an asymmetric proposal distribution.

Algorithm 8 Metropolis-Hastings method

Input. Initialize ψ^0
For $j = 1, \dots, N$:
Set $\psi = \psi^{j-1}$: Draw proposal $\tilde{\psi}$ from $q(\cdot | \psi)$
Calculate acceptance ratio $\alpha = \frac{p(\tilde{\psi})q(\psi|\tilde{\psi})}{p(\psi)q(\tilde{\psi}|\psi)}$
If $\alpha \geq 1$ then accept proposal $\tilde{\psi}$ by setting $\psi^j = \tilde{\psi}$
If $\alpha < 1$ then accept proposal $\tilde{\psi}$ with probability proportional to α :
by setting $\psi^j = \tilde{\psi}$ (acceptance), or,
by setting $\psi^j = \psi^{j-1}$ (rejection)
Output. $\{\psi^j\}_{j=1}^N$ a sample from $p(\psi)$

A special type of Metropolis-Hastings method is the Gibbs method whereby the proposal distribution is a full conditional distribution, which means that every variable is given except for one: $q(\cdot | \psi) = p(\cdot | \psi_{-k})$. At each iteration, one element $\tilde{\psi}_k$ is updated, and all others are kept fixed. The Gibbs method has acceptance ratio

$$\begin{aligned}\alpha &= \frac{p(\tilde{\psi}) \cdot q(\psi | \tilde{\psi})}{p(\psi) \cdot q(\tilde{\psi} | \psi)} \\ &= \frac{p(\tilde{\psi}) \cdot p(\psi_k | \tilde{\psi}_{-k})}{p(\psi) \cdot p(\tilde{\psi}_k | \psi_{-k})} \\ &= \frac{p(\tilde{\psi}_k, \tilde{\psi}_{-k}) \cdot p(\psi_k | \tilde{\psi}_{-k})}{p(\psi_k, \tilde{\psi}_{-k}) \cdot p(\tilde{\psi}_k | \psi_{-k})} \\ &= \frac{p(\tilde{\psi}_k | \tilde{\psi}_{-k}) \cdot p(\tilde{\psi}_{-k}) \cdot p(\psi_k | \tilde{\psi}_{-k})}{p(\psi_k | \tilde{\psi}_{-k}) \cdot p(\psi_{-k}) \cdot p(\tilde{\psi}_k | \psi_{-k})} \\ &= 1.\end{aligned}$$

which shows that every proposal is accepted with probability one.

Some checks when performing MCMC are in order. Since the outputs are random numbers, checking the convergence of the Markov chain is imperative. The most common way to do this is visually via a trace plot, where the number of search steps is plotted against sample realizations. This is done for a few Markov chains, typically three or four. Good convergence occurs when the trace plot shows high intensity and corresponds to adequate exploration of the distributional sample space. This is sometimes known as adequate mixing property. Poor convergence occurs when the Markov chain does not converge and there is minimal fluctuation in the overall pattern of the plot. This corresponds to inadequate searching of the target distribution and less mixing.

When performing smoothing using MCMC, the distribution of interest is the joint pos-

terior $p(\theta_{0:T}, \psi \mid y_{1:T})$. The Gibbs method is easily adapted to draw samples from this distribution by repeatedly sampling from the full conditional distribution as follows.

Algorithm 9 Gibbs method

Input. Initialize ψ^0
For $j = 1, \dots, N$:
Draw $\theta_{0:T}^j$ from $p(\theta_{0:T} \mid \psi = \psi^{j-1}, y_{1:T})$
Draw ψ^j from $p(\psi \mid \theta_{0:T} = \theta_{0:T}^j, y_{1:T})$
Output. $\{\theta_{0:T}^j, \psi^j\}_{j=1}^N$ a sample from $p(\theta_{0:T}, \psi \mid y_{1:T})$

Sampling can be improved by performing Kalman filtering forward in time followed by sampling via Kalman smoothing backward in time. This is known as forward filtering backward sampling (FFBS) [39] or simulation smoothing, as it is a simulation-based variant of Kalman smoothing.

Algorithm 10 Forward Filtering Backward Sampling

Run Kalman filtering with parameter ψ^{j-1}
Draw θ_T from $\mathcal{N}(m_T, C_T)$
For $t = T - 1, \dots, 0$:
Smoothing gain
 $A_t \leftarrow C_t G_{t+1}^\top R_{t+1}^{-1}$
Draw θ_t from $\mathcal{N}(h_t, H_t)$
where $h_t \leftarrow m_t + A_t[\theta_{t+1} - a_{t+1}]$ and
 $H_t \leftarrow C_t + A_t[0 - R_{t+1}]A_t^\top$
Output. $\{\theta_t\}_{t=T}^0$ a sample $\theta_{0:T}^j$ from $p(\theta_{0:T} \mid \psi = \psi^{j-1}, y_{1:T})$

To implement MCMC, a variety of software libraries are available. Some popular choices in the mathematical sciences include WinBUGS and JAGS which rely on the Gibbs method, and Stan, a probabilistic programming language which implements a hybrid Hamiltonian Monte Carlo method for enhanced sampling efficiency and flexibility.

As an illustration, consider the problem of smoothing in the linear Gaussian state-space model. Assuming a local-level model and, for the moment, known parameters, the smoothing operation can be performed using MCMC by drawing samples from the joint posterior distribution. The posterior distribution is specified in terms of the likelihood and prior

distributions using Bayes theorem:

$$\begin{aligned}
p(\theta_{0:T} \mid y_{1:T}) &= \frac{p(\theta_{0:T}, y_{1:T})}{p(y_{1:T})} \\
&\propto p(\theta_{0:T}, y_{1:T}) \\
&= p(y_{1:T} \mid \theta_{0:T}) \cdot p(\theta_{0:T}) \\
&= p(\theta_0) \cdot \prod_{t=1}^T p(y_t \mid \theta_t) \cdot p(\theta_t \mid \theta_{t-1}) \\
&= \prod_{t=1}^T p(y_t \mid \theta_t) \cdot \prod_{t=1}^T p(\theta_t \mid \theta_{t-1}) \cdot p(\theta_0). \tag{3.2.2}
\end{aligned}$$

The likelihood $p(y_{1:T} \mid \theta_{0:T})$ relates to the cumulative product of the observation equation as, $\prod_{t=1}^T p(y_t \mid \theta_t)$ and the prior distribution term $p(\theta_{0:T})$ relates to the cumulative product of the state equation times the prior distribution as $\prod_{t=1}^T p(\theta_t \mid \theta_{t-1}) \cdot p(\theta_0)$. This specification follows from the probability distribution formulation of the observation and state equations along with the prior distribution of the state.

Convergence of the Markov chain may be assessed using the effective sample size and the potential scale reduction factor, $\hat{R}$, which is related to the sample variance within Markov chain to between Markov chain, with a value less than one being desirable and indicative of adequate convergence of the Markov chain. Specifically, $\hat{R} = \sqrt{1 + \frac{B/W-1}{n_i}}$ where W and B are within and between Markov chain sequence variance and n_i is the sample size of one Markov chain sequence.

Now suppose a slightly different problem is encountered, whereby a general state space model is regarded as linear Gaussian with unknown stochastic parameters. The model must now be modified to consider the state along with the unknown parameters as random variables. The target distribution for estimation is the joint posterior distribution $p(\theta_{0:T}, \psi \mid y_{1:T})$. Further, suppose the variance of the state and observation noise are unknown and are to be jointly estimated. In specifying the model, the posterior is analogously manipulated using Bayes theorem:

$$\begin{aligned}
p(\theta_{0:T}, \psi \mid y_{1:T}) &= \frac{p(\theta_{0:T}, \psi, y_{1:T})}{p(y_{1:T})} \\
&\propto p(\theta_{0:T}, \psi, y_{1:T}) \\
&= p(y_{1:T} \mid \theta_{0:T}, \psi) \cdot p(\theta_{0:T}, \psi) \\
&= p(\theta_0 \mid \psi) \cdot p(\psi) \cdot \prod_{t=1}^T p(y_t \mid \theta_t, \psi) \cdot p(\theta_t \mid \theta_{t-1}, \psi) \\
&= \prod_{t=1}^T p(y_t \mid \theta_t, \psi) \cdot \prod_{t=1}^T p(\theta_t \mid \theta_{t-1}, \psi) \cdot p(\theta_0) \cdot p(\psi). \tag{3.2.3}
\end{aligned}$$

The likelihood term $p(y_{1:T} \mid \theta_{0:T}, \psi)$ relates to the cumulative product of the observation equation $\prod_{t=1}^T p(y_t \mid \theta_t, \psi)$, whereas the prior term $p(\theta_{0:T}, \psi)$ relates to the cumulative product of the state $\prod_{t=1}^T p(\theta_t \mid \theta_{t-1}, \psi)$ times the priors $p(\theta_0)$ and $p(\psi)$. For the prior distribution, it is reasonable to assume that the prior parameters are independent so that $p(\psi) = p(W, V) = p(W) \cdot p(V)$. Further, if no available information is available a priori on the distribution of W and V , a non-informative prior distribution such as a sufficiently wide uniform distribution can be used.

To obtain the marginal smoothing distribution $p(\theta_t \mid y_{1:T})$, one may perform, in two phases, MCMC smoothing $\int p(\theta_t, \psi \mid y_{1:T}) d\psi$ followed by Kalman smoothing $p(\theta_t \mid y_{1:T}, \psi = \psi^*)$ where ψ^* is the true value, typically fixed by MLE.

3.2.2 Sequential Monte Carlo method

Stochastic methods for the general state-space model include bootstrap filter [40], auxiliary particle filter [41], Liu & West filter and Rao-Blackwellized Liu & West filter [42]–[44]. These methods will be explained in detail.

MCMC operates as a batch method and can be inefficient if data is obtained sequentially in an online fashion due to the need for recalculation at each time point. Thus, a more intelligent sequential estimation in the general model, akin to the Kalman filter [31], [45] for the linear model, is required. Several proposals have been made to deal with this problem, including unscented Kalman filter, extended Kalman filter, ensemble Kalman filter and Gaussian-sum filter to note a few, all of which sequentially estimate mean and covariance statistics of the posterior distribution. Particle filter approximates the posterior using a collection of particles possessing weight probabilities for each realization value.

To begin, importance sampling is accomplished as follows. Consider the adapted problem

of sampling from proposal q instead of target distribution p :

$$\begin{aligned}
\mathbb{E}_{p(\theta)}[f(\theta)] &= \int f(\theta)p(\theta) \, \mathrm{d}\theta \approx \frac{1}{N} \sum_{n=1}^N f(\theta^{(n)}) \\
&= \int f(\theta) \frac{p(\theta)}{q(\theta)} q(\theta) \, \mathrm{d}\theta \\
&= \int f(\theta) w(\theta) q(\theta) \, \mathrm{d}\theta \\
&= \mathbb{E}_{q(\theta)}[w(\theta) f(\theta)] \\
&\approx \frac{1}{N} \sum_{n=1}^N w(\theta^{(n)}) f(\theta^{(n)}),
\end{aligned}$$

where $w = \frac{p}{q}$ is an importance function.

Sequential importance sampling (SIS) considers a weighted Monte Carlo approximation to $p(\theta_{0:t}, \psi \mid y_{1:t})$ as

$$\hat{p}(\theta_{0:t}) = \sum_{n=1}^N \omega_t^{(n)} \delta(\theta_{0:t} - \theta_{0:t}^{(n)}),$$

whereby in the sequential variation, weights are considered:

$$\begin{aligned}
\omega_t^{(n)} &= \frac{p(\theta_{0:t}^{(n)} \mid y_{1:t})}{q(\theta_{0:t}^{(n)} \mid y_{1:t})} \\
&= \frac{p(y_t \mid \theta_t^{(n)}) p(\theta_t^{(n)} \mid \theta_{t-1}^{(n)}) p(\theta_{0:t-1}^{(n)} \mid y_{1:t-1})}{q(\theta_t^{(n)} \mid \theta_{0:t-1}^{(n)}, y_t) q(\theta_{0:t-1}^{(n)} \mid y_{1:t-1})} \\
&= \omega_{t-1}^{(n)} \frac{p(\theta_t^{(n)} \mid \theta_{t-1}^{(n)}) p(y_t \mid y_{1:t-1}, \theta_t^{(n)})}{q(\theta_t^{(n)} \mid \theta_{0:t-1}^{(n)}, y_{1:t})}.
\end{aligned}$$

The method is a sequential Monte Carlo (SMC) [35]. SIS with resampling gives rise to particle filtering.

Some properties of particle filter are provided. Setting the proposal function q equal to the state equation produces a bootstrap filter or Monte Carlo filter. Particle degeneracy may occur when weights are biased toward only few particles. Resampling reuses a limited number of particles efficiently and is necessary due to finite particles. Resampling eliminates particles with small weights and accumulates those with large weights in the filtering distribution. Effective sample size (ESS) [46], a proxy of the variance of the importance weights, may be used to determine when to resample. ESS provides an approximation of the number of independent samples from the target distribution that would be required to provide an

Algorithm 11 Particle filtering

Input. filtering distribution at time $t - 1$: $\{\theta_{t-1}^{(n)}, \omega_{t-1}^{(n)}\}_{n=1}^N$

Time t update rule:

For $n = 1$ to N :

(realizations) Draw $\theta_t^{(n)}$ from proposal $q(\theta_t | \theta_{0:t-1}^{(n)}, y_{1:t})$

(weights) $\omega_t^{(n)} \leftarrow \omega_{t-1}^{(n)} \frac{p(\theta_t^{(n)} | \theta_{t-1}^{(n)}) p(y_t | y_{1:t-1}, \theta_t^{(n)})}{q(\theta_t^{(n)} | \theta_{0:t-1}^{(n)}, y_{1:t})}$

End

(normalization) $\omega_t^{(n)} \leftarrow \frac{\omega_t^{(n)}}{\sum_{n=1}^N \omega_t^{(n)}}$

(resampling) Sample from $\{1, \dots, N\}$ with replacement, with probability proportional to $\omega_t^{(n)}$ ($n = 1, \dots, N$) and set the index sequence $k = \{k_1, \dots, k_N\}$ for resampling

Relabel $\theta^{(k)} = \{\theta_t^{(k_1)}, \dots, \theta_t^{(k_N)}\}$ to obtain realizations $\theta_t^{(n)}$ and reset every weight $\omega_t^{(n)}$ to $\frac{1}{N}$

Output. filtering distribution at time t : $\{\theta_t^{(n)}, \omega_t^{(n)}\}_{n=1}^N$

estimate of comparable variance. An approximate solution that experiences accumulation of errors as it processes the series and the filter process should be restarted periodically to avoid deterioration of estimation accuracy.

Monte Carlo (bootstrap) filter [40] is depicted next.

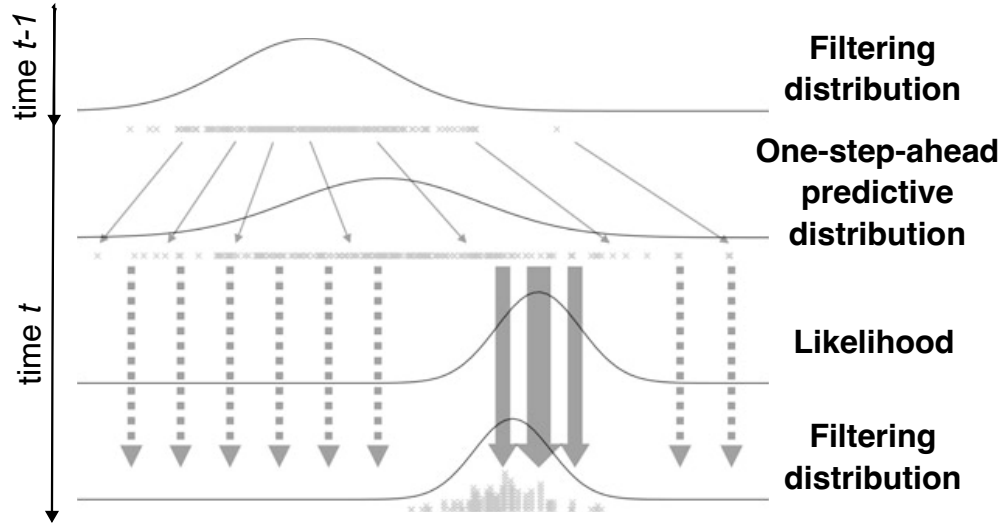


Figure 3.1: Graphical depiction of the bootstrap filter.

The mathematical foundation of SMC is grounded in probability theory. Target distribution p is approximated by weighted Monte Carlo samples, or particles

$$\{\theta_{0:t}^{(n)}, \psi^{(n)}, \omega_t^{(n)}\}_{n=1}^N \sim p(\theta_{0:t}, \psi | y_{1:t}),$$

such that $N \rightarrow \infty, \omega_t^{(n)} > 0, \sum_{n=1}^N \omega_t^{(n)} = 1$. Then for p -integrable function $\Phi_t : \mathbb{R}^{t \times p} \rightarrow \mathbb{R}$, the empirical distribution converges asymptotically to p as $N \rightarrow \infty$:

$$\sum_{n=1}^N \omega_t^{(n)} \Phi_t(\theta_{0:t}^{(n)}) \xrightarrow{a.s.} E_p(\Phi_t) = \int \Phi_t(\theta_{0:t}) p(\theta_{0:t} | y_{1:t}) d\theta_{0:t},$$

Some convergence results are as follows. Further, Chopin (2004) [47] and Del Moral (2004) [48] contain extensive results. Let

$$\begin{aligned} \bar{\Phi}_t &= E_p(\Phi_t) = \int \Phi_t(\theta_{0:t}) p(\theta_{0:t} | y_{1:t}) d\theta_{0:t}, \\ \hat{\Phi}_t &= \int \Phi_t(\theta_{0:t}) \hat{p}(\theta_{0:t} | y_{1:t}) d\theta_{0:t} = \sum_{n=1}^N \omega_t^{(n)} \Phi_t(\theta_{0:t}^{(n)}). \end{aligned}$$

Then there exists a constant C such that for any $r > 0$,

$$E[|\hat{\Phi}_t - \bar{\Phi}_t|^r]^{\frac{1}{r}} \leq \frac{C}{\sqrt{N}}, \quad (3.2.4)$$

$$\lim_{N \rightarrow \infty} \sqrt{N}(\hat{\Phi}_t - \bar{\Phi}_t) \xrightarrow{d} \mathcal{N}(0, \sigma_t^2). \quad (3.2.5)$$

Particle prediction is outlined below.

Algorithm 12 Particle prediction

Input. $k - 1$ -steps-ahead predictive distribution at time $t + k - 1 : \{\theta_{t-1}^{(n)}, \omega_{t+k-1}^{(n)}\}_{n=1}^N$
Time $t + k$ update rule:
For $n = 1$ to N :
(realizations) Draw $\theta_{t+k}^{(n)}$ from state equation $p(\theta_{t+k} | \theta_{t+k-1}^{(n)})$
(weights) $\omega_{t+k}^{(n)} \leftarrow \omega_{t+k-1}^{(n)}$
End
Output. k -steps-ahead predictive distribution at time $t + k : \{\theta_{t+k}^{(n)}, \omega_{t+k}^{(n)}\}_{n=1}^N$

Particle smoothing is accomplished by a forward filtering pass followed by a backward simulation.

Auxiliary particle filter (APF) resamples past particles with weights proportional to proposal $p(y_t | \hat{\theta}_t^{(n)})$, where $\hat{\theta}_t^{(n)} = E[\theta_t | \theta_{t-1}^{(n)}]$ is an estimate of the current state. The resampling is done by creating an index sequence, or auxiliary variable k , and then drawing realizations from the state equation. This prescreens the particles on the basis of the current observations. At random, it selects which of the particles from the previous recursion to propagate over time while conditioning on the current observation so that it is more likely

Algorithm 13 Forward Filtering Backward Simulation

Input. smoothing distribution at time $t + 1$: $\{s_{t+1}^{(n)}, \rho_{t+1}^{(n)}\}_{n=1}^N$

Time t update rule:

For $n = 1$ to N :

(smoothing weights) $\rho_t^{(n)} \leftarrow \omega_t^{(n)} p(s_{t+1}^{(n)} | \theta_t^{(n)})$

End

(normalization) $\rho_t^{(n)} \leftarrow \frac{\rho_t^{(n)}}{\sum_{n=1}^N \rho_t^{(n)}}$

(resampling) Sample from $\{1, \dots, N\}$ with replacement, with probability proportional to $\rho_t^{(n)}$ ($n = 1, \dots, N$) and set the index sequence $k = \{k_1, \dots, k_N\}$ for resampling

Relabel $s_t^{(k)} = \{s_t^{(k_1)}, \dots, s_t^{(k_N)}\}$ to obtain realizations $s_t^{(n)}$ and reset every weight $\rho_t^{(n)}$ to $\frac{1}{N}$

Output. smoothing distribution at time t : $\{s_t^{(n)}, \omega_t^{(n)}\}_{n=1}^N$

to select a particle with a high likelihood. The APF can sometimes experience performance degradation due to bias of the finite particles in the prior, which reduces diversity of the particles at the outset.

Algorithm 14 Auxiliary particle filtering

Input. filtering distribution at time $t - 1$: $\{\theta_{t-1}^{(n)}, \omega_{t-1}^{(n)}\}_{n=1}^N$

Time t update rule:

(resampling) Sample from $\{1, \dots, N\}$ with replacement, with probability proportional to $\omega_{t-1}^{(n)} p(y_t | y_{1:t-1}, \hat{\theta}_t^{(n)})$, $n = 1, \dots, N$ and set the index sequence $k = \{k_1, \dots, k_N\}$ for resampling, where $\hat{\theta}_t^{(n)} = E[\theta_t | \theta_{t-1}^{(n)}]$

For $n = 1$ to N :

(realizations) Draw $\theta_t^{(n)}$ from state equation $p(\theta_t | \theta_{t-1}^{(k_n)})$

(weights) $\omega_t^{(n)} \leftarrow \frac{p(y_t | y_{1:t-1}, \theta_t^{(n)})}{p(y_t | y_{1:t-1}, \hat{\theta}_t^{(k_n)})}$

End

(normalization) $\omega_t^{(n)} \leftarrow \frac{\omega_t^{(n)}}{\sum_{n=1}^N \omega_t^{(n)}}$

Output. filtering distribution at time t : $\{\theta_t^{(n)}, \omega_t^{(n)}\}_{n=1}^N$

For parameter estimation, a weighted kernel density method is used.

Algorithm 15 Kernel smoothing

Time t update rule:

moving average for parameters

(mean) $\mu^{(n)} \leftarrow a\psi_{t-1}^{(n)} + (1 - a)E_{\omega_{t-1}^{(n)}}[\psi_{t-1}^{(n)}]$

(variance reduction) $\gamma \leftarrow (1 - a^2)\text{Var}_{\omega_{t-1}^{(n)}}[\psi_{t-1}^{(n)}]$

For $n = 1$ to N :

(realizations) Draw $\psi_t^{(n)}$ from a continuous proposal with mean $\mu^{(n)}$ and variance γ

End

Combining APF with parameter estimation produces the widely-used Liu & West filter [49].

Algorithm 16 Liu and West filter

Input. filtering distribution at time $t - 1 : \{\psi_{t-1}^{(n)}, \theta_{t-1}^{(n)}, \omega_{t-1}^{(n)}\}_{n=1}^N$

Time t update rule:

moving average for parameters

(mean) $\mu^{(n)} \leftarrow a\psi_{t-1}^{(n)} + (1 - a)\mathbb{E}_{\omega_{t-1}^{(n)}}[\psi_{t-1}^{(n)}]$

(variance reduction) $\gamma \leftarrow (1 - a^2)\text{Var}_{\omega_{t-1}^{(n)}}[\psi_{t-1}^{(n)}]$

(resampling) Sample from $\{1, \dots, N\}$ with replacement, with probability proportional to $\omega_{t-1}^{(n)}p(y_t | y_{1:t-1}, \hat{\psi}_t^{(n)}, \hat{\theta}_t^{(n)})$, $n = 1, \dots, N$ and set the index sequence $k = \{k_1, \dots, k_N\}$ for resampling, where $\hat{\psi}_t^{(n)} = \mu^{(n)}$, $\hat{\theta}_t^{(n)} = \mathbb{E}[\theta_t | \theta_{t-1}^{(n)}, \mu^{(n)}]$

For $n = 1$ to N :

(realizations of parameters) Draw $\psi_t^{(n)}$ from a continuous proposal with mean $\mu^{(k_n)}$ and variance γ

(realizations of state) Draw $\theta_t^{(n)}$ from the state equation $p(\theta_t | \theta_{t-1}^{(k_n)}, \psi_t^{(n)})$

(weights) $\omega_t^{(n)} \leftarrow \frac{p(y_t | y_{1:t-1}, \psi_t^{(n)}, \theta_t^{(n)})}{p(y_t | y_{1:t-1}, \hat{\psi}_t^{(k_n)}, \hat{\theta}_t^{(k_n)})}$

End

(normalization) $\omega_t^{(n)} \leftarrow \frac{\omega_t^{(n)}}{\sum_{n=1}^N \omega_t^{(n)}}$

Output. filtering distribution at time $t : \{\psi_t^{(n)}, \theta_t^{(n)}, \omega_t^{(n)}\}_{n=1}^N$

For fixed parameter ψ , particles $\{\theta_t, \psi^{(i)}\}$ approximate $p(\theta_t, \psi | y_{1:t})$ such that

$$p(\psi | y_{1:t}) \approx \sum_{n=1}^N f_{\mathcal{N}}(\psi; \mu^{(n)}, (1 - a^2)\gamma),$$

where $a = \frac{3\Delta-1}{2\Delta}$ is the shrinkage factor for discount factor $\Delta \in (0, 1)$.

Rao-Blackwellization [36], [42] may be used to hybridize the Kalman and particle filters for the calculation of the joint filtering distribution $p(\theta_t, \psi | y_{1:t})$. By Bayes theorem, $p(\theta_t, \psi | y_{1:t}) = p(\theta_t | y_{1:t}, \psi) \cdot p(\psi | y_{1:t})$. Kalman and particle filter are applied to the first and second terms in the product, respectively. Rao-Blackwellization modifies the basic particle filter by treating realizations of the state θ_t as their parameters (mean m_t and covariance C_t of the filtering distribution) and Kalman filter is used to compute them at one time point. Further, the observation equation of the likelihood given state θ_t is treated as one-step-ahead predictive likelihood given parameters ψ whereby Kalman filter is used to compute them at one time point.

Rao-Blackwellized Liu & West particle filter is presented as follows.

Algorithm 17 Rao-Blackwellized Liu and West filter

Input. filtering distribution at time $t - 1$: $\{\psi_{t-1}^{(n)}, m_{t-1}^{(n)}, C_{t-1}^{(n)}, \omega_{t-1}^{(n)}\}_{n=1}^N$

Time t update rule:

moving average for parameters

(mean) $\mu^{(n)} \leftarrow a\psi_{t-1}^{(n)} + (1 - a)E_{\omega_{t-1}^{(n)}}[\psi_{t-1}^{(n)}]$

(variance reduction) $\gamma \leftarrow (1 - a^2)\text{Var}_{\omega_{t-1}^{(n)}}[\psi_{t-1}^{(n)}]$

(resampling) Sample from $\{1, \dots, N\}$ with replacement, with probability proportional to $\omega_{t-1}^{(n)}p(y_t \mid y_{1:t-1}, \hat{\psi}_t^{(n)}, \hat{\theta}_t^{(n)})$, $n = 1, \dots, N$ and set the index sequence $k = \{k_1, \dots, k_N\}$

for resampling, where $\hat{\psi}_t^{(n)} = \mu^{(n)}$ and calculate the one-step-ahead predictive likelihood $p(y_t \mid y_{t-1}, \hat{\psi}_t^{(n)})$ as $\mathcal{N}(y_t; f_t^{(n)}, Q_t^{(n)})$ using $f_t^{(n)}$ and $Q_t^{(n)}$ obtained from a Kalman filter with observations y_t , state $\{m_{t-1}^{(n)}, C_{t-1}^{(n)}\}$, parameters $\mu^{(n)}$

For $n = 1$ to N :

(realizations of parameters) Draw $\psi_t^{(n)}$ from a continuous proposal with mean $\mu^{(k_n)}$ and variance γ

(realizations of state) Obtain $m_t^{(n)}$ and $C_t^{(n)}$ from a Kalman filter with observations y_t , state $\{m_{t-1}^{(k_n)}, C_{t-1}^{(k_n)}\}$, parameters $\psi_t^{(n)}$.

(weights) $\omega_t^{(n)} \leftarrow \frac{p(y_t \mid y_{1:t-1}, \psi_t^{(n)})}{p(y_t \mid y_{1:t-1}, \psi_t^{(k_n)})}$ where the numerator is calculated as $\mathcal{N}(y_t; f_t^{(n)}, Q_t^{(n)})$

using realizations of state step and the denominator is calculated as $\mathcal{N}(y_t; f_t^{(k_n)}, Q_t^{(k_n)})$ using resampling step

End

(normalization) $\omega_t^{(n)} \leftarrow \frac{\omega_t^{(n)}}{\sum_{n=1}^N \omega_t^{(n)}}$

Output. filtering distribution at time t : $\{\psi_t^{(n)}, m_t^{(n)}, C_t^{(n)}, \omega_t^{(n)}\}_{n=1}^N$

3.2.3 Online structural change detection

Structural breaks are permanent disruptions of the level, slope, or seasonal component of the time series which indicates notable changes in the observed process. Change point detection [50] is a type of anomaly detection that is central to ascertaining structural change in time series. Treating the variance of the state noise w_t as time-varying via application of heavy-tailed distribution allows tracking of abrupt changes in the underlying state dynamics. Similarly, rare outliers may be handled by modelling the variance of the observation noise v_t as time-varying.

As a model for structural breaks, consider the random walk plus noise model:

$$\theta_t = \theta_{t-1} + w_t \quad w_t \sim \mathcal{N}(0, W\tau_t^2), \quad (3.2.6)$$

$$y_t = \theta_t + v_t \quad v_t \sim \mathcal{N}(0, V). \quad (3.2.7)$$

The state noise w_t contains a scale factor in its variance that is time-varying. The degree of ‘rarity’ of a structural break can not be assumed, thus an assumption of normality may be inappropriate depending on the application.

Shrinkage priors aim to shrink small effects to zero while maintaining true large effects and increase the bias of the model in order to reduce variance and improve overall model performance (lower generalization error). Here, consider the following hierarchical priors, formulated as scale mixture of normals [51], on the w_t : ridge, local Student’s t , Bayesian lasso, and horseshoe.

The ridge prior [52] parameter λ determines the degree of shrinkage, with larger values resulting in smaller prior variation leading to more shrinkage of the state noise toward zero.

$$w_t \mid \lambda, \sigma^2 \sim \text{Normal}\left(0, \frac{\sigma^2}{\lambda}\right), \quad \text{for } t = 1, \dots, T,$$

$$\lambda \sim \text{half-Cauchy}(0, 1).$$

Local Student’s t Prior [53] implies a Student’s t_ν prior on w_t [54].

$$w_t \mid \tau_t^2, \sigma^2 \sim \text{Normal}(0, \sigma^2 \tau_t^2),$$

$$\tau_t^2 \mid \nu, \lambda \sim \text{Inv-Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2\lambda}\right) \quad \text{for } t = 1, \dots, T,$$

$$\lambda \sim \text{half-Cauchy}(0, 1).$$

Alternatively, setting $\tau_t^2 \mid \nu \sim \text{Inv-Gamma}(\frac{\nu}{2}, \frac{\nu}{2})$ gives $w_t \sim \text{Student-}t(\nu)$ [54].

Bayesian lasso [55] is implemented as follows:

$$w_t \mid \tau_t^2, \sigma^2 \sim \text{Normal}(0, \sigma^2 \tau_t^2),$$

$$\tau_t^2 \mid \lambda^2 \sim \text{Exponential}\left(\frac{\lambda^2}{2}\right) \quad \text{for } t = 1, \dots, T,$$

$$\lambda \sim \text{half-Cauchy}(0, 1)$$

Integrating out τ_t^2 yields double-exponential (Laplace) priors on the state noise:

$$w_t \mid \lambda, \sigma \sim \text{Double-Exponential}\left(0, \frac{\sigma}{\lambda}\right),$$

for $t = 1 \dots, T$.

Specifying the horseshoe [38]:

$$\begin{aligned} w_t \mid \tau_t^2 &\sim \text{Normal}(0, \tau_t^2), \\ \tau_t \mid \lambda &\sim \text{half-Cauchy}(0, \lambda) \quad \text{for } t = 1, \dots, T, \\ \lambda &\sim \text{half-Cauchy}(0, \sigma). \end{aligned}$$

For fixed $\lambda = \sigma = 1$, the prior for the shrinkage factor $\kappa_t = \frac{1}{1+\tau_t^2}$ resembles a horseshoe shaped Beta(0.5,0.5) prior. For large coefficients: $\kappa_t \approx 0$ (no shrinkage) whereas for small coefficients: $\kappa_t \approx 1$ (heavy shrinkage). A global-local shrinkage prior: time-specific local shrinkage component τ_t as well as a global shrinkage component λ [56]. The so-called sparsity-inducing priors are depicted below.

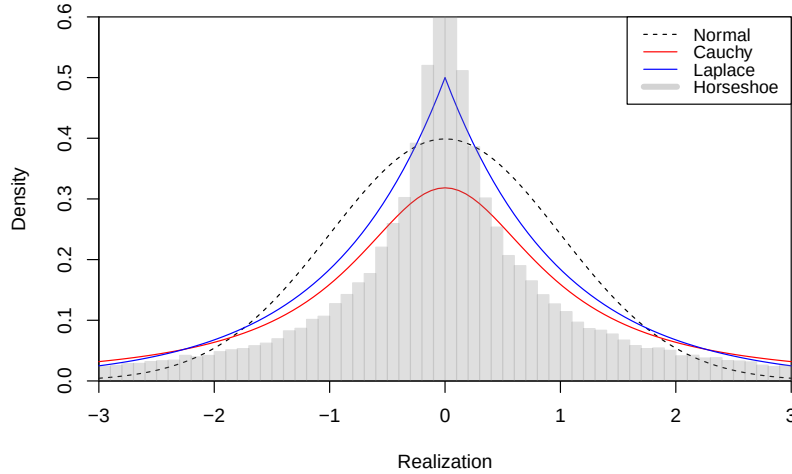


Figure 3.2: Shrinkage probability distributions.

3.3 Applications to understanding COVID-19 data

To illustrate the aforementioned concepts, in Ontario by first considering, a Kalman smoothing approach with MCMC used for the state sampling for a local level and seasonal model with COVID-19 case data from the province of Ontario, Canada [21].

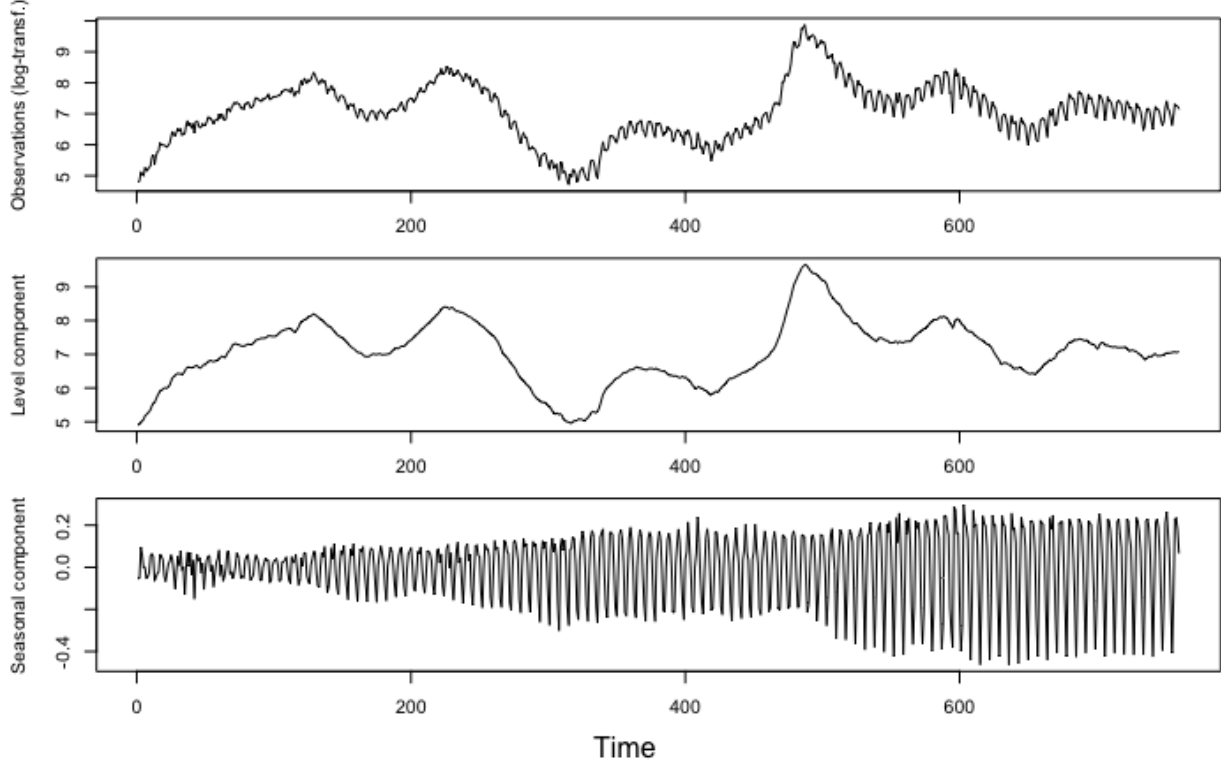


Figure 3.3: Local level and seasonal model for COVID-19 cases.

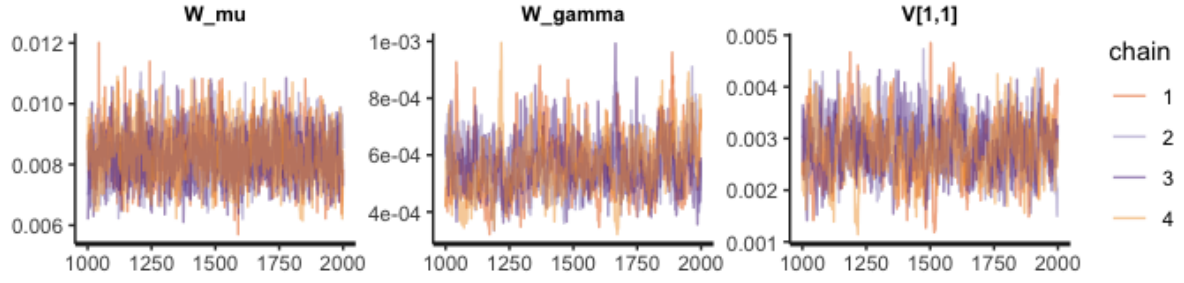


Figure 3.4: MCMC traceplots of the state noise variance of the level (W_μ) and seasonal components (W_γ) and observation noise variance V . State sampling by MCMC.

Parameter	Mean	SD	95% credible interval	n_{eff}	$\hat{R}$
W_μ	8.290e-03	8.114e-04	(6.817e-03, 1.001e-02)	1112.70	1.003537
W_γ	5.579e-04	9.392e-05	(3.977e-04, 7.693e-04)	244.82	1.014275
V	2.802e-03	5.289e-04	(1.831e-03, 3.897e-03)	440.24	1.007148

Table 3.1: Parameter estimate statistics for the inference approach with MCMC state sampling.

The results of the local level and seasonal model for COVID-19 cases without MCMC state sampling are shown below.

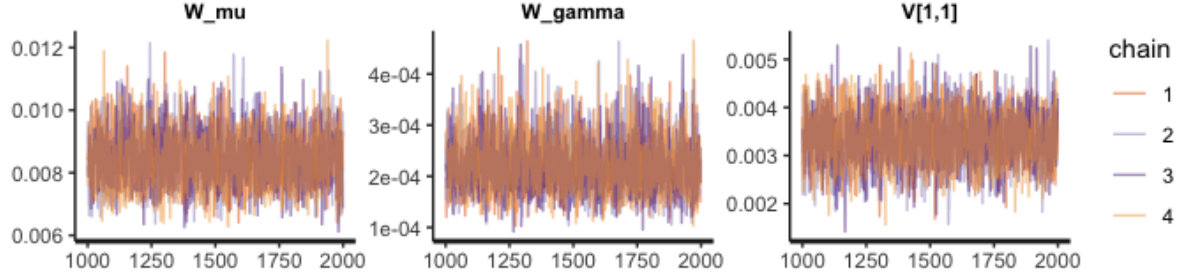


Figure 3.5: MCMC traceplots of the state noise variance of the level (W_μ) and seasonal components (W_γ) and observation noise variance V . State sampling by FFBS.

Parameter	Mean	SD	95% credible interval	n_{eff}	$\hat{R}$
W_μ	8.425e-03	8.554e-04	(6.939e-03, 1.024e-02)	2234.12	1.001947
W_γ	2.208e-04	5.299e-05	(1.345e-04, 3.433e-04)	2482.81	1.003342
V	3.327e-03	5.392e-04	(2.275e-03, 4.398e-03)	2186.27	1.001372

Table 3.2: Parameter estimate statistics for the inference approach without MCMC state sampling.

3.3.1 Bayesian inference for nonlinear dynamic regression models

Multivariate Bayesian dynamic models and applications are presented. Bayesian inference can be readily used to model multiple time series such as disease percent positivity by age over time, case counts and hospitalization common trend identification, or case counts and hospitalization stochastic volatilities. Throughout, consider COVID-19 data obtained from the province of Ontario, Canada [21].

First consider the following data of percent positivity by age over time.

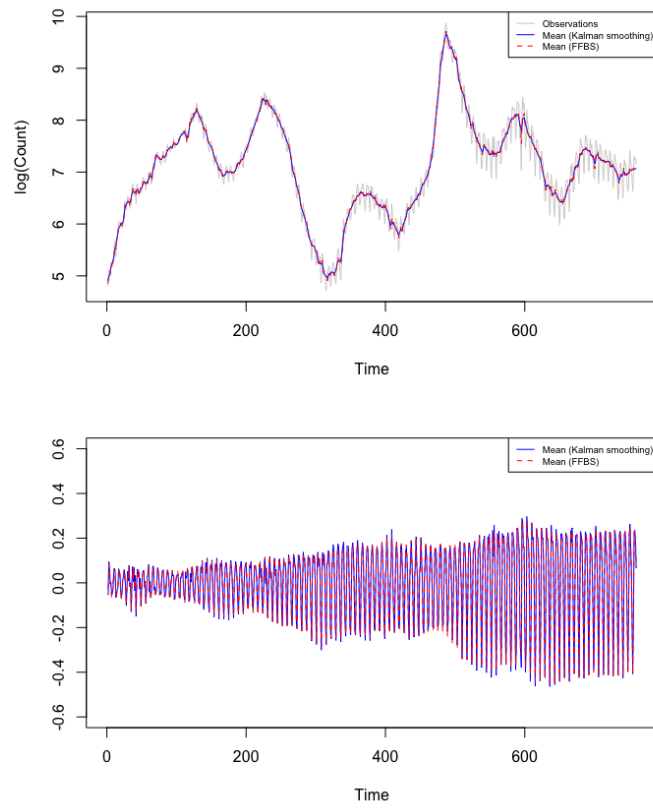


Figure 3.6: Local level and seasonal model for COVID-19 cases results. State sampling by FFBS.

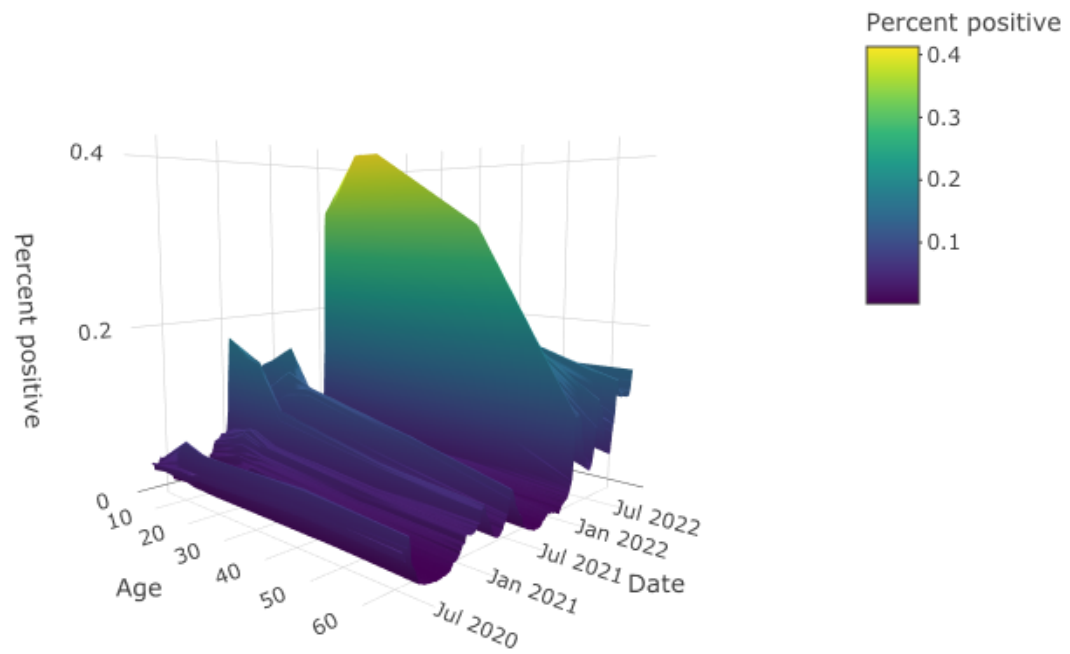


Figure 3.7: Percent positivity by age over time.

An alternative view of positivity after the large epidemic wave is shown below.

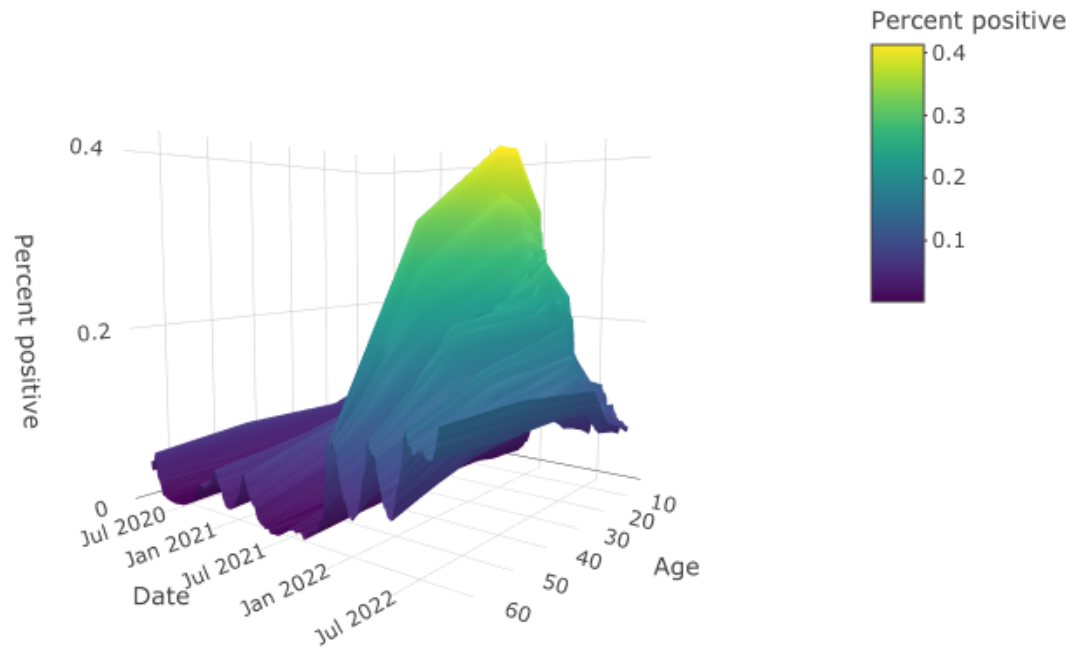


Figure 3.8: Alternative view of percent positivity by age over time after the large epidemic wave.

Observed and OLS fitted curves at selected dates are plotted and shown below.

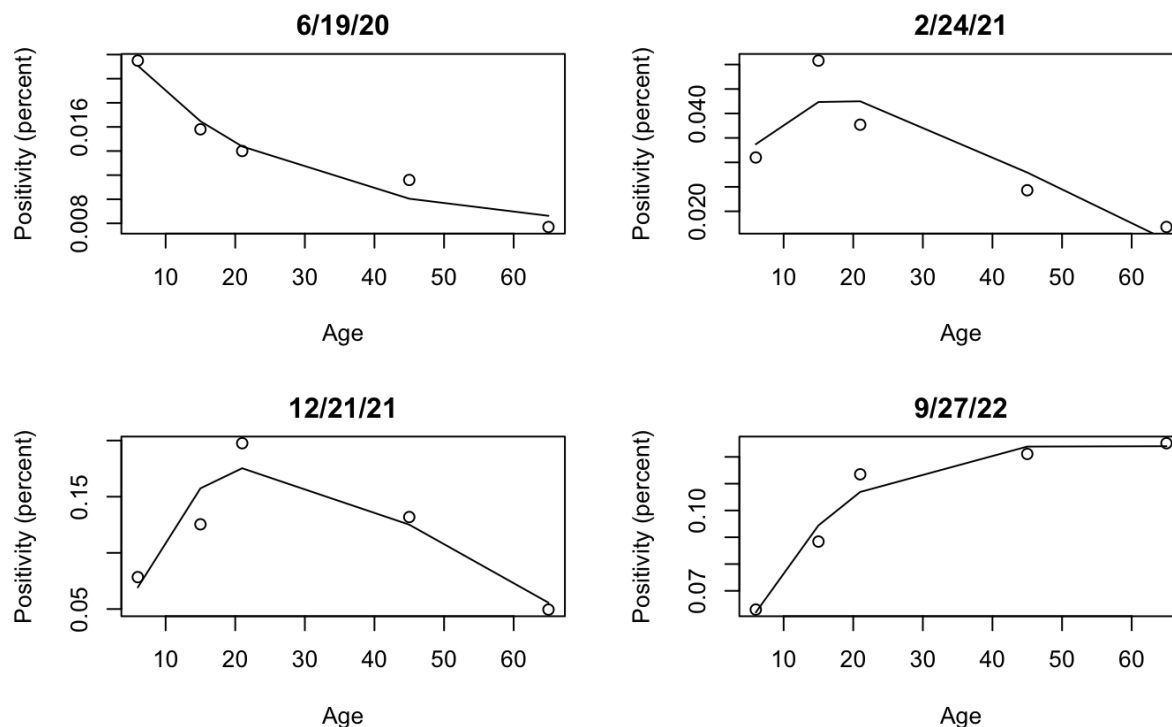


Figure 3.9: OLS fitted curves at selected dates.

Some diagnostics of the OLS fitting are as follows. The autocorrelation function (ACF) plots for Series 1, Series 2, and Series 3 display a gradual decline in correlation as the lag increases. This indicates positive autocorrelation, suggesting that each series exhibits persistence or memory over time. In other words, past values within each series have a measurable influence on future values.

The cross-correlation function (CCF) plots between the pairs of series, such as Series 1 and Series 2, Series 1 and Series 3, and Series 2 and Series 3, show weak or insignificant correlations at both positive and negative lags. This indicates that there are no strong linear relationships between the series at different lags, suggesting that the series evolve independently or are driven by separate factors.

These observations have implications for modelling. The independence observed between the series simplifies the modelling process, as there may be little need to include lagged covariates to account for cross-influences between series. However, the positive autocorrelation within each series indicates that their time-dependent structure should be modelled appropriately to capture their inherent persistence.

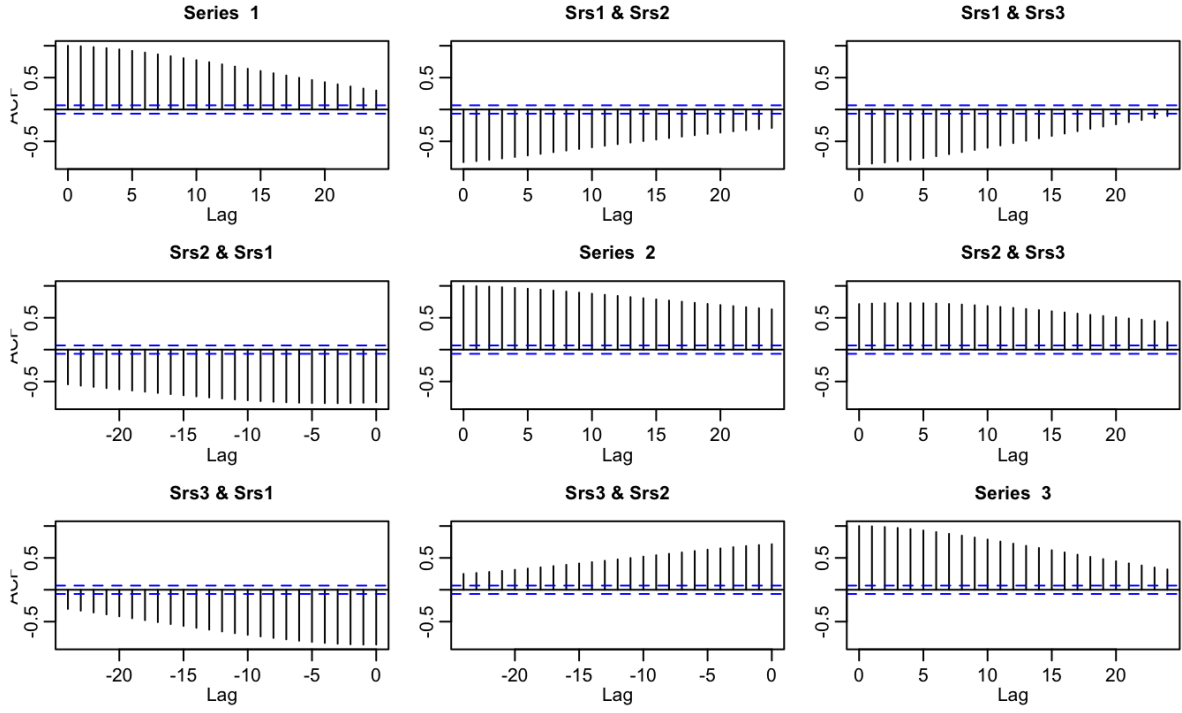


Figure 3.10: Autocorrelation (ACF) diagnostic plots.

Next, turn to dynamic estimation of positivity curve as follows.

$$y_t = F\theta_t + v_t, \quad v_t \sim \mathcal{N}(0, V),$$

$$\theta_t = G\theta_{t-1} + w_t, \quad w_t \sim \mathcal{N}(0, W),$$

where $y_t = [y_{1,t}, \dots, y_{m,t}]^\top$, $\theta_t = [\beta_{1,t}, \beta_{2,t}, \beta_{3,t}]^\top$, $F = \begin{bmatrix} 1 & h_2(x_1) & h_3(x_1) \\ 1 & h_2(x_2) & h_3(x_2) \\ \vdots & \vdots & \vdots \\ 1 & h_2(x_m) & h_3(x_m) \end{bmatrix}$, $G = \text{diag}(\psi_1, \psi_2, \psi_3)$,

$$V = \text{diag}(\phi_{y,1}^{-1}, \dots, \phi_{y,m}^{-1}), \quad W = \text{diag}(\phi_{\theta,1}^{-1}, \phi_{y,2}^{-1}, \phi_{y,3}^{-1}).$$

Computation of the MCMC smoothing and OLS estimates are conducted. The results are shown in the following plot and table.

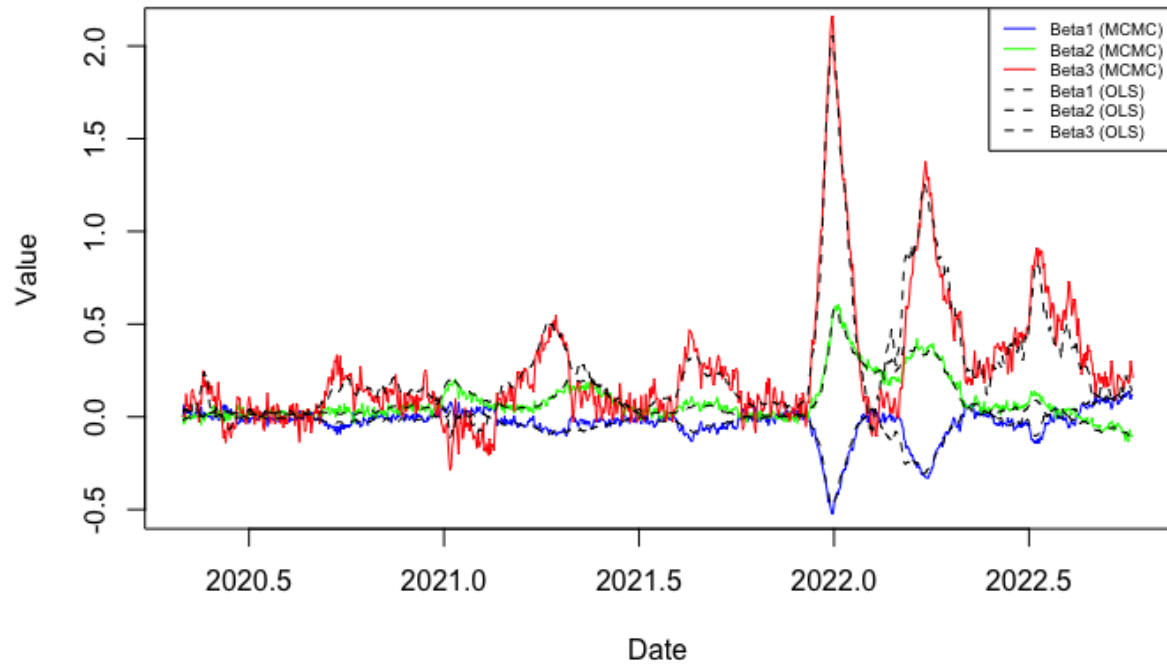


Figure 3.11: Comparison of MCMC and OLS smoothing estimates.

Parameter	Mean	SE
ψ_1	0.9737	0.0002
ψ_2	0.9841	0.0001
ψ_3	0.9885	0.0001
$\phi_{\theta,1}$	0.0520	0.0001
$\phi_{\theta,2}$	0.0535	0.0001
$\phi_{\theta,3}$	0.0802	0.0001
$\phi_{y,1}$	0.0096	0.0001
$\phi_{y,2}$	0.0245	0.0001
$\phi_{y,3}$	0.0090	0.0001
$\phi_{y,4}$	0.0130	0.0001
$\phi_{y,5}$	0.0088	0.0001

Table 3.3: MCMC posterior expectation approximations.

Next, consider the application of common trend identification between case counts and hospitalizations. This approach is useful for reducing the dimension of multiple correlated time series.

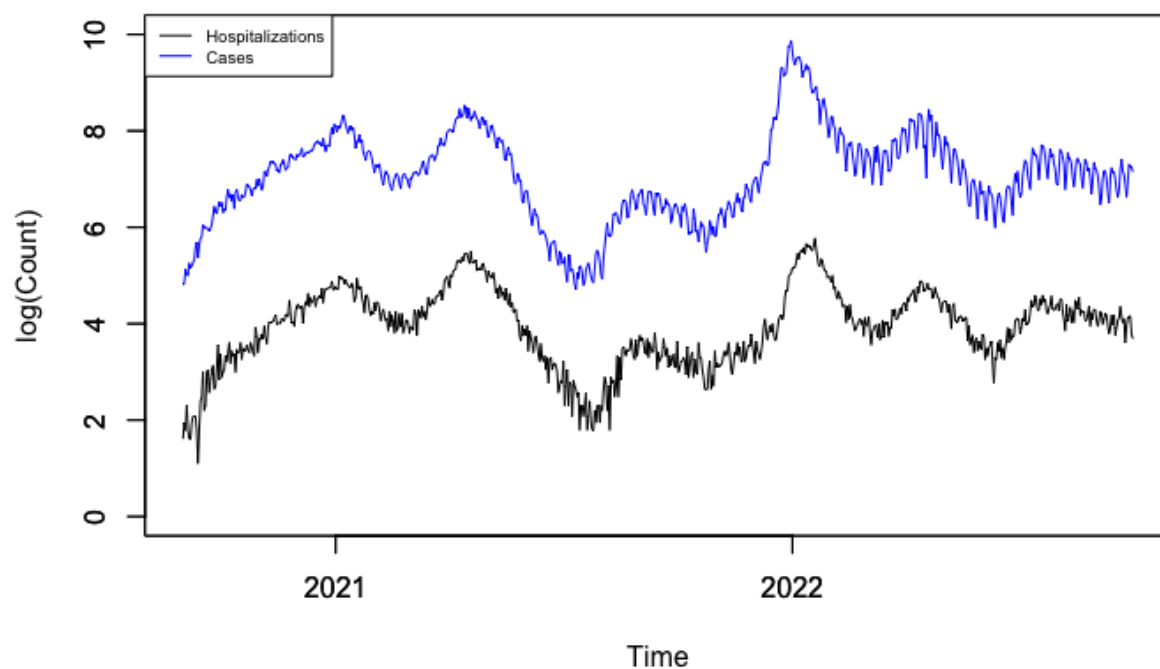


Figure 3.12: Case counts and hospitalizations.

For analysis, consider a dynamic factor model (DFM) which can be used to extract a common stochastic trend from multiple integrated time series. The goal is to explain fluctuations of various series by common latent factors that affect a set of variables simultaneously.

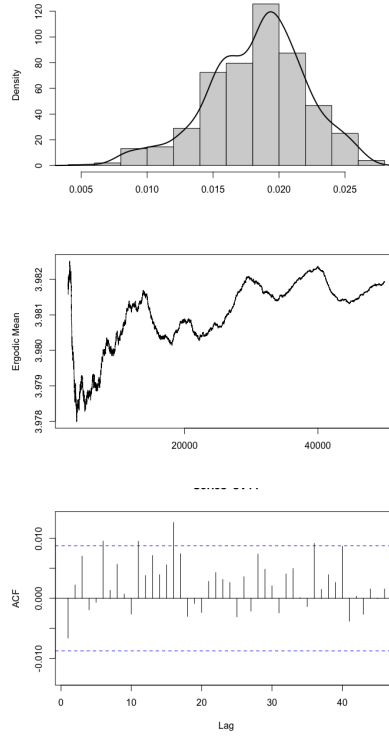


Figure 3.13: MCMC diagnostics for the DFM.

DFM parameter estimates are computed and recorded in the following table.

Parameter	α	σ_{μ}	σ_1	σ_2	σ_{12}
Posterior mean	222046.20	2.88e-05	3.9832	3299	-10126.43
Std. dev.	3.996223	9.26e-08	0.06	876.97	6.22
5% quantile	222038.4	2.86e-05	3.87	1580.12	-10138.62
95% quantile	222054.1	2.90e-05	4.10	5017.79	-10114.23

Table 3.4: DFM parameter estimates.

The posterior mean of common stochastic trend is shown in the following figure. The result shows that in this application, the trend closely follows the pattern of the hospitalizations.

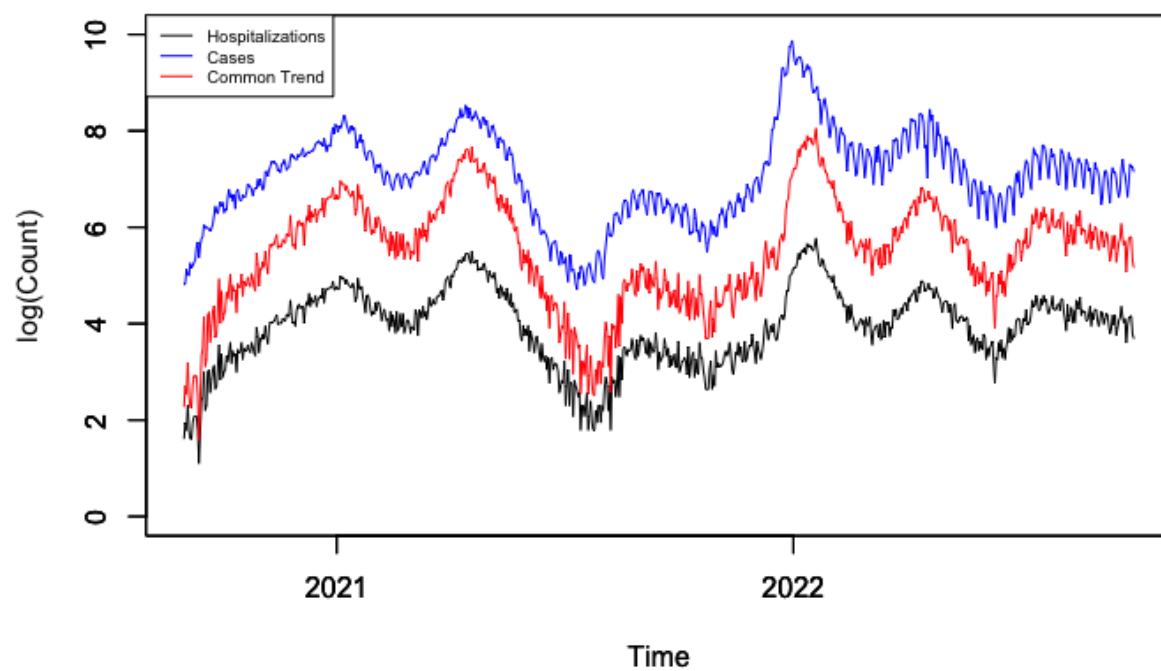


Figure 3.14: Original data with common stochastic trend.

Now, consider multivariate stochastic volatility model, which consists of correlated shocks in the count series and correlated log variances.

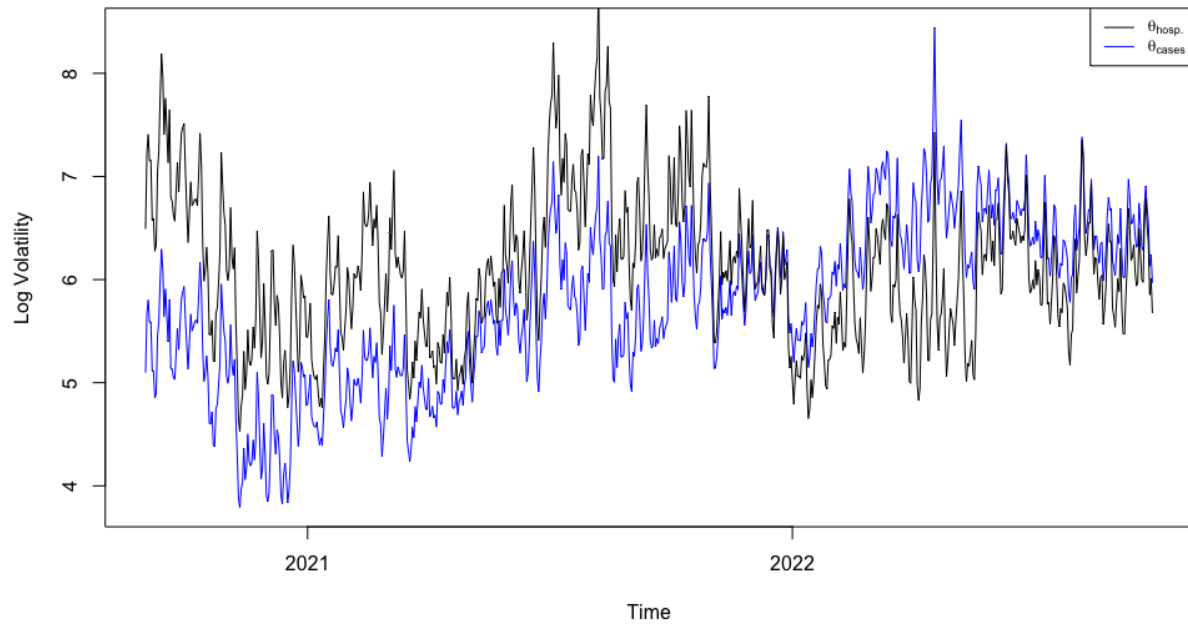


Figure 3.15: Posterior means of the log volatility.

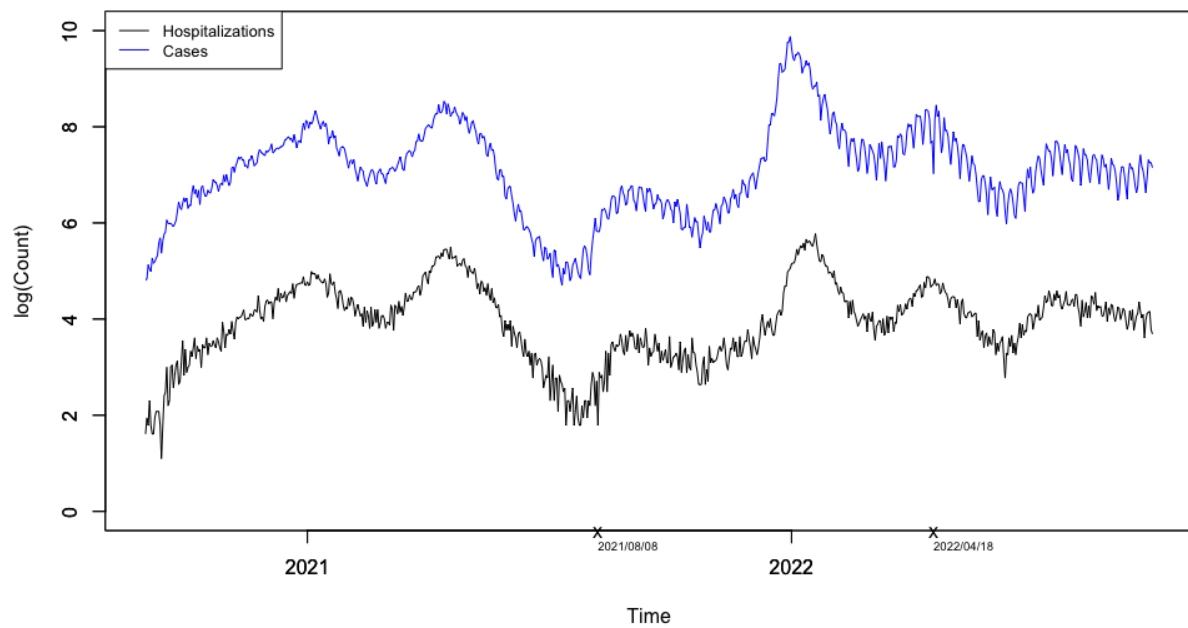


Figure 3.16: Original data series for the MSV model application.

Diagnostics of leave-one-out information criterion show extreme values for Aug. 8, 2021

when there was a rapid fall in cases and Apr. 18, 2022 when there was a sharp drop in hospitalizations.

3.3.2 Application of online structural change detection method

Next, consider a simulated data set with two notable change points at $t = 60$ and $t = 140$.

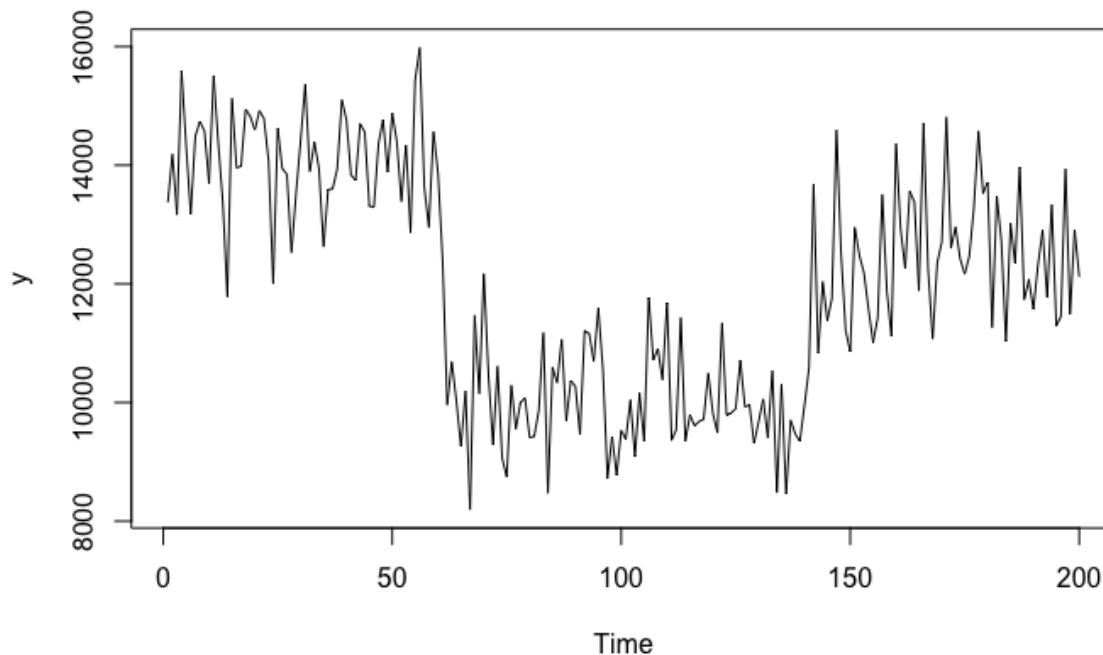


Figure 3.17: Simulated data set for change detection.

The normal distribution filter results show that the mean estimate from the particle filter closely follows the observations throughout the time series. The time-varying Kalman filter mean is almost identical to the particle filter mean, indicating a strong agreement between the two methods. The normal prior provides stable and accurate estimates, effectively capturing the overall trend without significant deviation.

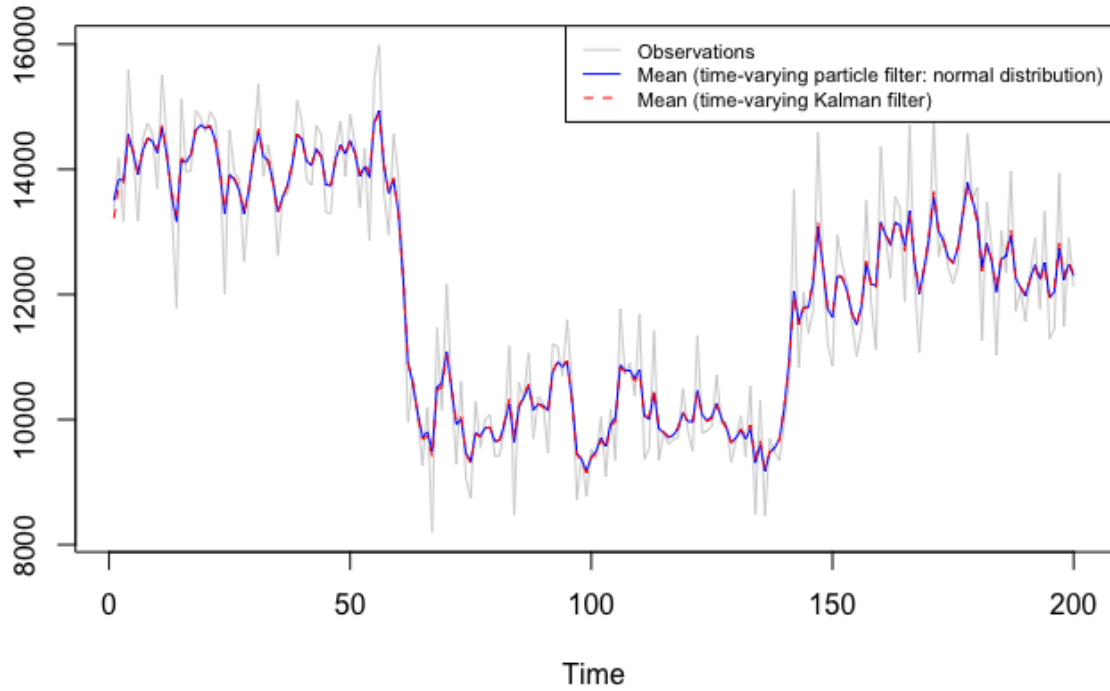


Figure 3.18: Normal distribution filter results.

The performance of the different priors is quantitatively evaluated by their log-likelihood values, which measure the fit of the models to the observed data. Higher log-likelihood values indicate a better fit. The following table summarizes the log-likelihoods for each prior:

Prior	$\ell(V, W)$
Normal	1579.879
Student's t	1910.972
Laplace	1675.949
Horseshoe	2346.218

Table 3.5: Log-likelihood of the models considered.

The horseshoe prior has the highest log-likelihood value (2346.218), indicating that it provides the best fit to the data among the priors considered.

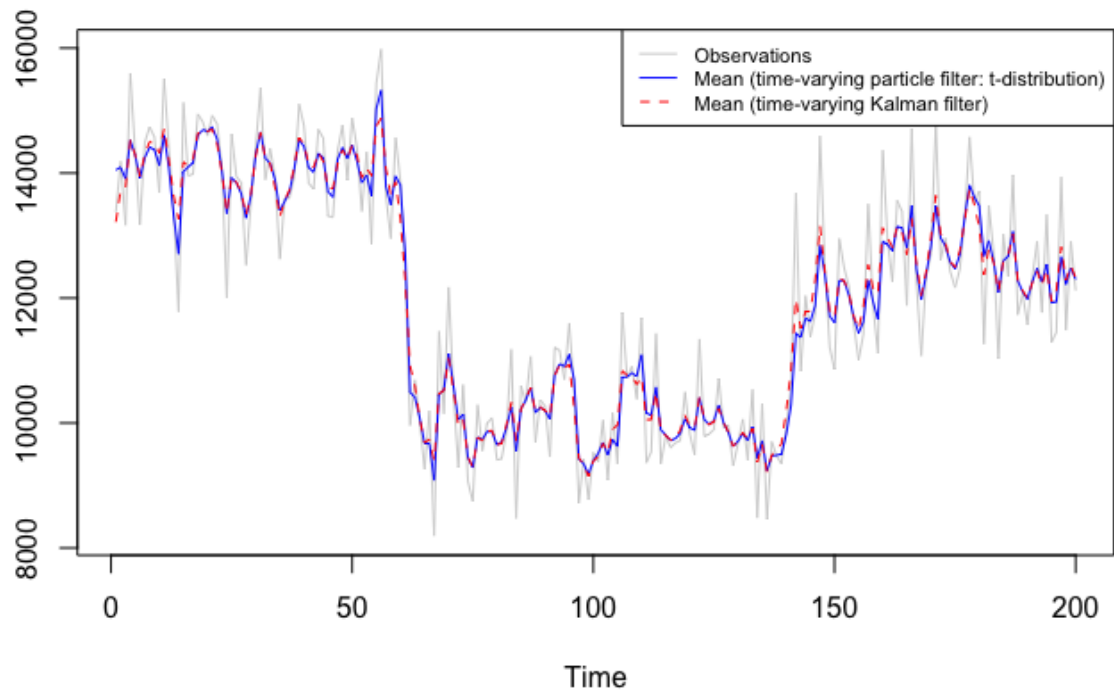


Figure 3.19: t -distribution filter results.

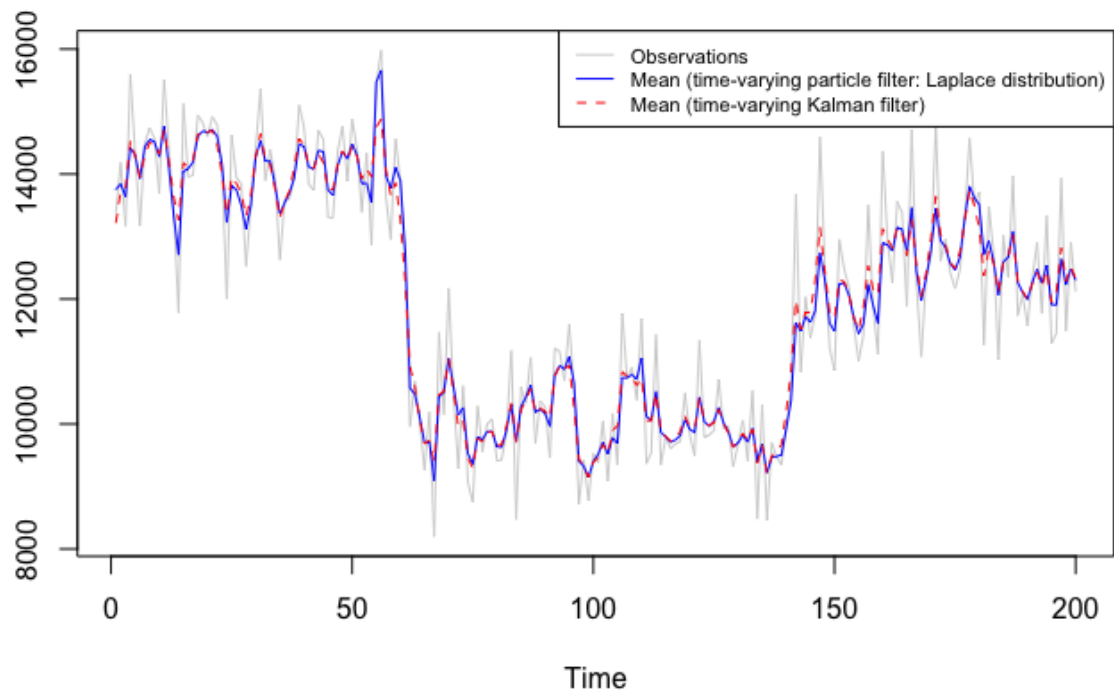


Figure 3.20: Laplace distribution filter results.

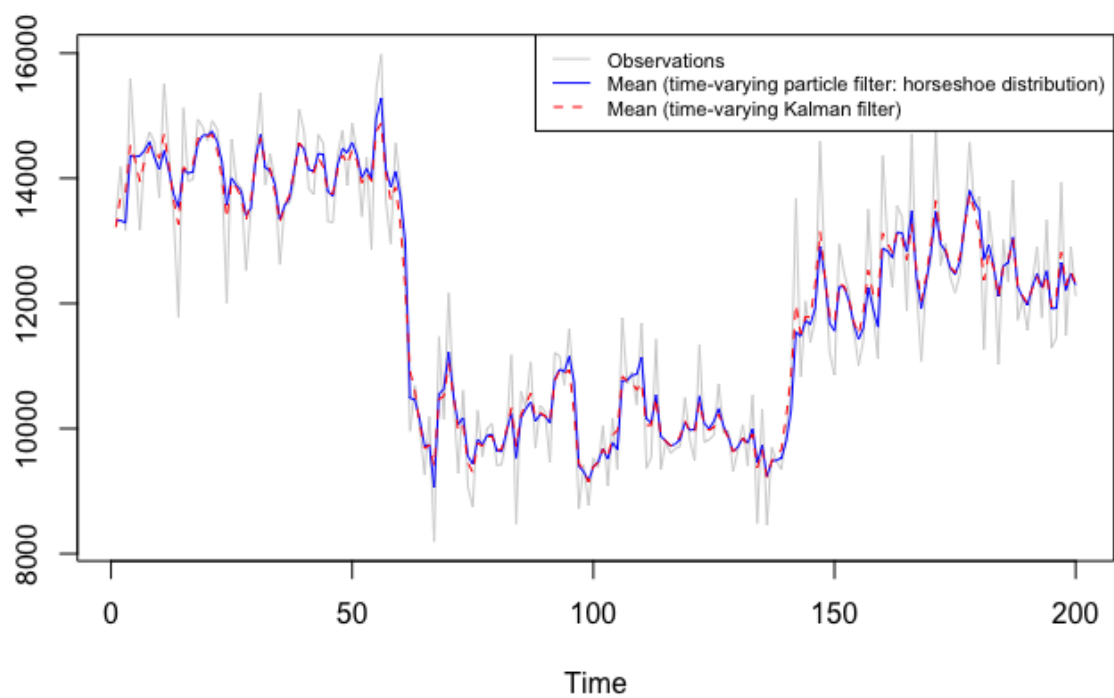


Figure 3.21: Horseshoe distribution filter results.

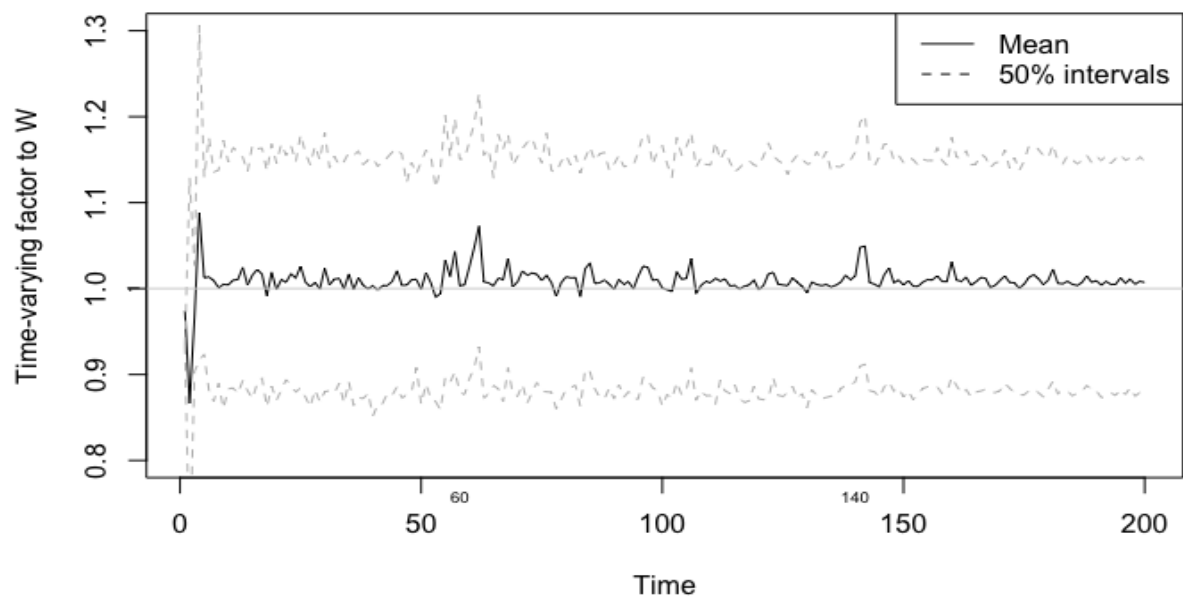


Figure 3.22: Parameter estimate for normal distribution model.

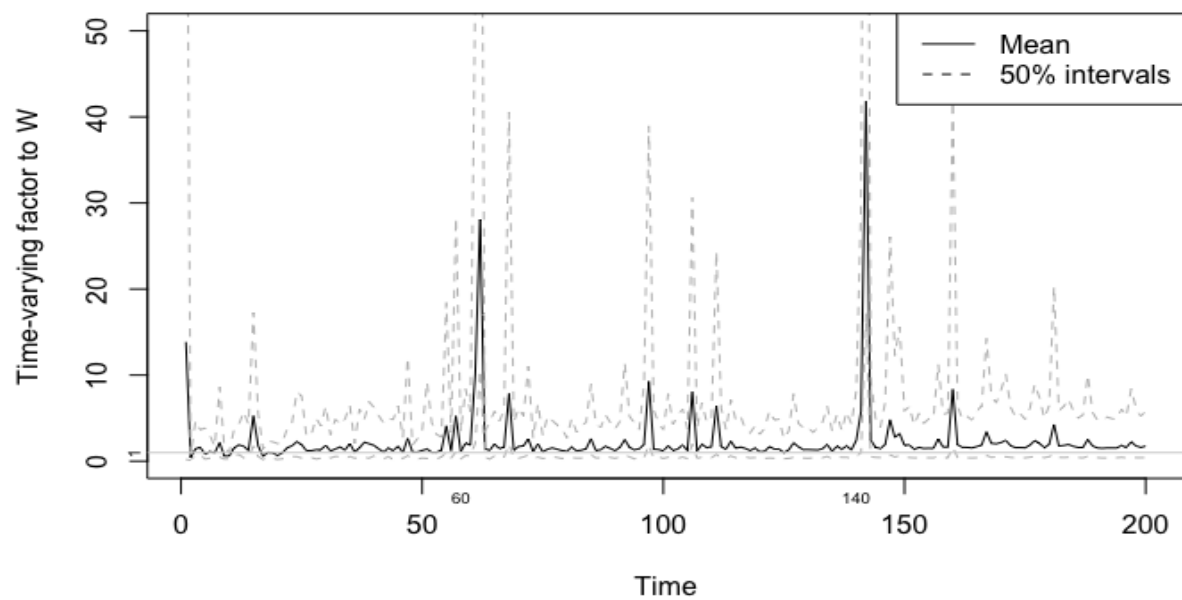


Figure 3.23: Parameter estimate for t -distribution model.

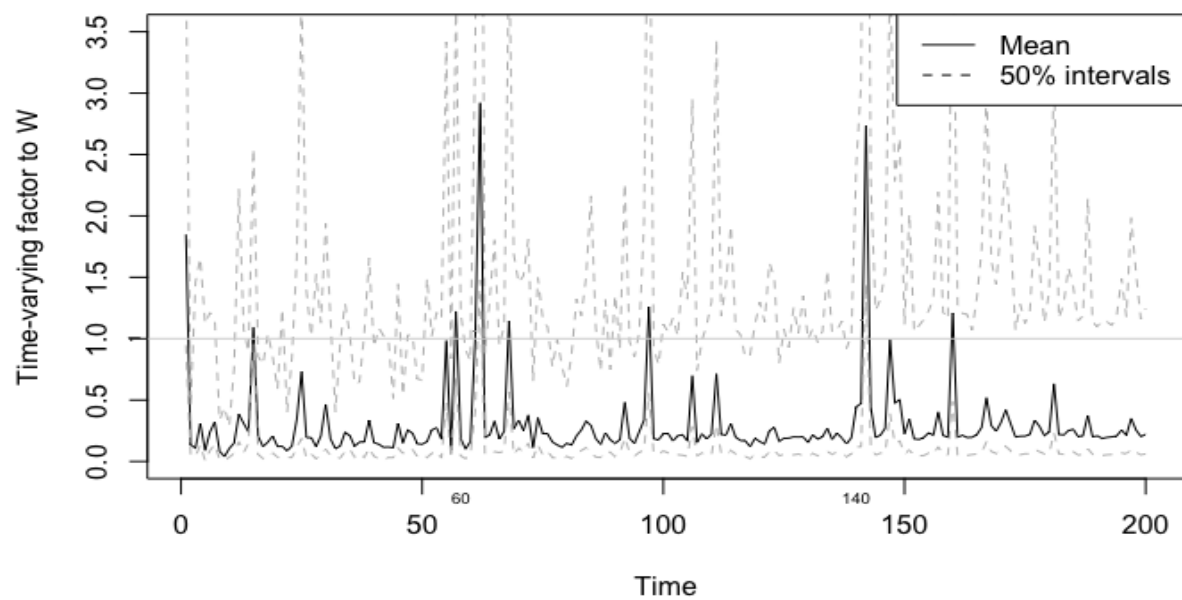


Figure 3.24: Parameter estimate for Laplace distribution model.

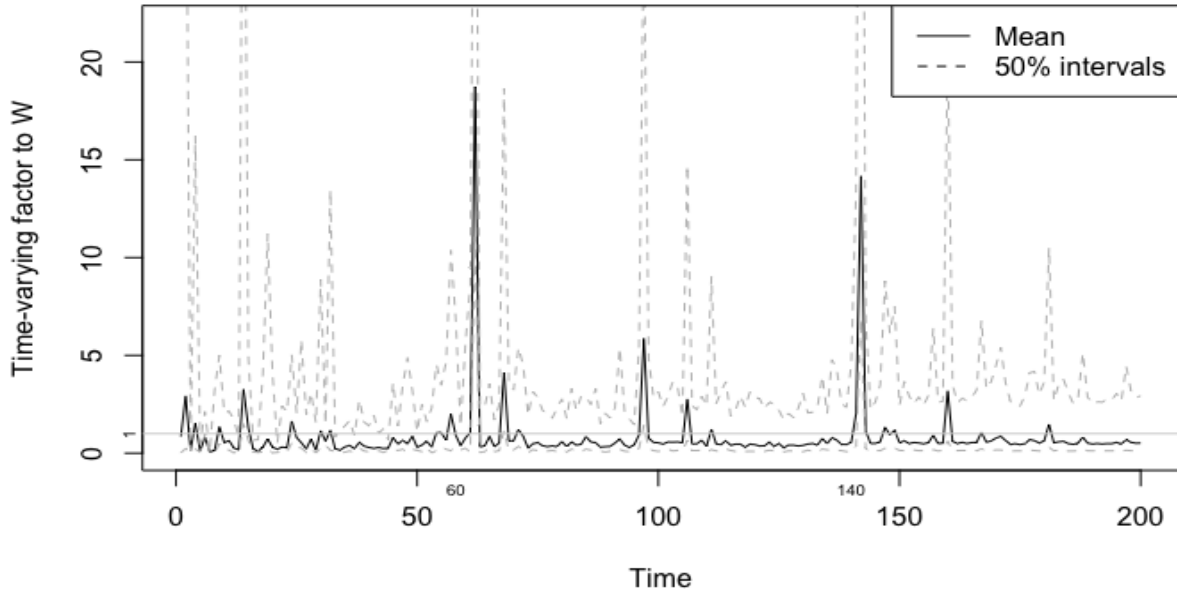


Figure 3.25: Parameter estimate for horseshoe distribution model.

The above plots of the time-varying factors for state noise variance τ_t using different distributional assumptions allows for the determination of the impact of these assumptions on the detection of structural changes in the data. The results are depicted in the plots, and show the mean and 50% intervals of the time-varying factors under various priors: normal, Student's t , Laplace, and horseshoe.

In summary, from the parameter estimates, it is clear that the normal distribution model struggles with identifying the change points while the heavy-tailed distributions are able to identify the change points. The Laplace model gives the correct change points if the parameter is thresholded at 1.5 while the t -distribution model at 10 and the horseshoe distribution model at 7, owing to the differences in the scales due to the distribution construction.

Finally, it is also interesting to observe that the data contains a less abrupt change at $t = 96$ for which there are notable, but lower parameter signals present in the parameter estimate plots as well.

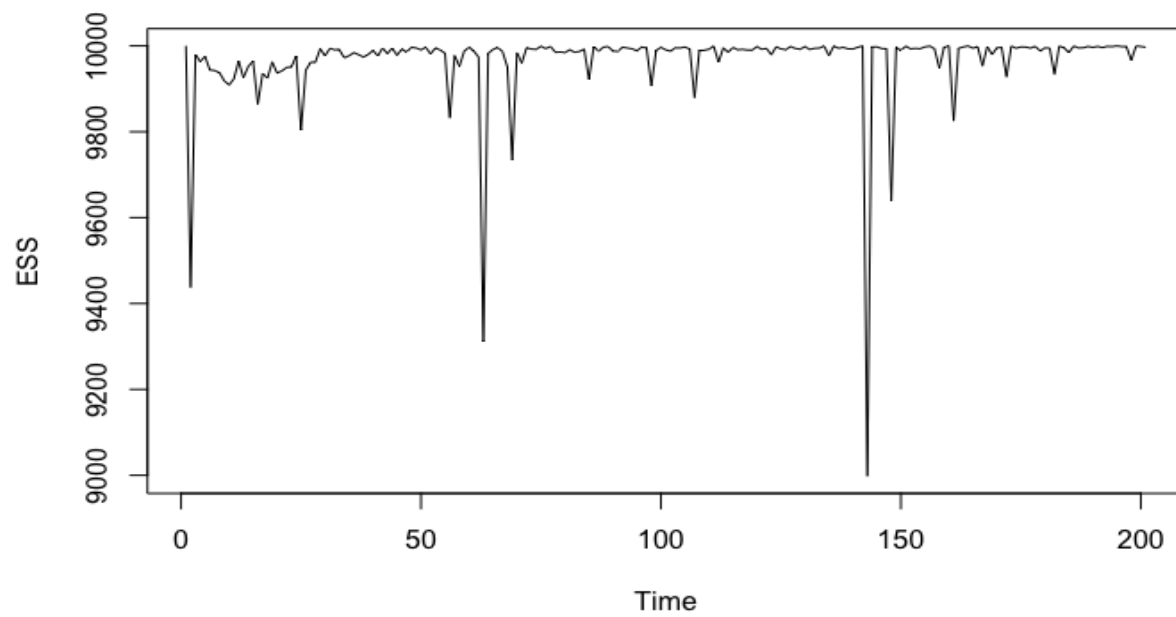


Figure 3.26: ESS for normal distribution model.

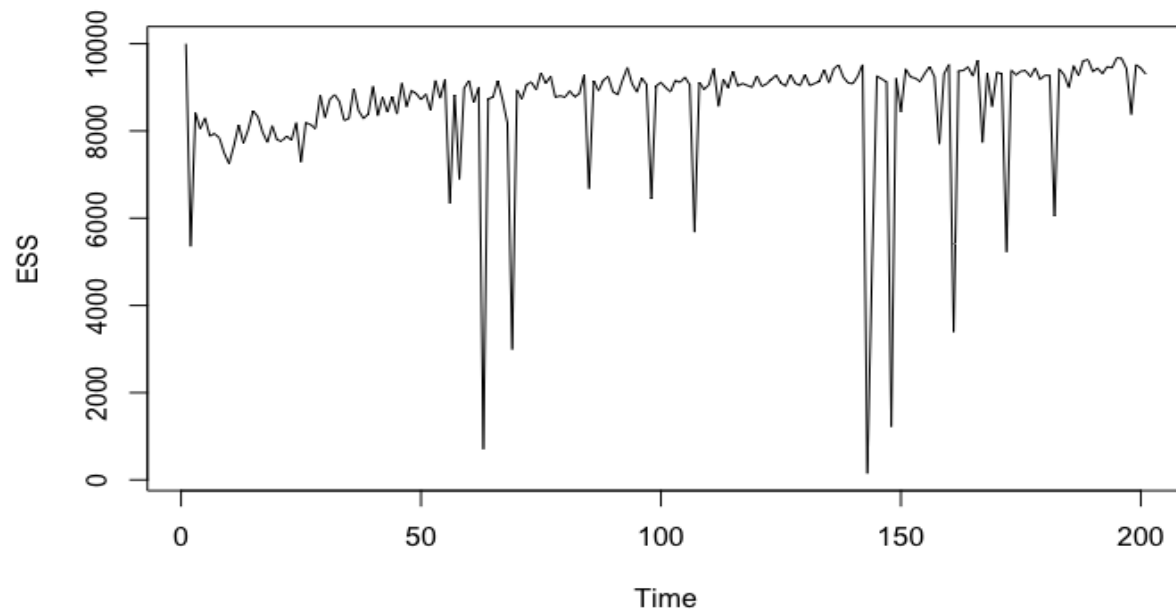


Figure 3.27: ESS for t -distribution model.

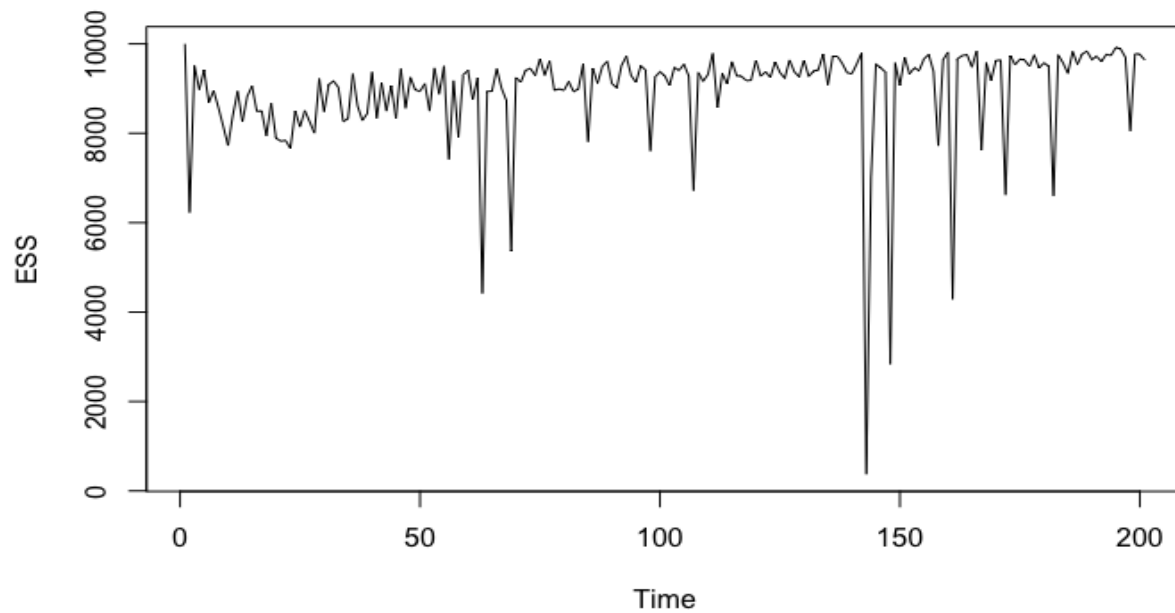


Figure 3.28: ESS for Laplace distribution model.

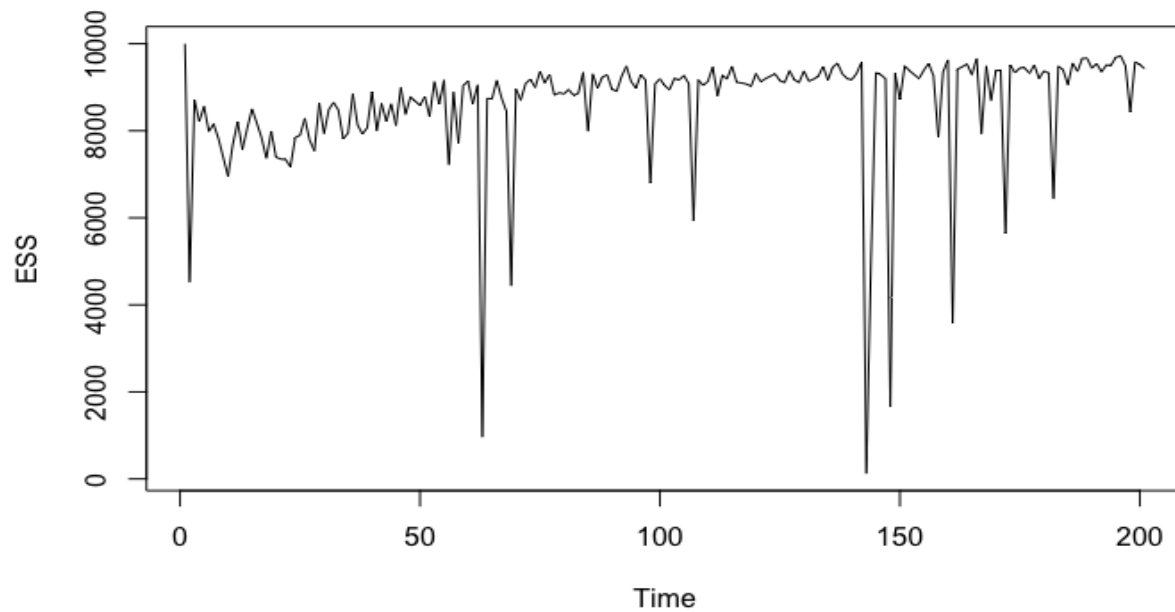


Figure 3.29: ESS for horseshoe distribution model.

The effective sample size (ESS) for state noise variance using different distributional assumptions is a crucial metric for evaluating the performance of sampling methods in capturing the underlying data distribution efficiently.

3.4 Discussion

Combining MCMC and Kalman filter may be used to draw a sample from $p(\theta_{0:T}, \psi \mid y_{1:T})$ as follows. By Bayes theorem, $p(\theta_{0:T}, \psi \mid y_{1:T}) = p(\theta_{0:T} \mid \psi, y_{1:T}) \cdot p(\psi \mid y_{1:T})$. Then, perform in the following order,

1. Parameter estimation: Draw a sample from $p(\psi \mid y_{1:T})$ by MCMC;
2. State estimation: Draw a sample from $p(\theta_{0:T} \mid \psi, y_{1:T})$ by FFBS.

In the above illustration, the marginal smoothing distribution is $p(\theta_t \mid y_{1:T})$. The MCMC method estimates $\int p(\theta_t, \psi \mid y_{1:T}) d\theta$ whereas Kalman smoothing calculates $p(\theta_t \mid y_{1:T}, \psi = \psi^*)$ where ψ^* is the true value. In theory, the two approaches should produce similar results, however there is no guarantee in practical application.

The estimation accuracy can be improved by various means. In theory, the performance gains can be made by increasing the number of iterations of the Markov chain. More recently, the replica exchange algorithm [57] has been shown to deal with the problem of having the Markov chain getting trapped in a localized region of the target distribution. Practically, the general state-space model can be estimated by resorting to combining the effectiveness of the Kalman filter in the linear Gaussian state-space model with the MCMC approach which can improve estimation accuracy. Initially, the linear Gaussian state-space model is used with unknown parameters regarded as random variables. The model can then be treated as a linear Gaussian state-space model if the parameters are known or can be estimated by maximum likelihood estimation. If the parameters are unknown and to be jointly estimated along with the underlying states, the target posterior distribution $p(\theta_{0:T}, \psi \mid y_{1:T})$ requires the general state-space model.

A combined algorithm is obtained by using MCMC to estimate the parameters ψ and then using this estimation result in a Kalman filter. MCMC estimates the posterior distribution $p(\psi \mid y_{1:T}) \propto p(y_{1:T} \mid \psi) \cdot p(\psi)$. The likelihood term of the linear Gaussian state-space model can be written according to:

$$\begin{aligned} \ell(\psi) &= \log p(y_1, \dots, y_T; \psi) \\ &= \sum_{t=1}^T \log p(y_t \mid y_{1:t-1}; \psi) \\ &= -\frac{1}{2} \sum_{t=1}^T \log |Q_t| - \frac{1}{2} \sum_{t=1}^T \frac{(y_t - f_t)^2}{Q_t}. \end{aligned}$$

and the prior distribution can be specified to be non-informative. The maximum likelihood estimates can then be fed into the Kalman filter for estimation. Alternatively, FFBS can efficiently provide samples from the joint posterior distribution $p(\theta_{0:T}, \psi \mid y_{1:T}) = p(\theta_{0:T} \mid \psi, y_{1:T}) \cdot p(\psi \mid y_{1:T})$.

Specifically, the term for the state follows linear Gaussian state-space model which can be efficiently sampled by using FFBS. As in the first approach, the distribution of the parameters given the observations is sampled from using the Kalman filter augmented MCMC method. The key when estimating random parameters in the linear Gaussian state-space model in tandem with the state is to utilize MCMC for the parameter estimation and this is what the two phase approach above accomplishes. That is, the two-step approach uses MCMC to draw samples from $p(\psi \mid y_{1:T})$ and then FFBS to draw samples from $p(\theta_{0:T} \mid \psi, y_{1:T})$ to provide the joint posterior distribution $p(\theta_{0:T}, \psi \mid y_{1:T}) = p(\theta_{0:T} \mid \psi, y_{1:T}) \cdot p(\psi \mid y_{1:T})$ as illustrated above. Such algorithm fusion proves to be helpful and will be returned to for stochastic simulation with the particle filter.

The comparative analysis of ESS under different priors provides valuable insights into their respective sampling efficiencies and robustness. The normal prior maintains a consistently high ESS with minimal variability, making it suitable for stable data with minimal structural changes. The Student's t prior, with its sensitivity to outliers, shows more variability in ESS but maintains high efficiency overall. The Laplace prior, promoting sparsity, experiences more frequent dips in ESS but quickly recovers, indicating a balance between stability and adaptability. The horseshoe prior, with its significant ESS fluctuations, excels in detecting structural changes and sparsity, making it highly effective for complex data structures with abrupt shifts.

This further highlights the importance of selecting appropriate priors based on the data characteristics and specific analysis requirements. The normal prior is ideal for stable data, while the Student's t and Laplace priors offer advantages in handling outliers and promoting sparsity, respectively. The horseshoe prior stands out for its ability to adapt to complex data structures, despite the greater fluctuations in ESS.

The application of the Rao-Blackwellized Liu & West filtering methodology with a horseshoe prior to the Ontario COVID-19 case count data [21] provides insightful results. The algorithm identifies two notable clusters of large values for the time-varying factor τ_t . By thresholding these values, two distinct change points are detected. These change points are plotted along with the original COVID-19 case count data, highlighting periods where significant shifts in the data occurred.

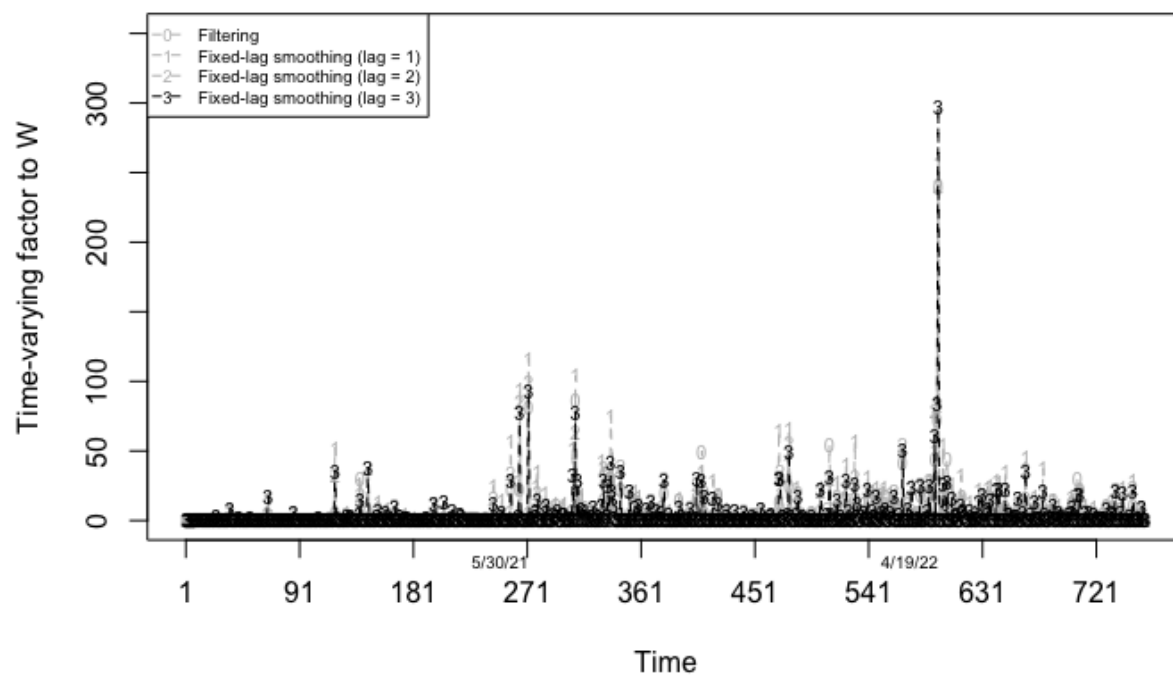


Figure 3.30: Parameter filtering and smoothing results.

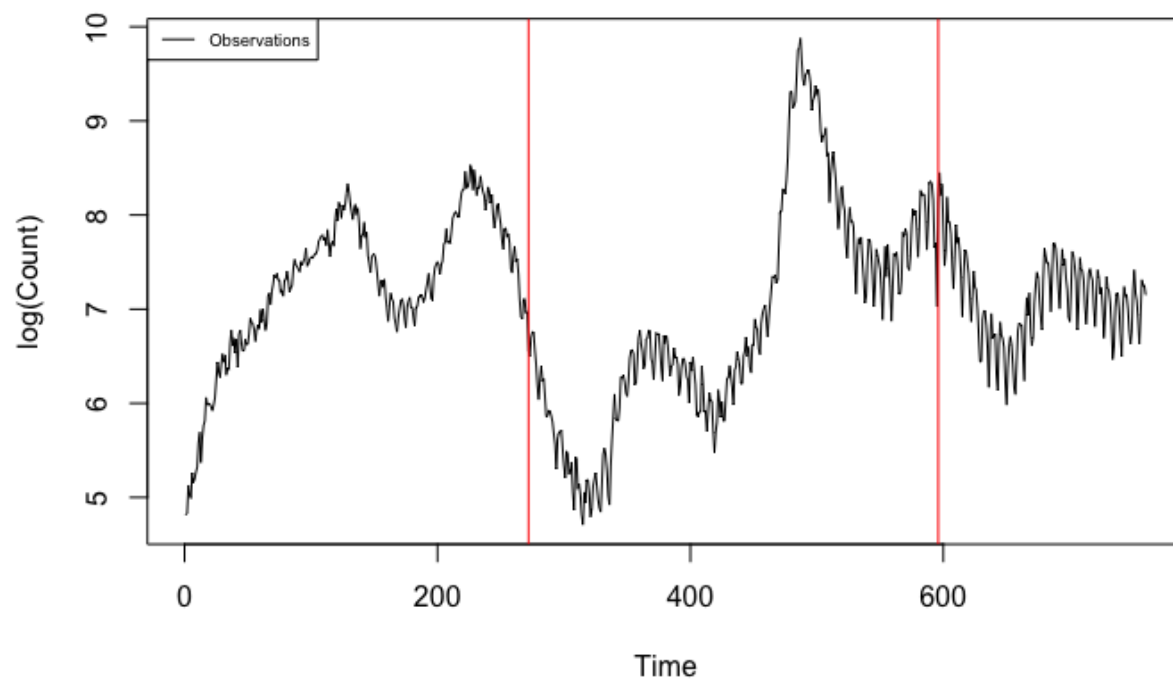


Figure 3.31: Change points identified via τ_t thresholding.

In contrast, the Pruned Exact Linear Time (PELT) and binary segmentation algorithm [58] identifies only one change point, located close to the largest τ_t signal. This suggests that while PELT is effective at detecting significant changes, it may overlook smaller yet important shifts in the data. The detected change point by PELT closely aligns with the most prominent spike in the τ_t values, corroborating the significance of this specific change.

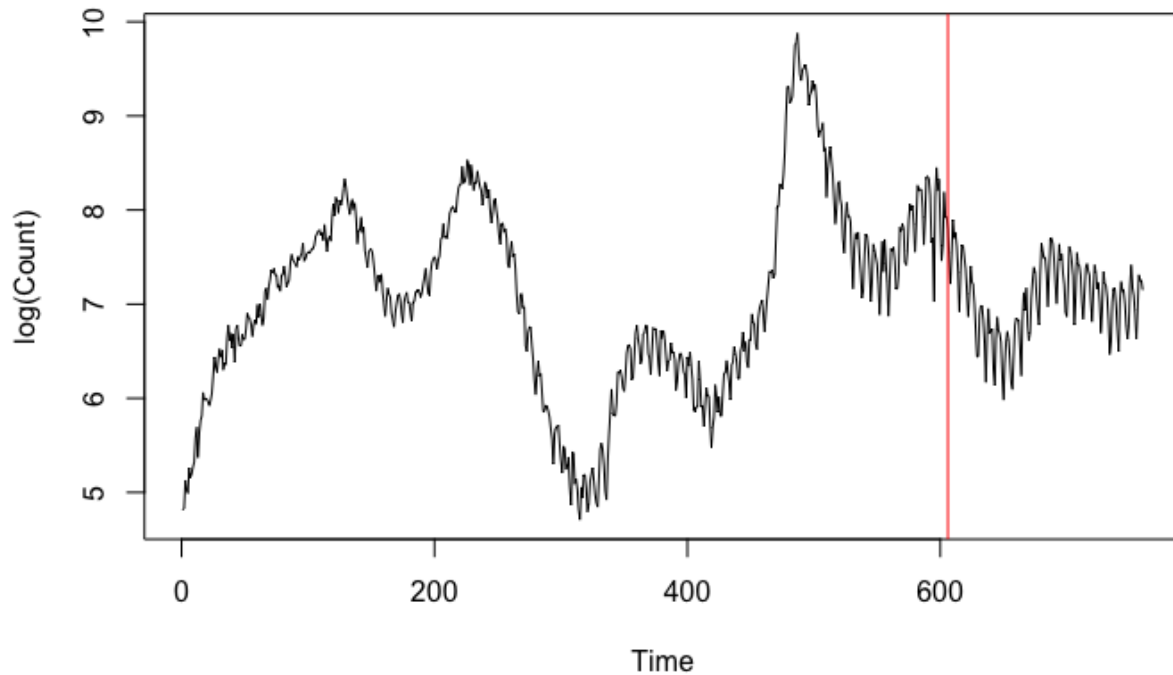


Figure 3.32: The result of change detection using PELT algorithm/binary segmentation.

The Bayesian change point detection method [59] experiences difficulty in clearly identifying change points within the original dataset. The posterior probability exceeds 0.5 in multiple regions, indicating uncertainty and potential over-detection of change points. This method's sensitivity leads to multiple identified regions, making it challenging to pinpoint distinct change points with high confidence.

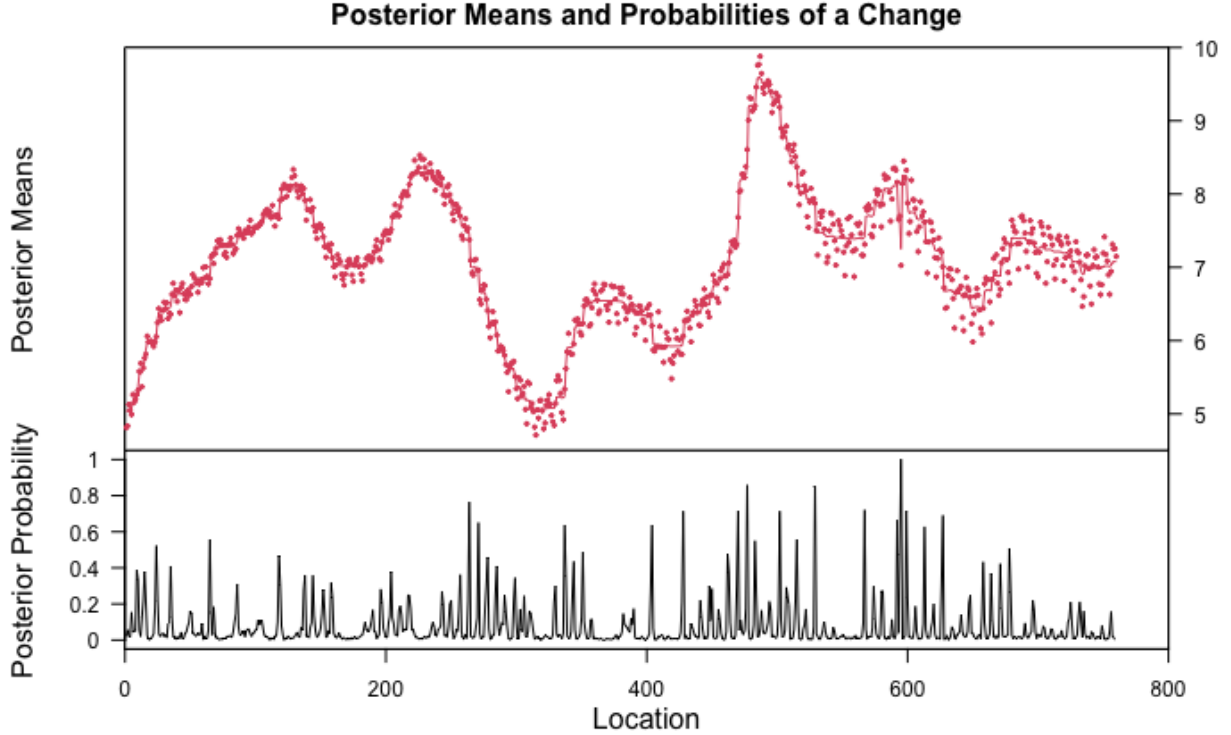


Figure 3.33: Posterior means and probabilities of change points using Bayesian change point method.

Comparing the methods, the horseshoe prior in the Rao-Blackwellized Liu & West filter shows a balanced ability to detect significant change points while maintaining specificity. The PELT algorithm, although robust, may miss smaller changes, focusing on the most pronounced shifts. The Bayesian method's high sensitivity can lead to over-detection, identifying multiple regions of change with lower confidence.

In practical applications, such as analyzing COVID-19 case counts, selecting the appropriate change point detection method is crucial. The prior-based filter provides a clear advantage by identifying significant changes without over-detection, offering a reliable tool for understanding and responding to shifts in epidemic data. This comparison illustrates the importance of method selection based on the data characteristics and the specific needs of the analysis, ensuring accurate and actionable insights.

3.5 Conclusion

The methodology presented here is widely applicable and can be used as part of multivariate time series analysis. For example, the method is adaptable to time-varying coefficient regression prediction for which piecewise linear regression methods have been implemented for the

change detection [60]. In addition, previous works have provided an integrated quantitative approach for data assimilation, prediction, and anomaly detection in real-time public health surveillance [61], emphasizing the development of dynamic probabilistic models to predict future disease cases and detect anomalies using Bayesian methods and advanced information technology. Another line of research has employed Bayesian model averaging with change-point models to estimate the timing and magnitude of vaccine-associated changes, while controlling for seasonality and other covariates [62].

We have demonstrated how dynamic regression models provide a framework for analyzing relationships between time series data and state-space models are able to capture the signal from the data, determine unexpected movement, perform anomaly detection and forecast with uncertainty quantification. Stochastic numerical methods for estimation such as Markov chain Monte Carlo and particle filters may be used for the general state-space model.

Our state-space modelling analysis on Ontario’s COVID-19 surveillance data specifically identified two critical structural change points, corresponding to significant shifts in case count trajectories, using a Rao-Blackwellized particle filter augmented by a horseshoe prior. These change points, occurring during heightened transmission periods, align with key COVID-19 waves observed in Ontario. The model’s sensitivity to abrupt shifts allowed it to accurately capture periods of rapid case increase, which are crucial for timely public health interventions. By integrating dynamic regression models with MCMC and SMC methods, we explored relationships between case counts, hospitalizations, and wastewater signals, yielding a detailed view of how these covariates interact and drive transmission. The horseshoe prior, selected for its ability to shrink small effects while highlighting meaningful changes, effectively isolated high-magnitude changes in case counts, providing reliable indicators of emerging transmission waves. Compared to other priors like normal and Laplace, the horseshoe prior demonstrated superior fit with higher log-likelihood values, reinforcing its utility in epidemic modelling where structural changes are frequent and impactful. These findings suggest that adaptive state-space modelling can enhance surveillance systems by identifying key inflection points in epidemic trends, thereby supporting proactive public health responses in Ontario.

Chapter 4

State-Space Model for Epidemic Disease: Assessing the Impact of Vaccination and Behavioural Effects on COVID-19 Dynamics

4.1 Introduction

A comprehensive framework for modeling disease transmission dynamics is presented, utilizing a state-space model to link a system of ordinary differential equations governing disease evolution to surveillance targets such as case, hospitalization, and death counts. This framework is specifically adapted to COVID-19, incorporating time-varying epidemiological parameters derived from relevant clinical data and flexible non-parametric statistical models.

Key to this adaptation is the modeling of the disease transmission rate, $\beta(t)$, using a log-normal stochastic process. Other epidemiological parameters are captured through Bayesian shrinkage splines, providing computational efficiency and allowing for the inclusion of exogenous covariates such as vaccination rates, waning vaccine effectiveness, and community mobility. The framework also accounts for human demography and incorporates factors like testing capacity, positivity rates, wastewater signals, and is also suited for incorporating environmental factors such as temperature, humidity, and air quality.

To predict surveillance targets, a nonlinear and non-Gaussian dynamic lognormal and gamma mixed negative binomial regression model is employed. This model captures the heterogeneity and contagion effects, recovering the timing and intensity of major epidemic peaks more accurately than standard approaches. The framework addresses the complexity

of COVID-19's spread, which is influenced by factors such as pre-diagnosis infectiousness and asymptomatic cases that often go undiagnosed but still contribute to transmission.

Mathematical analysis of the model yields the basic reproduction number, $\mathcal{R}_0$, using the next-generation matrix technique. The model's properties, including the existence and uniqueness of a non-trivial endemic equilibrium when $\mathcal{R}_0 > 1$ and the stability of the disease-free equilibrium, are thoroughly characterized. Additionally, the model addresses the stochastic nature of real-world data, accounting for variations due to local outbreaks and measurement errors.

Implemented using Hamiltonian Markov chain Monte Carlo (MCMC) sampling via Stan, this probabilistic programming approach facilitates full Bayesian inference on parameters. The model has been applied to data from Ontario, Canada, demonstrating its practical application.

This framework introduces several novel advancements: detailed vaccination analysis, the incorporation of demography, comprehensive mathematical analysis including $\mathcal{R}_0$ derivation and stability characterization, and a sophisticated state-space model framework. These elements collectively enhance the model's ability to assess the impact of human mobility and government policies on the dynamics of COVID-19 and other respiratory diseases, providing a robust tool for public health decision-making.

4.2 Methodology

4.2.1 Compartmental model

The compartmental modeling approach [63]–[65] is widely used in epidemiology to understand the dynamics of diseases, including COVID-19. This model is characterized by a system of differential equations and serves as the foundation for the statistical model.

First, begin with the deterministic compartmental model. The flow chart of the base model is shown below.

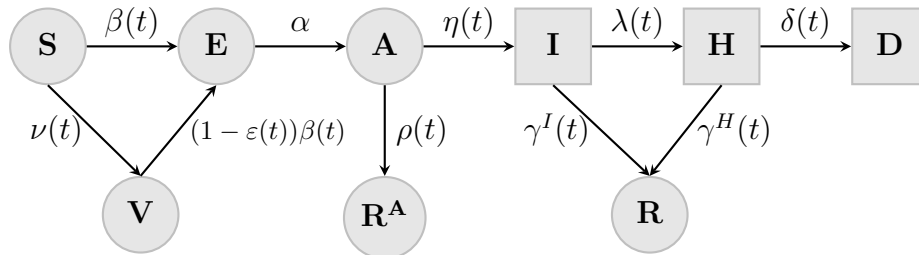


Figure 4.1: Flow chart of the base model.

The parameter $\beta(t)$ is the transmission rate between the asymptomatic infectious and susceptible individuals and it is the number of effective contacts that are sufficient for disease transmission made by an average member of the asymptomatic infectious individuals per unit time. The probability of each of these contacts being with a susceptible individual is $S(t)/N$ and therefore, $A(t)$ asymptomatic infectious individuals lead to a rate of new infections $\beta(t)A(t)S(t)/N$. For the symptomatic infectious individuals, the rate of effective contacts is reduced by a factor of $\tau \in [0, 1]$. Individuals diagnosed with disease are likely to have reduced contact with others due to quarantine or self-isolation, leading to a rate of new infections $\tau\beta(t)I(t)S(t)/N$:

$$\underbrace{\beta(t)A(t)S(t)/N}_{\text{asympt. rate of new infect.}} + \underbrace{\tau\beta(t)I(t)S(t)/N}_{\text{symp. rate of new infect.}}$$

The model incorporates nine compartments, each representing a specific stage of COVID-19:

1. Susceptible individuals (**S**): those who have not been infected by the disease and are at risk of becoming infected;
2. Exposed individuals (**E**): those who have been exposed to the disease but are not yet capable of infecting others;
3. Undetected infectious individuals (**A**): those who have the disease, are able to infect others, but have not yet been detected as having the disease;
4. Detected infectious individuals (**I**): those who have the disease (either with or without symptoms), are able to infect others, and have been detected as having the disease;
5. Individuals removed from **A** without being detected (**R^A**): those who had the disease but are then removed from the possibility of being infected again or spreading the disease for any reason (e.g., significant reduction in viral load, recovery, or death), without ever being diagnosed with the disease;
6. Hospitalized individuals (**H**): those who have been identified as cases and are hospitalized;
7. Individuals removed after having been detected (**R**): those who have been through the **I** state (and possibly **H** state as well) but are then removed from the possibility of being infected again or spreading the disease for any reason except death. They are assumed to eventually recover from the disease;

8. Deceased individuals (**D**): those who have been through the **H** state and are eventually deceased;
9. Vaccinated individuals (**V**): those who have been vaccinated and are thus removed from the susceptible pool.

The transitions among these compartments capture important features of COVID-19 dynamics, including asymptomatic and presymptomatic spread. Due to data limitations, symptomatic and asymptomatic cases are not separately modeled. As time progresses, undetected infectious individuals either become detected or get removed without ever being detected. Detected infectious individuals may be hospitalized or removed without ever being hospitalized. Hospitalized individuals can recover or die, and individuals who die are assumed to have gone through hospitalization.

The SEAIHRDV model equations, with time-varying parameters, are shown below. These differential equations represent the flow of individuals through compartments:

$$\left\{ \begin{array}{l} \frac{dS(t)}{dt} = \Lambda - \frac{(\nu(t) + \beta(t)(A(t) + \tau I(t)))S(t)}{N(t)} - \mu S(t) \\ \frac{dE(t)}{dt} = \frac{\beta(t)(A(t) + \tau I(t))(S(t) + (1 - \varepsilon(t))V(t))}{N(t)} - \alpha E(t) - \mu E(t) \\ \frac{dA(t)}{dt} = \alpha E(t) - (\eta(t) + \rho(t))A(t) - \mu A(t) \\ \frac{dR^A(t)}{dt} = \rho(t)A(t) - \mu R^A(t) \\ \frac{dI(t)}{dt} = \eta(t)A(t) - (\gamma^I(t) + \lambda(t))I(t) - \mu I(t) \\ \frac{dR(t)}{dt} = \gamma^I(t)I(t) + \gamma^H(t)H(t) - \mu R(t) \\ \frac{dH(t)}{dt} = \lambda(t)I(t) - (\gamma^H(t) + \delta(t))H(t) - \mu H(t) \\ \frac{dD(t)}{dt} = \delta(t)H(t) \\ \frac{dV(t)}{dt} = \nu(t)S(t) - \frac{(1 - \varepsilon(t))\beta(t)(A(t) + \tau I(t))V(t)}{N(t)} - \mu V(t) \end{array} \right. \quad \begin{array}{l} (4.2.1a) \\ (4.2.1b) \\ (4.2.1c) \\ (4.2.1d) \\ (4.2.1e) \\ (4.2.1f) \\ (4.2.1g) \\ (4.2.1h) \\ (4.2.1i) \end{array}$$

The time-varying coefficients in these equations add flexibility to capture and adapt to the complex and rapidly changing dynamics of COVID-19. Specifically, $\beta(t)$ represents the disease transmission rate, and $\eta(t)$ represents the detection rate, both of which are expected to vary over time as the pandemic evolves. Other parameters, such as $\lambda(t)$ for hospitalization

rate and $\delta(t)$ for death rate, also change over time to reflect real-world dynamics. The birth rate is denoted by Λ .

The reproductive ratio, $\mathcal{R}_0$, is a crucial metric for predicting the spread of infection. It is computed based on the equilibrium reproduction number, indicating whether the pandemic will persist or subside. If $\mathcal{R}_0(t) < 1$, the number of infectious individuals will decrease, leading to a disease-free state. If $\mathcal{R}_0(t) > 1$, the pandemic will persist with a non-trivial equilibrium.

Including the effect of demography such as births, deaths, immigration, and emigration can be accomplished by modifying the base model. The flow chart of the model with demography is shown below.

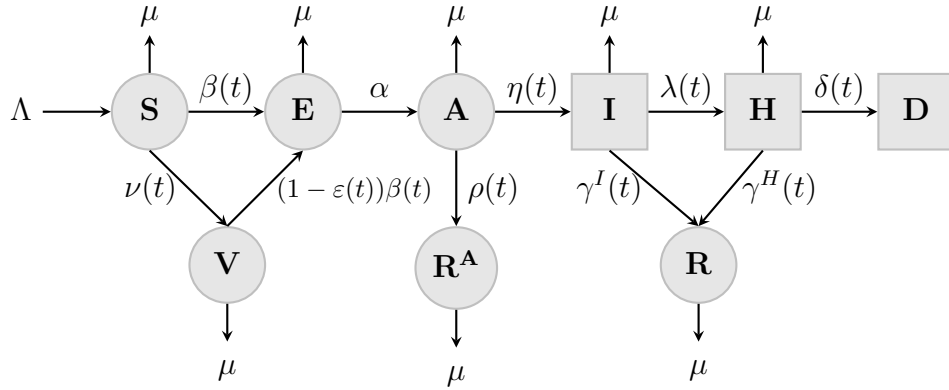


Figure 4.2: Flow chart of the model with demography.

Regarding model assumptions and simplifications, the base model assumes a fast-spreading pandemic, which justifies ignoring births and deaths due to natural causes, as well as immigration and emigration. Reinfection is assumed to be negligible during the study period, meaning that individuals who recover from COVID-19 are considered immune and do not return to the susceptible compartment. The transitions between compartments are assumed to be instantaneous, and the rates of transition are time-varying to capture the dynamic nature of the pandemic.

4.2.2 Mathematical analysis

Here, the following results are provided: determination of disease-free equilibrium and reproduction number for the model, demonstration of non-negative trajectory of the state model compartments, boundedness of the model, existence and uniqueness of non-trivial asymptotic endemic equilibrium, characterization of the stability of the disease-free equilibrium and sensitivity indexes for vaccination parameters.

First, to set-up of $\mathcal{R}_0$ calculation, let the state compartment vector be denoted by

$$\Theta = [S, E, A, R^A, I, H, R, D, V]^\top,$$

the states that represent infected cases be denoted by

$$\mathbf{X} = [E, A, I, H]^\top; \quad \mathbf{X}^0 = [0, 0, 0, 0]^\top,$$

and define the following vectors:

$$F(\mathbf{X}) = [\beta(A + \tau I)(S + (1 - \varepsilon)V)/N, 0, 0, 0]^\top$$

representing the rate of new infections entering each state, and

$$U(\mathbf{X}) = [(\alpha + \mu)E, (\eta + \rho + \mu)A - \alpha E, (\gamma^I + \lambda + \mu)I - \eta A, (\gamma^H + \delta + \mu)H - \lambda I]^\top$$

representing the transfer rate out of each state by way of non-infection.

Then,

$$\frac{d\mathbf{X}(t)}{dt} = F(\mathbf{X}) - U(\mathbf{X}).$$

Assume, for the moment, that the parameters are independent of t . The Jacobian of F and U are given by

$$D_{\mathbf{X}}F = \begin{bmatrix} 0 & \beta(S + (1 - \varepsilon)V)/N & \tau\beta(S + (1 - \varepsilon)V)/N & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$D_{\mathbf{X}}U = \begin{bmatrix} \alpha + \mu & 0 & 0 & 0 \\ -\alpha & \eta + \rho + \mu & 0 & 0 \\ 0 & -\eta & \gamma^I + \lambda + \mu & 0 \\ 0 & 0 & -\lambda & \gamma^H + \delta + \mu \end{bmatrix}.$$

The inverse of Jacobian of U is given by

$$[D_{\mathbf{X}}U]^{-1} = \begin{bmatrix} \frac{1}{\alpha + \mu} & 0 & 0 & 0 \\ \frac{\alpha}{(\alpha + \mu)(\eta + \rho + \mu)} & \frac{1}{\eta + \rho + \mu} & 0 & 0 \\ \frac{\alpha\eta}{(\alpha + \mu)(\gamma^I + \lambda + \mu)(\eta + \rho + \mu)} & \frac{\eta}{(\gamma^I + \lambda + \mu)(\eta + \rho + \mu)} & \frac{1}{\gamma^I + \lambda + \mu} & 0 \\ \frac{\alpha\eta\lambda}{(\alpha + \mu)(\gamma^H + \delta + \mu)(\gamma^I + \lambda + \mu)(\eta + \rho + \mu)} & \frac{\eta\lambda}{(\gamma^H + \delta + \mu)(\gamma^I + \lambda + \mu)(\eta + \rho + \mu)} & \frac{1}{(\gamma^H + \delta + \mu)(\gamma^I + \lambda + \mu)} & \frac{1}{(\gamma^H + \delta + \mu)} \end{bmatrix}$$

Evaluation at DFE is as follows. At the disease-free equilibrium, $\frac{d\theta(t)}{dt} = 0$ and $I = 0$: $\Theta^0 = [S^0 = \frac{\Lambda}{\nu+\mu}, 0, 0, 0, 0, 0, 0, 0, V^0 = \frac{\nu\Lambda}{\mu(\nu+\mu)}]^\top$. Then,

$$\left([D_{\mathbf{X}}F |_{\Theta^0}][D_{\mathbf{X}}U]^{-1}\right)_{11} = \frac{\alpha}{(\alpha+\mu)(\eta+\rho+\mu)}\beta N^{-1}(S^0 + (1-\varepsilon)V^0)\left(1 + \frac{\tau\eta}{\gamma^I + \lambda + \mu}\right)$$

$$\left([D_{\mathbf{X}}F |_{\Theta^0}][D_{\mathbf{X}}U]^{-1}\right)_{12} = \beta N^{-1}(S^0 + (1-\varepsilon)V^0)\left(1 + \frac{\tau\eta}{\gamma^I + \lambda + \mu}\right)$$

$$\left([D_{\mathbf{X}}F |_{\Theta^0}][D_{\mathbf{X}}U]^{-1}\right)_{13} = \frac{\tau\beta N^{-1}(S^0 + (1-\varepsilon)V^0)}{\gamma^I + \lambda + \mu}$$

$$\left([D_{\mathbf{X}}F |_{\Theta^0}][D_{\mathbf{X}}U]^{-1}\right)_{ij} = 0 \text{ for all } ij \neq 11, 12, 13.$$

Eigenvalues of $[D_{\mathbf{X}}F |_{\Theta^0}][D_{\mathbf{X}}U]^{-1}$ are $0, 0, 0$ and $\left([D_{\mathbf{X}}F |_{\Theta^0}][D_{\mathbf{X}}U]^{-1}\right)_{11}$.

The effective reproduction number can be identified with the spectral radius (dominant eigenvalue) of the next-generation matrix $[D_{\mathbf{X}}F |_{\Theta^0}][D_{\mathbf{X}}U]^{-1}$ [66]. That is,

$$\begin{aligned} \mathcal{R}_0 &= \beta \frac{\alpha}{\alpha + \mu} \frac{S^0 + (1 - \varepsilon)V^0}{N} \frac{\gamma^I + \lambda + \tau\eta + \mu}{(\eta + \rho + \mu)(\gamma^I + \lambda + \mu)} \\ &= \beta \frac{\alpha}{\alpha + \mu} \frac{\Lambda(\nu(1 - \varepsilon) + \mu)}{\mu(\nu + \mu)N} \frac{\gamma^I + \lambda + \tau\eta + \mu}{(\eta + \rho + \mu)(\gamma^I + \lambda + \mu)}, \end{aligned} \quad (4.2.2)$$

a measure of the transmission potential of the disease, equivalent to the number of secondary cases or new infections produced by a single infectious individual in a completely susceptible population. An epidemic is expected to increase exponentially if $\mathcal{R}_0 > 1$, and to end if $\mathcal{R}_0 < 1$.

A bound on the impact of vaccination is readily obtained.

$$\begin{aligned} \mathcal{R}_0 &= \frac{\Lambda(\nu(1 - \varepsilon) + \mu)}{\mu(\nu + \mu)N} \times \beta \frac{\alpha}{\alpha + \mu} \frac{\gamma^I + \lambda + \tau\eta + \mu}{(\eta + \rho + \mu)(\gamma^I + \lambda + \mu)} \\ &= f_{\text{vacc}}(\nu, \varepsilon; \mu) \times \mathcal{R}_0^{\text{unvacc}} \end{aligned} \quad (4.2.3)$$

where the prefactor due to vaccination,

$$f_{\text{vacc}}(\nu, \varepsilon; \mu) = \frac{\nu(1 - \varepsilon) + \mu}{\nu + \mu} \quad (4.2.4)$$

satisfies

$$1 \geq f_{\text{vacc}}(\nu, \varepsilon; \mu) \geq \frac{\mu}{\nu + \mu}, \quad (4.2.5)$$

for $\nu, \varepsilon \in [0, 1]$.

If $0 < \varepsilon < 1$ then $\mathcal{R}_0 < \mathcal{R}_0^{\text{unvacc}}$ which analytically implies that a vaccine diminishes the spread of the virus disease.

Now, vaccine effectiveness, containment, and herd immunity are addressed. Consider the scenario whereby $\nu \gg \mu$:

$$f_{\text{vacc}}(\nu, \varepsilon; \mu) = \frac{\nu(1 - \varepsilon) + \mu}{\nu + \mu} = \frac{\frac{\nu}{\mu}(1 - \varepsilon) + 1}{\frac{\nu}{\mu} + 1} \approx 1 - \varepsilon \quad (4.2.6)$$

Then,

$$\mathcal{R}_0 \approx (1 - \varepsilon) \times \mathcal{R}_0^{\text{unvacc}} \quad (4.2.7)$$

The vaccine will stop the spread of the virus if its optimal effectiveness ε^* exceeds herd immunity:

$$\varepsilon^* > 1 - \frac{\mathcal{R}_0}{\mathcal{R}_0^{\text{unvacc}}} \quad (4.2.8)$$

Without a vaccine, the virus will die out when the population reaches herd immunity. To stop the spread of the virus, the optimal vaccine effectiveness must be greater than the herd immunity. The optimal vaccine effectiveness will move the endemic disease equilibrium to the disease-free equilibrium.

To show non-negative trajectory for $t \geq 0$, let k_1 and k_2 be non-negative constants and denote state E, A, R^A, I, H, R , or D of the model as $\theta(t)$. Note the following:

$$\begin{aligned} \frac{d\theta(t)}{dt} &= k_1 - k_2\theta(t) \\ &\geq -k_2\theta(t) \\ \frac{d \log \theta(t)}{dt} &\geq -k_2 \\ \theta(t) &\geq \exp(-k_2 \int dt) \\ &= \theta(0) \exp(-k_2 t) \\ &\geq 0 \end{aligned} \quad (4.2.9)$$

For the demonstration of non-negative trajectory of $\theta = S, V$ for $t \geq 0$, proceed as follows. Let k_1, k_2, k_3 be non-negative constants, $f(A(t))$ be a function of $A(t)$, and $g(I(t))$

be a function of $I(t)$. Then,

$$\begin{aligned}
\frac{d\theta(t)}{dt} &= k_1 - (f(A(t)) + g(I(t)))\theta(t) \\
&\geq -(f(A(t)) + g(I(t)))\theta(t) \\
&\geq -(k_3 \max_t |A(t)| + k_4 \max_t |I(t)|)\theta(t) \\
\frac{d \log \theta(t)}{dt} &\geq -(k_3 \max_t |A(t)| + k_4 \max_t |I(t)|) \\
\theta(t) &\geq \theta(0) \exp(-(k_3 \max_t |A(t)| + k_4 \max_t |I(t)|)t) \\
&\geq 0
\end{aligned} \tag{4.2.10}$$

To show boundedness of the model, observe that the growth rate of the total population is the sum of the derivatives of the state compartment components:

$$\frac{dN(t)}{dt} = \sum \frac{d\theta(t)}{dt} = \Lambda - \mu N(t).$$

Therefore,

$$N(t) = \frac{\Lambda}{\mu} + \left(N(0) - \frac{\Lambda}{\mu} \exp(-\mu t) \right) \leq \frac{\Lambda}{\mu}$$

as $t \rightarrow \infty$. Thus,

$$0 \leq N(t) \leq \frac{\Lambda}{\mu} \tag{4.2.11}$$

and the compartment components remain bounded for all $t > 0$.

Calculation of the endemic equilibrium is obtained by first setting each of the state derivatives to zero, which gives:

$$\left\{ \begin{array}{l} S(t) = \frac{\Lambda}{\beta(t)N^{-1}(t)(A(t) + \tau I(t)) + \nu(t) + \mu} \quad (4.2.12a) \\ 0 = \beta(t)N^{-1}(t)(A(t) + \tau I(t))(S(t) + (1 - \varepsilon(t))V(t)) - (\alpha + \mu)E(t) \quad (4.2.12b) \\ E(t) = \frac{\eta(t) + \rho(t) + \mu}{\alpha}A(t) \quad (4.2.12c) \\ R^A(t) = \frac{\rho(t)}{\mu}A(t) \quad (4.2.12d) \\ A(t) = \frac{\gamma^I(t) + \lambda(t) + \mu}{\eta(t)}I(t) \quad (4.2.12e) \\ R(t) = \frac{\gamma^I(t)I(t) + \gamma^H(t)H(t)}{\mu} \quad (4.2.12f) \\ H(t) = \frac{\lambda(t)}{\gamma^H(t) + \delta(t) + \mu}I(t) \quad (4.2.12g) \\ V(t) = \frac{\nu(t)S(t)}{(1 - \varepsilon(t))\beta(t)N^{-1}(t)(A(t) + \tau I(t)) + \mu} \quad (4.2.12h) \end{array} \right.$$

The equation for the $E(t)$ derivative becomes:

$$\begin{aligned} 0 &= \beta(t)N^{-1}(t)(A(t) + \tau I(t))(S(t) + (1 - \varepsilon(t))V(t)) - (\alpha + \mu)E(t) \\ &= \beta(t)N^{-1}(t)(A(t) + \tau I(t)) \left(S(t) + \frac{(1 - \varepsilon(t))\nu(t)S(t)}{(1 - \varepsilon(t))\beta(t)N^{-1}(t)(A(t) + \tau I(t)) + \mu} \right) - (\alpha + \mu)E(t) \\ &= \beta(t)N^{-1}(t)(A(t) + \tau I(t)) \left(\frac{(1 - \varepsilon(t))[\beta(t)N^{-1}(t)(A(t) + \tau I(t)) + \nu(t)] + \mu}{(1 - \varepsilon(t))\beta(t)N^{-1}(t)(A(t) + \tau I(t)) + \mu} \right) \times \\ &\quad \times \left(\frac{\Lambda}{\beta(t)N^{-1}(t)(A(t) + \tau I(t)) + \nu(t) + \mu} \right) - \frac{(\alpha + \mu)(\eta(t) + \rho(t) + \mu)(\gamma^I(t) + \lambda(t) + \mu)}{\alpha\eta(t)}I(t) \\ &= \Lambda\beta(t)N^{-1}(t)(A(t) + \tau I(t)) \left((1 - \varepsilon(t))[\beta(t)N^{-1}(t)(A(t) + \tau I(t)) + \nu(t)] + \mu \right) - \\ &\quad - \frac{(\alpha + \mu)(\eta(t) + \rho(t) + \mu)(\gamma^I(t) + \lambda(t) + \mu)}{\alpha\eta(t)} \times \\ &\quad \times ((1 - \varepsilon(t))\beta(t)N^{-1}(t)(A(t) + \tau I(t)) + \mu)(\beta(t)N^{-1}(t)(A(t) + \tau I(t)) + \nu(t) + \mu)I(t) \end{aligned} \quad (4.2.13)$$

$$\begin{aligned}
0 &= \Lambda\beta(t)N^{-1}(t)(A(t) + \tau I(t)) \left((1 - \varepsilon(t)) [\beta(t)N^{-1}(t)(A(t) + \tau I(t)) + \nu(t)] + \mu \right) \\
&\quad - \frac{(\alpha + \mu)(\eta(t) + \rho(t) + \mu)(\gamma^I(t) + \lambda(t) + \mu)}{\alpha\eta(t)} \\
&\quad \times ((1 - \varepsilon(t))\beta(t)N^{-1}(t)(A(t) + \tau I(t)) + \mu) \\
&\quad \times (\beta(t)N^{-1}(t)(A(t) + \tau I(t)) + \nu(t) + \mu)I(t)
\end{aligned} \tag{4.2.14}$$

$$\begin{aligned}
0 &= \Lambda\beta(t)N^{-1}(t) \frac{\gamma^I(t) + \lambda(t) + \mu + \tau\eta(t)}{\eta(t)} I(t) \\
&\quad \times \left((1 - \varepsilon(t)) [\beta(t)N^{-1}(t) \frac{\gamma^I(t) + \lambda(t) + \mu + \tau\eta(t)}{\eta(t)} I(t) + \nu(t)] + \mu \right) \\
&\quad - \frac{(\alpha + \mu)(\eta(t) + \rho(t) + \mu)(\gamma^I(t) + \lambda(t) + \mu)}{\alpha\eta(t)} \\
&\quad \times \left((1 - \varepsilon(t))\beta(t)N^{-1}(t) \frac{\gamma^I(t) + \lambda(t) + \mu + \tau\eta(t)}{\eta(t)} I(t) + \mu \right) \\
&\quad \times \left(\beta(t)N^{-1}(t) \frac{\gamma^I(t) + \lambda(t) + \mu + \tau\eta(t)}{\eta(t)} I(t) + \nu(t) + \mu \right) I(t)
\end{aligned} \tag{4.2.15}$$

$$\begin{aligned}
0 &= \left(\beta(t)N^{-1}(t) \frac{\gamma^I(t) + \lambda(t) + \mu + \tau\eta(t)}{\eta(t)} \right)^2 \\
&\quad \times \frac{(\alpha + \mu)(\eta(t) + \rho(t) + \mu)(\gamma^I(t) + \lambda(t) + \mu)}{\alpha\eta(t)} (1 - \varepsilon(t)) I^2(t) \\
&\quad + \left(\beta(t)N^{-1}(t) \frac{\gamma^I(t) + \lambda(t) + \mu + \tau\eta(t)}{\eta(t)} \right. \\
&\quad \times \frac{(\alpha + \mu)(\eta(t) + \rho(t) + \mu)(\gamma^I(t) + \lambda(t) + \mu)}{\alpha\eta(t)} (1 - \varepsilon(t)) (\nu(t) + \mu) \\
&\quad + \beta(t)N^{-1}(t) \frac{\gamma^I(t) + \lambda(t) + \mu + \tau\eta(t)}{\eta(t)} \\
&\quad \times \frac{(\alpha + \mu)(\eta(t) + \rho(t) + \mu)(\gamma^I(t) + \lambda(t) + \mu)}{\alpha\eta(t)} \mu \\
&\quad \left. - \left(\beta(t)N^{-1}(t) \frac{\gamma^I(t) + \lambda(t) + \mu + \tau\eta(t)}{\eta(t)} \right)^2 \Lambda(1 - \varepsilon(t)) \right) I(t) \\
&\quad + \frac{(\alpha + \mu)(\eta(t) + \rho(t) + \mu)(\gamma^I(t) + \lambda(t) + \mu)}{\alpha\eta(t)} \mu (\nu(t) + \mu) \\
&\quad - \beta(t)N^{-1}(t) \frac{\gamma^I(t) + \lambda(t) + \mu + \tau\eta(t)}{\eta(t)} \Lambda((1 - \varepsilon(t))\nu(t) + \mu)
\end{aligned} \tag{4.2.16}$$

The roots of the infectious compartment I are determined as follows.

$$\begin{aligned}
0 = & \left(\beta(t)N^{-1}(t) \frac{\gamma^I(t) + \lambda(t) + \mu + \tau\eta(t)}{\eta(t)} \right)^2 \frac{(\alpha + \mu)(\eta(t) + \rho(t) + \mu)(\gamma^I(t) + \lambda(t) + \mu)}{\alpha\eta(t)} (1 - \varepsilon(t)) I^2(t) \\
& + \left(\beta(t)N^{-1}(t) \frac{\gamma^I(t) + \lambda(t) + \mu + \tau\eta(t)}{\eta(t)} \frac{(\alpha + \mu)(\eta(t) + \rho(t) + \mu)(\gamma^I(t) + \lambda(t) + \mu)}{\alpha\eta(t)} (1 - \varepsilon(t)) (\nu(t) + \mu) \right. \\
& + \beta(t)N^{-1}(t) \frac{\gamma^I(t) + \lambda(t) + \mu + \tau\eta(t)}{\eta(t)} \frac{(\alpha + \mu)(\eta(t) + \rho(t) + \mu)(\gamma^I(t) + \lambda(t) + \mu)}{\alpha\eta(t)} \mu \\
& - \left. \left(\beta(t)N^{-1}(t) \frac{\gamma^I(t) + \lambda(t) + \mu + \tau\eta(t)}{\eta(t)} \right)^2 \Lambda(1 - \varepsilon(t)) \right) I(t) \\
& + \frac{(\alpha + \mu)(\eta(t) + \rho(t) + \mu)(\gamma^I(t) + \lambda(t) + \mu)}{\alpha\eta(t)} \mu(\nu(t) + \mu) \\
& - \beta(t)N^{-1}(t) \frac{\gamma^I(t) + \lambda(t) + \mu + \tau\eta(t)}{\eta(t)} \Lambda((1 - \varepsilon(t))\nu(t) + \mu) \tag{4.2.17}
\end{aligned}$$

By Vieta's formula, the roots of I , I_1 and I_2 , have opposite signs satisfying:

$$\begin{aligned}
0 > & \frac{\left(\frac{(\alpha + \mu)(\eta(t) + \rho(t) + \mu)(\gamma^I(t) + \lambda(t) + \mu)}{\alpha\eta(t)} \right) \mu(\nu(t) + \mu) - \left(\beta(t)N^{-1}(t) \frac{\gamma^I(t) + \lambda(t) + \mu + \tau\eta(t)}{\eta(t)} \right) \Lambda((1 - \varepsilon(t))\nu(t) + \mu)}{\left(\beta(t)N^{-1}(t) \frac{\gamma^I(t) + \lambda(t) + \mu + \tau\eta(t)}{\eta(t)} \right)^2 \frac{(\alpha + \mu)(\eta(t) + \rho(t) + \mu)(\gamma^I(t) + \lambda(t) + \mu)}{\alpha\eta(t)} (1 - \varepsilon(t))} \tag{4.2.18}
\end{aligned}$$

To show unique endemic equilibrium, observe the following:

$$\begin{aligned}
& \left(\beta(t)N^{-1}(t) \frac{\gamma^I(t) + \lambda(t) + \mu + \tau\eta(t)}{\eta(t)} \right) \Lambda((1 - \varepsilon(t))\nu(t) + \mu) > \\
& \left(\frac{(\alpha + \mu)(\eta(t) + \rho(t) + \mu)(\gamma^I(t) + \lambda(t) + \mu)}{\alpha\eta(t)} \right) \mu(\nu(t) + \mu) \\
\implies & \beta(t) \frac{\alpha}{\alpha + \mu} \frac{\Lambda(\nu(t)(1 - \varepsilon(t)) + \mu)}{\mu(\nu(t) + \mu)N} \frac{\gamma^I(t) + \lambda(t) + \tau\eta(t) + \mu}{(\eta(t) + \rho(t) + \mu)(\gamma^I(t) + \lambda(t) + \mu)} > 1 \tag{4.2.19}
\end{aligned}$$

Thus, if $\mathcal{R}_0 > 1$, solutions have opposite signs with the negative solution inadmissible due to the positive restriction on the trajectories. Therefore, there is a unique endemic equilibrium.

For stability analysis, consider the Jacobian at the disease-free equilibrium (DFE):

$$\mathbf{J}(\Theta^0) = \begin{bmatrix} -(\nu + \mu) & 0 & -\beta N^{-1} S^0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -(\alpha + \mu) & \beta N^{-1} (S^0 + (1 - \varepsilon) V^0) & 0 & \tau \beta N^{-1} (S^0 + (1 - \varepsilon) V^0) & 0 & 0 & 0 & 0 \\ 0 & \alpha & -(\eta + \rho + \mu) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \rho & -\mu & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \eta & 0 & -(\gamma^I + \lambda + \mu) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda & -(\gamma^H + \delta + \mu) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \gamma^I & \gamma^H & -\mu & 0 & 0 \\ \nu & 0 & -(1 - \varepsilon) - \beta N^{-1} V^0 & 0 & -(1 - \varepsilon) - \beta N^{-1} \tau V^0 & 0 & 0 & 0 & -\mu \end{bmatrix}$$

$$p(r) = \det(\mathbf{J}(\Theta^0) - rI) = -(\gamma^H + \delta + \mu + r) \left[-\mu^3 - (3r + \eta + \rho + \alpha + \gamma^I + \lambda)\mu^2 + \left(-3r^2 - 2(\eta + \rho + \alpha + \gamma^I + \lambda)r + \left(-\eta - \gamma^I + \beta N^{-1}((1 - \varepsilon)V^0 + S^0) - \rho - \lambda \right) \alpha - (\gamma^I + \lambda)(\eta + \rho) \right) \mu - r^3 - (\eta + \rho + \gamma^I + \alpha + \lambda)r^2 + \left(\left(-\eta - \gamma^I + \beta N^{-1}((1 - \varepsilon)V^0 + S^0) - \rho - \lambda \right) \alpha - (\gamma^I + \lambda)(\eta + \rho) \right) r + \alpha \left((-\gamma^I - \lambda + (\tau(1 - \varepsilon)V^0 + \tau S^0)\beta N^{-1})\eta + (\gamma^I + \lambda)(-\rho + \beta N^{-1}((1 - \varepsilon)V^0 + S^0)) \right) \right] (r + \eta + \mu)(r + \mu)^3$$

The roots include $r = -(\gamma^H + \delta + \mu)$, $-(\eta + \mu)$, and $-\mu$ (multiplicity of 3) which are all negative.

For the expanded cubic term of the form $q(r) := r^3 + a_1 r^2 + a_2 r + a_3 = 0$, Routh-Hurwitz criterion gives the requirement for determining whether all roots have negative real part required for asymptotic stability at an equilibrium as $a_3 > 0, a_1 > 0, a_1 a_2 > a_3$.

For the stability analysis, a cubic characteristic equation is encountered.

$$\begin{aligned}
0 &= q(r) \\
&= -\mu^3 - (3r + \eta + \rho + \alpha + \gamma^I + \lambda)\mu^2 + \left(-3r^2 - 2(\eta + \rho + \alpha + \gamma^I + \lambda)r + \right. \\
&\quad \left(-\eta - \gamma^I + \beta N^{-1}((1 - \varepsilon)V^0 + S^0) - \rho - \lambda \right) \alpha - (\gamma^I + \lambda)(\eta + \rho) \Big) \mu - r^3 \\
&\quad - (\eta + \rho + \gamma^I + \alpha + \lambda)r^2 + \left(\left(-\eta - \gamma^I + \beta N^{-1}((1 - \varepsilon)V^0 + S^0) - \rho - \lambda \right) \alpha - (\gamma^I + \lambda)(\eta + \rho) \right) r + \\
&\quad \alpha \left((-\gamma^I - \lambda + (\tau(1 - \varepsilon)V^0 + \tau S^0)\beta N^{-1})\eta + (\gamma^I + \lambda) \left(-\rho + \beta N^{-1}((1 - \varepsilon)V^0 + S^0) \right) \right) \\
&= -r^3 - \left(\alpha + \gamma^I + \lambda + 3\mu + \eta + \rho \right) r^2 \\
&\quad + \left(\alpha \beta N^{-1}(1 - \varepsilon)V^0 + \alpha \beta N^{-1}S^0 - \alpha(\gamma^I + \lambda + 2\mu + \eta + \rho) - \lambda(2\mu + \eta + \rho) - \mu(3\mu + 2\eta + 2\rho) \right) r \\
&\quad + \left(\alpha \beta N^{-1}((1 - \varepsilon)V^0 + S^0)(\eta\tau + \gamma^I + \lambda + \mu) - (\alpha + \mu)(\mu + \eta + \rho)(\gamma^I + \lambda + \mu) \right) \\
&= r^3 + \left(\alpha + \gamma^I + \lambda + 3\mu + \eta + \rho \right) r^2 \\
&\quad + \left(\alpha(\gamma^I + \lambda + 2\mu + \eta + \rho) + \lambda(2\mu + \eta + \rho) + \mu(3\mu + 2\eta + 2\rho) - \alpha \beta N^{-1}((1 - \varepsilon)V^0 + S^0) \right) r \\
&\quad + \left((\alpha + \mu)(\mu + \eta + \rho)(\gamma^I + \lambda + \mu) - \alpha \beta N^{-1}((1 - \varepsilon)V^0 + S^0)(\eta\tau + \gamma^I + \lambda + \mu) \right)
\end{aligned} \tag{4.2.20}$$

The Routh-Hurwitz criteria on a_3 and a_1 is as follows.

$$\begin{aligned}
0 &= q(r) \\
&= r^3 + \left(\alpha + \gamma^I + \lambda + 3\mu + \eta + \rho \right) r^2 \\
&\quad + \left(\alpha(\gamma^I + \lambda + 2\mu + \eta + \rho) + \lambda(2\mu + \eta + \rho) + \mu(3\mu + 2\eta + 2\rho) - \alpha\beta N^{-1}((1 - \varepsilon)V^0 + S^0) \right) r \\
&\quad + \left((\alpha + \mu)(\mu + \eta + \rho)(\gamma^I + \lambda + \mu) - \alpha\beta N^{-1}((1 - \varepsilon)V^0 + S^0)(\eta\tau + \gamma^I + \lambda + \mu) \right)
\end{aligned} \tag{4.2.21}$$

$$\begin{aligned}
0 &< a_3 \\
&= (\alpha + \mu)(\mu + \eta + \rho)(\gamma^I + \lambda + \mu) - \alpha\beta N^{-1}((1 - \varepsilon)V^0 + S^0)(\eta\tau + \gamma^I + \lambda + \mu) \\
1 &> \frac{\alpha\beta N^{-1}((1 - \varepsilon)V^0 + S^0)(\eta\tau + \gamma^I + \lambda + \mu)}{(\alpha + \mu)(\mu + \eta + \rho)(\gamma^I + \lambda + \mu)} \\
\Rightarrow 1 &> \mathcal{R}_0
\end{aligned} \tag{4.2.22}$$

$$\begin{aligned}
0 &< a_1 \\
&= \alpha + \gamma^I + \lambda + 3\mu + \eta + \rho
\end{aligned} \tag{4.2.23}$$

The Routh-Hurwitz criterion, $a_1 a_2 > a_3$ is satisfied, as shown in the following calculation.

$$\begin{aligned}
a_1 a_2 - a_3 &= \left(\alpha + \gamma^I + \lambda + 3\mu + \eta + \rho \right) \times \\
&\quad \left(\alpha(\gamma^I + \lambda + 2\mu + \eta + \rho) + \lambda(2\mu + \eta + \rho) + \mu(3\mu + 2\eta + 2\rho) - \alpha\beta N^{-1}((1 - \varepsilon)V^0 + S^0) \right) \\
&\quad - (\alpha + \mu)(\mu + \eta + \rho)(\gamma^I + \lambda + \mu) + \alpha\beta N^{-1}((1 - \varepsilon)V^0 + S^0)(\eta\tau + \gamma^I + \lambda + \mu) \\
&= \left(\alpha + \gamma^I + \lambda + 3\mu + \eta + \rho \right) \times \\
&\quad \left(\alpha(\gamma^I + \lambda + 2\mu + \eta + \rho) + \lambda(2\mu + \eta + \rho) + \mu(3\mu + 2\eta + 2\rho) - \alpha\beta \frac{\Lambda}{\mu N} f_{\text{vacc}} \right) \\
&\quad - (\alpha + \mu)(\mu + \eta + \rho)(\gamma^I + \lambda + \mu) + \alpha\beta(\eta\tau + \gamma^I + \lambda + \mu) \frac{\Lambda}{\mu N} f_{\text{vacc}} \\
&= \left(\alpha + \gamma^I + \lambda + 3\mu + \eta + \rho \right) \times \\
&\quad \left(\alpha(\gamma^I + \lambda + 2\mu + \eta + \rho) + \lambda(2\mu + \eta + \rho) + \mu(3\mu + 2\eta + 2\rho) \right. \\
&\quad \left. - (\alpha + \mu) \frac{(\eta + \rho + \mu)(\gamma^I + \lambda + \mu)}{(\gamma^I + \lambda + \tau\eta + \mu)} \mathcal{R}_0 \right) + (\alpha + \mu)(\mu + \eta + \rho)(\gamma^I + \lambda + \mu)(\mathcal{R}_0 - 1) \\
&= \left(\alpha + \gamma^I + \lambda + 3\mu + \eta + \rho \right) \times \\
&\quad \left(\alpha(\gamma^I + \lambda + 2\mu + \eta + \rho) + \lambda(2\mu + \eta + \rho) + \mu(3\mu + 2\eta + 2\rho) \right) \\
&\quad - (\alpha + \mu)(\mu + \eta + \rho)(\gamma^I + \lambda + \mu) \\
&\quad + (\alpha + \mu)(\eta + \rho + \mu)(\gamma^I + \lambda + \mu) \left(1 - \frac{\alpha + \gamma^I + \lambda + 3\mu + \eta + \rho}{\gamma^I + \lambda + \tau\eta + \mu} \right) \mathcal{R}_0 \tag{4.2.24}
\end{aligned}$$

Local asymptotic stability of DFE is considered next.

$$\begin{aligned}
a_1 a_2 - a_3 &= \left(\alpha + \gamma^I + \lambda + 3\mu + \eta + \rho \right) \times \\
&\quad \left(\alpha(\gamma^I + \lambda + 2\mu + \eta + \rho) + \lambda(2\mu + \eta + \rho) + \mu(3\mu + 2\eta + 2\rho) \right) \\
&\quad + (\alpha + \mu)(\mu + \eta + \rho)(\gamma^I + \lambda + \mu) \times \\
&\quad \left[\left(1 - \frac{\alpha + \gamma^I + \lambda + 3\mu + \eta + \rho}{\gamma^I + \lambda + \tau\eta + \mu} \right) \mathcal{R}_0 - 1 \right] \tag{4.2.25}
\end{aligned}$$

To satisfy the criterion, it is required that:

$$0 < \left(1 - \frac{\alpha + \gamma^I + \lambda + 3\mu + \eta + \rho}{\gamma^I + \lambda + \tau\eta + \mu}\right) \mathcal{R}_0 - 1$$

$$\iff \mathcal{R}_0 > \frac{1}{1 - \frac{\alpha + \gamma^I + \lambda + 3\mu + \eta + \rho}{\gamma^I + \lambda + \tau\eta + \mu}} = -\frac{\gamma^I + \lambda + \tau\eta + \mu}{(1 - \tau)\eta + \alpha + \rho + 2\mu} \quad (4.2.26)$$

Since $\mathcal{R}_0 > 0$, the Routh-Hurwitz criteria are satisfied if $\mathcal{R}_0 < 1$ and in this situation, the disease-free equilibrium of the system is locally asymptotically stable.

The concept of a normalized forward sensitivity index is a key tool in sensitivity analysis, used to quantify how a model output, such as the basic reproduction number $\mathcal{R}_0$, changes in response to variations in a parameter ψ . Normalization involves scaling the sensitivity index by multiplying by the parameter and dividing by the output, allowing for meaningful comparisons across parameters with differing magnitudes.

In this context, ψ represents a generic parameter of the model, such as ν (the vaccination rate) or ε (the vaccine efficacy), whose sensitivity we aim to analyze. The normalized forward sensitivity index for ψ , denoted by $\Sigma_V(\psi)$, measures the relative rate of change of the model output $\mathcal{R}_0$ with respect to changes in ψ . Specifically, $\Sigma_V(\psi)$ is defined as:

$$\Sigma_V(\psi) = \psi(t) \frac{\partial \log \mathcal{R}_0(t)}{\partial \psi(t)}.$$

This definition is derived directly from the general sensitivity analysis relationship:

$$\frac{\partial \log \mathcal{R}_0}{\partial \psi} = \frac{1}{\mathcal{R}_0} \frac{\partial \mathcal{R}_0}{\partial \psi}.$$

By multiplying through by ψ , the sensitivity index is normalized to enable relative comparisons.

The term $\Sigma(\psi)$ represents additional contributions to the sensitivity index that may arise from other components of the system not explicitly detailed in the simplified equations. These contributions reflect the structural complexity of the model.

This extends the normalized forward sensitivity index to account for the vaccination-related term $f_{\text{vacc}}(\psi)$, which represents the influence of parameters ν and ε on $\mathcal{R}_0$. Substituting $f_{\text{vacc}}(\psi)$ into the general sensitivity formula provides a more specific expression that

captures the relationship between $\mathcal{R}_0$ and ψ . The term $f_{\text{vacc}}(\psi)$ is defined as:

$$f_{\text{vacc}}(\psi = (\nu, \varepsilon); \mu) = \frac{\nu(1 - \varepsilon) + \mu}{\nu + \mu}.$$

The sensitivity index for specific parameters, such as ν and ε , can then be derived from the general formula. For the parameter ν , the sensitivity index is:

$$\Sigma_V(\nu) = \frac{\nu\mu(\nu - \varepsilon + \mu)}{(\nu(1 - \varepsilon) + \mu)(\nu + \mu)}.$$

This is obtained by substituting $f_{\text{vacc}}(\nu, \varepsilon)$ into the sensitivity formula, taking the derivative of f_{vacc} with respect to ν , and simplifying the resulting expression. Similarly, for ε , the sensitivity index is given by:

$$\Sigma_V(\varepsilon) = \frac{\varepsilon(\mu - \nu)}{\nu(1 - \varepsilon) + \mu}.$$

This expression is derived by taking the derivative of f_{vacc} with respect to ε and simplifying.

These equations provide specific measures of how the basic reproduction number $\mathcal{R}_0$ is influenced by the parameters ν and ε , normalized for relative comparison. The derivation of these results from the general definition highlights the importance of normalized sensitivity analysis in understanding the role of different parameters in complex models.

To summarize, ψ is a parameter of interest, such as ν or ε , $\Sigma_V(\psi)$ is the normalized forward sensitivity index quantifying the relative impact of ψ on $\mathcal{R}_0$, and $\Sigma(\psi)$ represents additional contributions arising from the model's structure. The definition of the normalized forward sensitivity index is derived from general principles of sensitivity analysis. The third equation, for ν , is obtained by substituting $f_{\text{vacc}}(\nu, \varepsilon)$ into the sensitivity formula and computing the derivative with respect to ν .

4.2.3 Probabilistic observation model

Incorporating count data into the model is a crucial step in elevating the deterministic compartmental model into a data-driven engine. This approach enhances the deterministic compartmental model by accounting for the stochastic nature of disease spread and measurement errors in the data. It allows for capturing the dynamics and transitions among compartments [32], [33], [67], which include key features of COVID-19 dynamics such as asymptomatic and presymptomatic spread.

Desirable properties of the data model include heterogeneity (differences between individuals) and contagion (dependence between the occurrence of events). Heterogeneity arises due to genuine ambiguity in the relationship between the underlying stochastic process and

the count data distribution if the population is heterogeneous rather than homogeneous. The probability of an occurrence becomes itself a random variable.

Contagion, or state dependence, exists when previous outcomes influence the current experiment. Positive contagion indicates that the occurrence of an event makes further occurrences more likely; for negative contagion, the opposite holds. Both cases lead to a contagious distribution for the number of counts, the Poisson distribution being an example of a non-contagious distribution. Further details can be found in [68].

As an illustration of types of contagion, consider the following, suppose a red balls and b blue balls are placed into an urn. One ball is drawn at random. If it is a red ball (success), it is replaced together with s red balls. If it is a blue ball, the proportion $\frac{a}{a+b}$ is unchanged, and the blue ball is replaced. Repeat this procedure n times and let X represent the total number of red balls drawn. Then $X \sim \text{Pólya-Eggenberger}$. Occurrence dependence may develop as follows. If the number of red balls is increased after a success ($s > 0$), then an occurrence increases the probability of further occurrences, and the urn model reflects positive contagion. It can be shown that the negative binomial distribution is obtained as a limiting form; for $s = 0$, the model reduces to the binomial model with independent trials, and for $s = -1$, the urn scheme corresponds to drawing without replacement, leading to a hypergeometric distribution, a distribution for negative contagion.

There also exists a spectrum of dependence relation, as pure birth process

$$P\{N(0, t + \Delta) = k + 1 \mid N(0, t) = k\} = \lambda_k \Delta + o(\Delta)$$

generalizes the applicability to continuous time. As a function of k , λ_k increasing implies positive occurrence dependence, linear increasing implies negative binomial model, constant gives the Poisson case (no occurrence dependence), and decreasing indicates negative occurrence dependence.

Duration dependence and non-stationarity can also arise. In the urn model for occurrence dependence, the composition of the urn was left unchanged when a blue ball (failure) occurred. If failures matter, then the outcome of an experiment depends on the time (number of draws) that has elapsed since the last success, generating “duration dependence”. Duration dependence can be analyzed either in discrete time, as represented by the urn model, or in continuous time using the concept of (continuous) waiting times. The composition of the urn may change over consecutive trials due to exogenous effects, while being unaffected by previous trials. This is the case if the underlying process is non-stationary, which does not necessarily invalidate the Poisson distribution.

Next, a non-linear and non-Gaussian dynamic regression for count data is provided. The

dynamic random effect Poisson model serves as the basis for the more elaborate dynamic hierarchical negative binomial dual random effect model.

For $t = 1, \dots, T$, let $y(t)$ denote the count time series of interest and let $x(t) = \tilde{\theta}(t)$ denote the corresponding solution of the compartmental model. Consider models for counts with non-negative over-dispersion multiplicative random effect $\epsilon(t)$ to account for individual heterogeneity and contagion dependence [69]:

$$\begin{aligned} y(t) &\sim \text{Poisson}(\theta(t)), \\ \theta(t) &= \exp([x(t)]^\top \beta) \epsilon(t). \end{aligned} \quad (4.2.27)$$

To form the dynamic random effect Poisson model, distributional priors are placed on the multiplicative effect, which yield compound models with

$$\mathbb{E}[y(t) \mid x(t)] = \theta(t) \mathbb{E}[\epsilon(t)], \quad (4.2.28)$$

$$\text{Var}[y(t) \mid x(t)] = \mathbb{E}[y(t) \mid x(t)] + \frac{\text{Var}[\epsilon(t)]}{\mathbb{E}^2[\epsilon(t)]} \mathbb{E}^2[y(t) \mid x(t)]. \quad (4.2.29)$$

The compound Gamma-Poisson model for counts follows by setting

$$\epsilon(t) \sim \text{Gamma}\left(r, \frac{1}{r}\right) = \frac{r^r}{\Gamma(r)} \epsilon(t)^{r-1} \exp(-r\epsilon(t)) \quad (4.2.30)$$

such that $\mathbb{E}[\epsilon(t)] = 1$ and $\text{Var}[\epsilon(t)] = r^{-1}$ yields a gamma-Poisson model. Integrating out $\epsilon(t)$ gives a negative binomial distribution with mean $\theta(t)$ and inverse dispersion parameter $\phi = \frac{1}{r}$ with probability density function

$$f(y(t)) = \frac{\Gamma(\phi^{-1} + y(t))}{y(t)! \Gamma(\phi^{-1})} \left(\frac{\phi^{-1}}{\phi^{-1} + \theta(t)} \right)^{\phi^{-1}} \left(\frac{\theta(t)}{\phi^{-1} + \theta(t)} \right)^{y(t)}, \quad (4.2.31)$$

$$\mathbb{E}[y(t) \mid x(t)] = \theta(t), \quad (4.2.32)$$

$$\text{Var}[y(t) \mid x(t)] = \mathbb{E}[y(t) \mid x(t)] + \phi \mathbb{E}^2[y(t) \mid x(t)]. \quad (4.2.33)$$

Similarly, lognormal-Poisson model for counts is created by setting

$$\epsilon(t) \sim \log \mathcal{N}(0, \sigma^2) \quad (4.2.34)$$

which yields a heavy-tailed lognormal-Poisson model where

$$\mathbb{E}[\epsilon(t)] = \exp\left(\frac{\sigma^2}{2}\right), \quad (4.2.35)$$

$$\text{Var}[\epsilon(t)] = \exp(\sigma^2)(\exp(\sigma^2) - 1), \quad (4.2.36)$$

$$\mathbb{E}[y(t) \mid x(t)] = \theta(t) \exp\left(\frac{\sigma^2}{2}\right), \quad (4.2.37)$$

$$\text{Var}[y(t) \mid x(t)] = \mathbb{E}[y(t) \mid x(t)] + (\exp(\sigma^2) - 1)\mathbb{E}^2[y(t) \mid x(t)]. \quad (4.2.38)$$

Finally, the dynamic lognormal and gamma mixed negative binomial (DLGNB) regression model is presented. Let $p(t) = \frac{\exp(\psi(t))}{1+\exp(\psi(t))} = \frac{\exp([x(t)]^\top \beta \varepsilon(t))}{1+\exp([x(t)]^\top \beta \varepsilon(t))}$, $\text{logit}(p(t)) = \log \frac{p(t)}{1-p(t)}$. Then, the observation model reads:

$$y(t) \sim \text{NB}(r, p(t)), \quad \psi(t) = \text{logit}(p(t)) = [x(t)]^\top \beta + \log \varepsilon(t) \quad (4.2.39)$$

$$\varepsilon(t) \sim \log \mathcal{N}(0, \varphi^{-1}), \quad \varphi \sim \text{Gamma}(e_0, 1/f_0) \quad (4.2.40)$$

$$\beta \sim \prod_{p=0}^P \mathcal{N}(0, \alpha_p^{-1}), \quad \alpha_p \sim \text{Gamma}(c_0, 1/d_0) \quad (4.2.41)$$

$$r \sim \text{Gamma}(a_0, 1/h), \quad h \sim \text{Gamma}(b_0, 1/g_0) \quad (4.2.42)$$

where $\varphi = \sigma^{-2}$ and $a_0, b_0, c_0, d_0, e_0, f_0$, and g_0 are gamma hyperparameters. The DLGNB model properties are as follows. By law of total expectation and law of total variance:

$$\mathbb{E}[y(t) \mid x(t)] = \mathbb{E}_{\varepsilon(t)}[\mathbb{E}[y(t) \mid x(t), \varepsilon(t)]] \quad (4.2.43)$$

$$= \exp([x(t)]^\top \beta + \sigma^2/2 + \log r), \quad (4.2.44)$$

$$\text{Var}[y(t) \mid x(t)] = \mathbb{E}_{\varepsilon(t)}[\text{Var}[y(t) \mid x(t), \varepsilon(t)]] + \quad (4.2.45)$$

$$\text{Var}_{\varepsilon(t)}[\mathbb{E}[y(t) \mid x(t), \varepsilon(t)]] \quad (4.2.46)$$

$$= \mathbb{E}[y(t) \mid x(t)] + \quad (4.2.47)$$

$$\left(\exp(\sigma^2)(1 + r^{-1}) - 1 \right) \mathbb{E}^2[y(t) \mid x(t)]. \quad (4.2.48)$$

Quasi-dispersion parameter $\kappa = \exp(\sigma^2)(1 + r^{-1}) - 1$ shows that the DLGNB model is a generalization of the NB model ($\sigma^2 = 0$) and the lognormal-Poisson model ($\phi = r^{-1} = 0$) which includes one extra degree of freedom to incorporate both kinds of random effects, with their proportion automatically inferred.

In the regression model, explanatory variables and coefficients are defined as follows. The

solution of the compartmental model, $\tilde{\theta}^C(t)$, $\tilde{\theta}^H(t)$ and $\tilde{\theta}^D(t)$, corresponding to the respective I , H , and D compartments may be used as explanatory variables in the regression:

$$x(t) = [1, x_1(t), \dots, x_P(t)]^\top = [1, \tilde{\theta}^C(t), \tilde{\theta}^H(t), \tilde{\theta}^D(t)]^\top, \quad (4.2.49)$$

$$\beta = [\beta_0, \dots, \beta_P]^\top = [\beta^0, \beta^C, \beta^H, \beta^D]^\top. \quad (4.2.50)$$

4.2.4 Non-parametric Bayesian basis expansion model

Modelling parameters $\psi(t)$ as time-varying allows for accounting of infection rate variation due to human behaviour, non-pharmaceutical interventions (such as masking and governmental policies), and emergence of new viral strains.

In the Bayesian framework, shrinkage priors G_h , $h = 1, \dots, H$, have high density at zero and heavy tails to avoid over-shrinking which emulate the role of a smoothing parameter:

$$\begin{aligned} \psi(t) &= \xi_0 + \sum_{h=1}^H \xi_h b_h(t) \\ \xi_0 &\sim \mathcal{N}(0, \sigma_\psi^2) \\ \xi_h &\sim G_h, \quad h = 1, \dots, H. \end{aligned} \quad (4.2.51)$$

In the temporal Bayesian shrinkage spline, the choice of shrinkage prior G_h can take on various distributions, including t -distribution, Laplace distribution, horseshoe distribution, Pearson type VII, or generalized double Pareto. Local Student's t prior [53] may be defined using a scale mixture of normals formulation [51]:

$$\begin{aligned} G_h &| \tau_h^2, \sigma^2 \sim \text{Normal}(0, \sigma^2 \tau_h^2), \\ \tau_h^2 &| \nu, \lambda \sim \text{Inv-Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2\lambda}\right) \quad \text{for } h = 1, \dots, H, \\ \lambda &\sim \text{half-Cauchy}(0, 1) \end{aligned}$$

This implies a Student's t_ν prior on w_h [54].

Alternatively, setting $\tau_t^2 | \nu \sim \text{Inv-Gamma}(\frac{\nu}{2}, \frac{\nu}{2})$ gives $w_t \sim \text{Student-}t(\nu)$ [54].

The Bayesian Lasso [55] is as follows:

$$\begin{aligned} G_h \mid \tau_h^2, \sigma^2 &\sim \text{Normal}(0, \sigma^2 \tau_h^2), \\ \tau_h^2 \mid \lambda^2 &\sim \text{Exponential}\left(\frac{\lambda^2}{2}\right) \quad \text{for } h = 1, \dots, H, \\ \lambda &\sim \text{half-Cauchy}(0, 1) \end{aligned}$$

Integrating out τ_h^2 yields double-exponential (Laplace) priors on the coefficients:

$$G_h \mid \lambda, \sigma \sim \text{Double-Exponential}\left(0, \frac{\sigma}{\lambda}\right)$$

for $h = 1 \dots, H$.

Finally, the horseshoe prior [38] is:

$$\begin{aligned} G_h \mid \tau_h^2 &\sim \text{Normal}(0, \tau_h^2), \\ \tau_h \mid \lambda &\sim \text{half-Cauchy}(0, \lambda) \quad \text{for } h = 1, \dots, H, \\ \lambda &\sim \text{half-Cauchy}(0, \sigma) \end{aligned}$$

For fixed $\lambda = \sigma = 1$, the prior for the shrinkage factor $\kappa_h = \frac{1}{1+\tau_h^2}$ resembles a horseshoe shaped Beta(0.5,0.5) prior. Large coefficients occur for $\kappa_t \approx 0$ (no shrinkage), whereas small coefficients occur for $\kappa_h \approx 1$ (heavy shrinkage). This is an example of a global-local shrinkage prior with time-specific local shrinkage component τ_h as well as a global shrinkage component λ .

Each of these shrinkage priors are sparsity-inducing in the sense that some coefficients in the regression will be shrunk toward, or set to, zero.

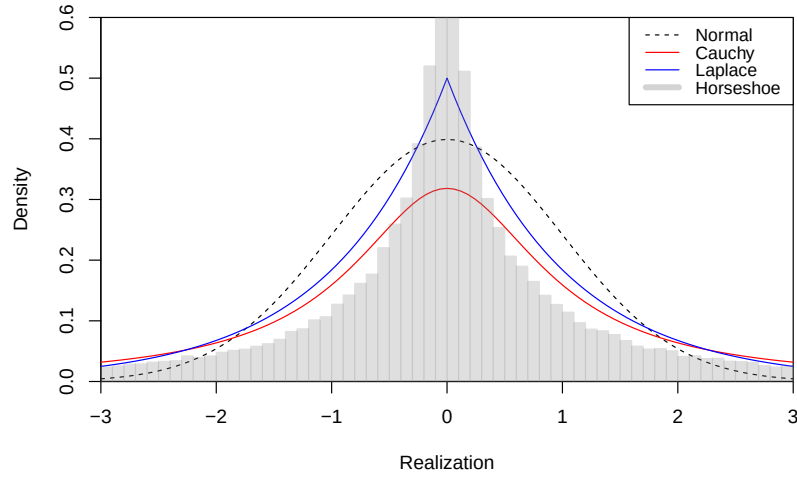


Figure 4.3: Shrinkage priors for the Bayesian spline functions

Fixing a multi-dimensional spline basis and placing Gaussian prior over the coefficients results in a Gaussian process [70], albeit one that lives in finite-dimensional space. Thus, the present setting utilizing shrinkage priors over the coefficients can be regarded as a generalization of Gaussian processes [71].

Regarding the epidemiological parameter specification, a timeline of COVID-19 progression and model parameters is shown below [72].

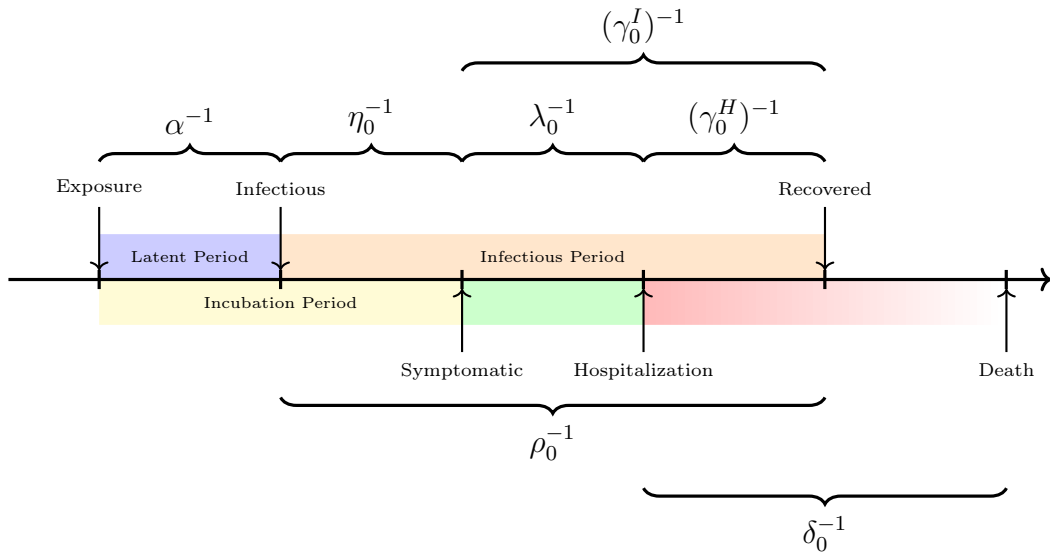


Figure 4.4: Timeline of COVID-19 progression and model parameters.

Parameter	Interpretation	Prior	Mean	95% coverage
α	Inverse of latent period	Beta(31.5, 58.5)	0.35	(0.27, 0.43)
η_0	Inverse of infectious period before symp.	Beta(32.4, 58.4)	0.36	(0.28, 0.44)
λ_0	Inverse of infectious period (symp. to hosp.)	Beta(82.0, 163.9)	0.33	(0.28, 0.38)
γ_0^I	Inverse of infectious period after symp.	Beta(19.3, 115.8)	0.14	(0.10, 0.20)
γ_0^H	Inverse of infectious period after hosp.	$[(\gamma_0^I)^{-1} - \lambda_0^{-1}]^{-1}$	0.25	(0.14, 0.49)
ρ_0	Inverse of entire infectious period	$[\eta_0^{-1} + (\gamma_0^I)^{-1}]^{-1}$	0.10	(0.075, 0.127)
δ_0	Inverse of time to death after infectious per.	Beta(29.1, 494.6)	0.056	(0.040, 0.073)
κ	Ratio factor of exposed asymp. to symp.	Gamma(25, 5)	5	(3.47, 6.75)

Table 4.1: Informing model parameters using clinical knowledge [67].

Moving on to the parameter submodels, time-varying epidemiological parameter $\beta(t)$ is modelled as follows.

Since $\beta(t) > 0$ for all t , let

$$\log[\beta(t)] = \xi_{0\beta} + [b_h(t)]^\top \xi_h \quad (4.2.52)$$

where

$$\begin{aligned} \xi_{0\beta} &\sim \mathcal{N}(0, 1) \\ \xi_h &\sim G_{h\beta}, \quad h = 1, \dots, H. \end{aligned} \quad (4.2.53)$$

Assimilating covariates $x_i, i = 1, \dots, P$, into the model can be accomplished as follows:

$$\log[\beta(t)] = \xi_{0\beta}(t) + [b_h(t)]^\top \xi_h + \sum_{i=1}^P x_i(t) \xi_{\beta; x_i} \quad (4.2.54)$$

where

$$\begin{aligned}\xi_{0\beta} &\sim \mathcal{N}(0, 1) \\ \xi_h &\sim G_{h\beta} \\ \xi_{\beta;x_i} &\sim G_{h\beta;x_i}, \quad h = 1, \dots, H.\end{aligned}\tag{4.2.55}$$

At last, the finalized flow chart of the model is shown below.

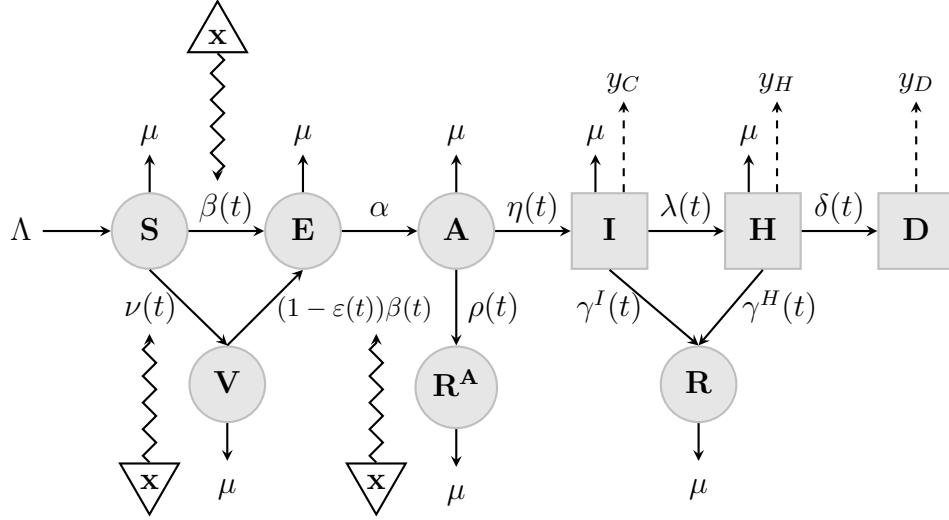


Figure 4.5: Flow chart of the model with covariate inclusion.

Parameter α denotes the rate of exposed individuals becoming infectious, and α^{-1} represents the latent or pre-infectious period (time period between exposure to the disease and being able to infect others).

$$\alpha \sim \text{Beta}(31.5, 58.5)$$

with mean $0.35 = \frac{31.5}{31.5+58.5} = \frac{1}{2.8571}$ and standard deviation $0.05 = \sqrt{\frac{31.5 \times 58.5}{(31.5+58.5)^2(31.5+58.5+1)}}$. Due to the asymptomatic condition, incubation period (time period between exposure and symptom onset) is at least the latent period, with their difference equal to the detection rate. For COVID-19, individuals usually become infectious before the onset of symptoms.

Parameter $\eta(t)$ denotes the rate of detection of undetected infectious individuals and η_0^{-1} represents the time between being infectious and becoming symptomatic. The average time to detection for undetected infectious individuals is assumed to be similar to the time to symptoms onset. Thus, undetected infectious individuals (who are eventually diagnosed) are typically diagnosed with the disease when their symptoms start to appear.

Define $\eta(t) = \eta_0 \psi(t)$, where $\psi(t)$ represents the proportion of undetected infectious individuals that are eventually diagnosed with the disease. The incubation period is approximately $\alpha^{-1} + \eta_0^{-1}$. Let $\eta_0 \sim \text{Beta}(32.4, 58.4)$, with mean $0.3568 = \frac{32.4}{32.4+58.4} = \frac{1}{2.8024}$ and standard deviation $0.05 = \sqrt{\frac{32.4 \times 58.4}{(32.4+58.4)^2(32.4+58.4+1)}}$. Let

$$\psi(t) = a_\psi - b_\psi \exp(-c_\psi t) \quad (4.2.56)$$

where $0 < b_\psi < a_\psi < 1$ and $c_\psi > 0$. Here, $\psi(0) = a_\psi - b_\psi$, $\psi(t) = a_\psi$ as $t \rightarrow \infty$ and c_ψ characterizes testing capacity increase.

Model $a_\psi \sim \text{Beta}(55.5, 18.5)$ (mean=0.75, SD=0.05), $\frac{b_\psi}{a_\psi} \sim \text{Unif}(0, 1)$, $c_\psi \sim \text{Gamma}(5, 100)$.

Regarding the hospitalization rate, let parameter $\lambda(t)$ denote the rate of detected infectious individuals becoming hospitalized:

$$\lambda(t) = \nu(t)\lambda_0$$

where $\nu(t)$ is the proportion of detected infectious individuals who will eventually be hospitalized, and λ_0^{-1} represents the time interval between disease detection and hospitalization such that $\lambda_0 \sim \text{Beta}(82, 163.9)$ (mean= $\frac{1}{3}$, SD=0.03).

For the hospitalization rate priors, proportion of hospitalizations is modelled as:

$$\text{logit}[\nu(t)] = \zeta_{0\nu} + [b_h(t)]^\top \zeta_\nu \quad (4.2.57)$$

where $\zeta_{0\nu}$ is the intercept and $b_h(t)$ is a vector of basis functions at time t and ζ_ν is the vector of regression coefficients. Let $\zeta_{0\nu} \sim \mathcal{N}(-2, 1)$, $\zeta_\nu \sim \mathcal{N}(0, \sigma_\nu^2 I)$ and $\sigma_\nu \sim \mathcal{N}_+(0.5, 0.1^2)$.

Parameter $\gamma^I(t)$ denotes the rate of detected infectious individuals becoming noninfectious without being hospitalized:

$$\gamma^I(t) = (1 - \nu(t))\gamma_0^I$$

where $\gamma_0^I \sim \text{Beta}(19.3, 115.8)$ (mean= $\frac{1}{7}$, SD=0.03) such that $(\gamma_0^I)^{-1}$ represents the time interval between disease detection and when the individual is no longer able to infect others.

Parameter $\rho(t)$ denotes the rate of undetected infectious individuals becoming noninfectious without ever being detected as having the disease:

$$\rho = (1 - \psi(t))\rho_0$$

where $1 - \psi(t)$ represents the proportion of undetected infectious individuals that are never di-

agnosed with the disease and ρ_0^{-1} is the time interval during which a never-detected infectious individual is capable of infecting others (the whole infectious period): $\rho_0^{-1} = \eta_0^{-1} + (\gamma_0^I)^{-1}$.

That is, the whole infectious period is the time interval between when an individual is capable of infecting others and when the individual is diagnosed with the disease, plus the time interval between diagnosis and when the individual is no longer able to infect others.

Denote the rate of death for hospitalized individuals by (time-varying) parameter $\delta(t)$, and (time-varying) recovery rate of hospitalized individuals by $\gamma^H(t)$. Further, assume $\omega(t)$ represents the proportion of deaths among hospitalized individuals, and $1 - \omega(t)$ the proportion of recoveries.

$$\gamma^H(t) = (1 - \omega(t))\gamma_0^H, \quad \delta(t) = \omega(t)\delta_0 \quad (4.2.58)$$

where $\delta_0 \sim \text{Beta}(29.1, 494.6)$ with mean $\frac{1}{18}$ and standard deviation 0.01.

For the death rate priors, let

$$(\gamma_0^H)^{-1} + \lambda_0^{-1} = (\gamma_0^I)^{-1}.$$

The proportion of deaths is modelled as:

$$\text{logit}[\omega(t)] = \zeta_{0\omega} + [b_h(t)]^T \zeta_\omega \quad (4.2.59)$$

where $\zeta_{0\omega}$ is the intercept and $b_h(t)$ is a vector of basis functions at time t and ζ_ω is the vector of regression coefficients. Let $\zeta_{0\omega} \sim \mathcal{N}(-1, 1)$, $\zeta_\omega \sim \mathcal{N}(0, \sigma_\omega^2 I)$ and $\sigma_\omega \sim \mathcal{N}_+(0.5, 0.1^2)$.

Typically, at the individual level, vaccine effectiveness ε is modelled as an exponential decay function of time. Here, the assumption is that at the population level, vaccine effectiveness is, on average, relatively constant with the passage of time for the analysis period. The vaccine breakthrough (failure) rate σ is modelled as the complement of vaccine effectiveness:

$$\sigma = 1 - \varepsilon \quad (4.2.60)$$

where $\varepsilon \sim \text{Beta}(55.5, 18.5)$ (mean=75%, SD=5%). Viral variants responds differently to vaccination. To take such differences into account, another setting whereby $\varepsilon \sim \text{Beta}(31.5, 3.5)$ (mean=90%, SD=5%) is analyzed.

Alternative modelling techniques could be used if vaccine effectiveness is known to follow specified temporal patterns. For example, a Brownian bridge could be used to account for initial and terminal boundary time points of the analysis period. Additionally, waning immunity could be incorporated in a time-varying fashion into a function of the vaccine

effectiveness.

4.2.5 Markov chain Monte Carlo implementation

Parameter estimation can be accomplished by the Monte Carlo method [34]. Monte Carlo simulation is a sampling process from a target distribution $p(\psi \mid y)$. Once sufficient samples are obtained, the inferences of target distribution can be approximated by sample statistics, such as sample mean, sample standard error and sample percentile. Traditional Monte Carlo simulation methods involve calculation of the inverse cumulative distribution function, the rejection sampling method, which generates independent samples. In contrast, Markov chain Monte Carlo (MCMC) methods generate a Markov chain whose stationary distribution is equivalent to the target distribution. In MCMC, the next sampled value typically depends on the previous sampled value.

In contrast, Hamiltonian dynamics can be used to intelligently guide the iterative process. Hamiltonian Monte Carlo (HMC) [7] proposes a parameter value by computing a trajectory according to Hamiltonian dynamics that takes the target posterior distribution into account by considering position (target distribution) ψ and momentum ϕ in a Hamiltonian function as

$$\begin{aligned} H(\psi, \phi) &= U(\psi) + K(\phi) \\ -\log p(\psi, \phi) &= -\log p(\psi \mid y) - \log p(\phi) \end{aligned} \quad (4.2.61)$$

Taking $p(\phi)$ to be zero-mean multivariate Gaussian gives $\frac{1}{2}\phi^\top \Sigma^{-1}\phi$ where the mass matrix Σ^{-1} is symmetric positive-definite, typically set to identity. By Hamilton's equation of motion:

$$\begin{aligned} \frac{d\psi_i}{dt} &= \frac{\partial H}{\partial \phi_i} = [\Sigma^{-1}\phi]_i \\ \frac{d\phi_i}{dt} &= -\frac{\partial H}{\partial \psi_i} = -\frac{\partial U}{\partial \psi_i} \end{aligned} \quad (4.2.62)$$

For implementation of Hamilton's equation, Euler's method is as follows:

$$\begin{aligned} \phi_i(t + \varepsilon) &= \phi_i(t) + \varepsilon \frac{d\phi_i(t)}{dt} = \phi_i(t) - \varepsilon \frac{\partial U}{\partial \psi_i}(\psi_i(t)) \\ \psi_i(t + \varepsilon) &= \psi_i(t) + \varepsilon \frac{d\psi_i(t)}{dt} = \psi_i(t) + \varepsilon [\Sigma^{-1}\phi]_i \end{aligned} \quad (4.2.63)$$

Even more efficient is the trajectory volume-preserving Leapfrog method:

$$\begin{aligned}
\phi_i(t + \frac{1}{2}\varepsilon) &= \phi_i(t) - \frac{1}{2}\varepsilon \frac{\partial U}{\partial \psi_i}(\psi_i(t)) \\
\psi_i(t + \varepsilon) &= \psi_i(t) + \varepsilon[\Sigma^{-1}\phi(t + \frac{1}{2}\varepsilon)]_i \\
\phi_i(t + \varepsilon) &= \phi_i(t + \frac{1}{2}\varepsilon) - \frac{1}{2}\varepsilon \frac{\partial U}{\partial \psi_i}(\psi_i(t + \varepsilon))
\end{aligned} \tag{4.2.64}$$

Posterior distribution simulation by Hamiltonian MCMC is outlined below.

Algorithm 1 Hamiltonian MCMC

- 1: Input: (ψ^{t-1}, ϕ^{t-1})
 - 2: Time t update rule:
 - 3: • draw auxiliary momentum variable
 - 4: $\phi^{t-1} \sim \mathcal{N}(0, \Sigma)$
 - 5: • evolve joint (ψ^{t-1}, ϕ^{t-1}) via L -step leapfrog method
 - 6: $\phi_i(t - 1 + \frac{1}{2}\varepsilon) = \phi_i(t - 1) + \frac{1}{2}\varepsilon \frac{\partial \log p(\psi(t-1))}{\partial \psi_i}$
 - 7: $\psi_i(t - 1 + \varepsilon) = \psi_i(t - 1) + \varepsilon[\Sigma^{-1}\phi(t - 1 + \frac{1}{2}\varepsilon)]_i$
 - 8: $\phi_i(t - 1 + \varepsilon) = \phi_i(t - 1 + \frac{1}{2}\varepsilon) + \frac{1}{2}\varepsilon \frac{\partial \log p(\psi(t-1+\varepsilon))}{\partial \psi_i}$
 - 9: Obtain proposed $(\psi^* = \psi^{t-1+\varepsilon L}, \phi^* = \phi^{t-1+\varepsilon L})$.
 - 10: • Metropolis accept step
 - 11: $A = \min\{1, \exp[H(\psi^{t-1}, \phi^{t-1}) - H(\psi^*, \phi^*)]\}$
 - 12: Accept (ψ^*, ϕ^*) and set $(\psi^t, \phi^t) = (\psi^*, \phi^*)$ with probability A .
 - 13: Otherwise, reject (ψ^*, ϕ^*) and set $(\psi^t, \phi^t) = (\psi^{t-1}, \phi^{t-1})$.
 - 14: Output: (ψ^t, ϕ^t)
-

No-U-Turn-Sampler (NUTS) [73], [74] dynamically adjusts the number of leapfrog steps at each iteration to send the trajectory as far as it can go during that iteration. It travels backward and forward along the trajectory in a way that satisfies detailed balance to ensure convergence to the target distribution.

Hamiltonian MCMC tuning parameters (fixed in the inferential computation) include: mass matrix Σ , covariance of the target distribution and a Euclidean metric that rotates and scales the target distribution; step size ε , ε large: the resulting trajectory will be inaccurate and too many proposals will be rejected and ε small: too many steps will be taken by the leapfrog method leading to long simulation times per iteration; number of steps L : L small: the trajectory traced out will be too short and sampling will devolve to a random walk and L large: the algorithm will spend too much time in one iteration.

To assess convergence, a potential scale reduction factor $\hat{R}$ is used. It is created as follows. For I chains, each of length J , split each chain into 2 parts to get $2I$ batches, each of length $J/2$ and label the simulations $\psi_{ij}, i = 1, \dots, 2I, j = 1, \dots, J/2$ and calculate between- and

within-batch variances as a measure of posterior variability:

$$W_{\text{Var}} = \sum_{i=1}^{2I} s_i^2, \quad s_i^2 \text{ the sample variance of } i\text{th batch}, \quad (4.2.65)$$

$$B_{\text{Var}} = \frac{J/2}{2I-1} \sum_{i=1}^{2I} (\bar{\psi}_i - \bar{\psi})^2 \quad (4.2.66)$$

where $\bar{\psi}_i$ is the sample mean of the i th batch and $\bar{\psi}$ is the pooled mean. Both W_{Var} and B_{Var} estimate posterior variance $\text{Var}(\psi \mid y)$. An accurate estimate is given by [7]:

$$\widetilde{\text{Var}}(\psi \mid y) = \frac{J/2-1}{J} W_{\text{Var}} + \frac{1}{J} B_{\text{Var}}. \quad (4.2.67)$$

This is an unbiased estimate under stationarity, and over-estimate including warm-up iterations. However, W_{Var} underestimates $\widetilde{\text{Var}}(\psi \mid y)$ due to sample size $J/2$ and dependent iterations. Thus, convergence is monitored by estimating

$$\hat{R} = \sqrt{\frac{\widetilde{\text{Var}}(\psi \mid y)}{W_{\text{Var}}}} \quad (4.2.68)$$

which declines to 1 as $J \rightarrow \infty$.

For efficiency, batch means method is used. For I chains, each of length J , where each chain is split into M batches, each of length J/M such that CLT holds for each batch. Then, for simulations $\psi_{ij}, i = 1, \dots, IM, j = 1, \dots, J/M$ the batch means $\bar{\psi}_i$ are i.i.d. with posterior mean and posterior variance over J/M . The sample variance of batch means is used to estimate the posterior variance:

$$\widehat{\text{Var}}(\psi \mid y) = \frac{J/M}{IM-1} \sum_{i=1}^{IM} (\bar{\psi}_i - \bar{\psi})^2. \quad (4.2.69)$$

Batch means method can be used to create for confidence intervals for the parameter estimates. The standard error of the posterior mean $\bar{\psi} = \sum_{ij} \frac{\psi_{ij}}{IJ}$ can be approximated by the Monte Carlo standard error:

$$\sqrt{\frac{\widehat{\text{Var}}(\psi \mid y)}{IJ}} = \sqrt{\frac{1}{IM(IM-1)} \sum_{i=1}^{IM} (\bar{\psi}_i - \bar{\psi})^2}. \quad (4.2.70)$$

Thus, a 95% CI of $\hat{\psi}$ is

$$\left(\bar{\psi} \pm 1.96 \sqrt{\frac{\widehat{\text{Var}(\psi | y)}}{IJ}} \right). \quad (4.2.71)$$

Efficiency can also be assessed by effective sample size (ESS). An estimate of effective sample size is [7]:

$$n_{\text{eff}} = \frac{IJ}{1 + 2 \sum_{t=1}^{\infty} \rho_t} \quad (4.2.72)$$

for I chains of length J and ρ_t is the autocorrelation of the Monte Carlo sample at lag t .

For the implementation, NUTS is accessed via **Stan** [73], [74], a probabilistic programming language for statistical inference written in C++ with interfaces to Python, R, and Julia. It applies NUTS and can also do approximate Bayesian inference by variational Bayes or by penalized maximum likelihood estimation. The algorithm includes data and model input and computation of the log posterior density (up to an arbitrary constant that cannot depend on the parameters in the model) and its gradients. In warm-up phase, the tuning parameters ε and L are set and NUTS is used to move through the (possibly constrained) parameter space. Convergence monitoring and inferential summaries are also provided.

4.3 Applications to COVID-19 mobility and vaccination data

The analysis period is taken to be 12/14/2020 to 09/16/2022 (642 days). Daily data is obtained from Ontario Agency for Health Protection and Promotion (Public Health Ontario) COVID-19 Data Tool [21], Ministry of Health COVID-19 Vaccine Data Catalogue [75], and Google Community Mobility Reports [76].

In terms of model settings and initialization values, note that the susceptible, exposed, and undetected infectious (asymptomatic) individuals are not observed at the beginning of analysis and are estimated as: $E(0) + A(0) = \kappa I(0)$ where $\kappa \sim \text{Gamma}(25, 5)$ i.e., about 80% of all infections were undocumented at beginning of epidemic, and $\frac{E(0)}{E(0)+A(0)} \sim \text{Unif}(0, 1)$. The initial number of susceptible individuals is $S(0) = N(0) - E(0) - I(0) - A(0)$ where $N(0)$ is the population size.

Surveillance targets of case, hospitalization, and death counts, vaccination rates and mobility signals are used in the analysis. The data over the analysis period are shown in the following plots.

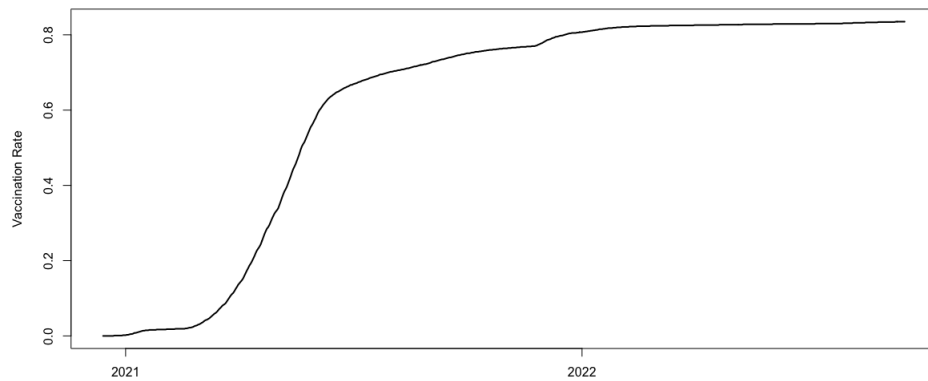


Figure 4.6: Vaccination rate over time.

The first plot illustrates the vaccination rate in Ontario from December 2020 to September 2022. The vaccination rate begins very low at the start of 2021 and rises significantly around January 2021. There is a steep increase in the vaccination rate from January to mid-2021, which indicates a rapid vaccine rollout and high uptake among the population. This increase continues at a slower pace from mid-2021 to the end of the year, suggesting a saturation point where most individuals willing to be vaccinated had already received their doses. By 2022, the rate of increase slows even further, indicating that the majority of the population intending to get vaccinated had done so. The plot demonstrates the successful vaccination efforts and high coverage, which are crucial for controlling the spread of COVID-19.

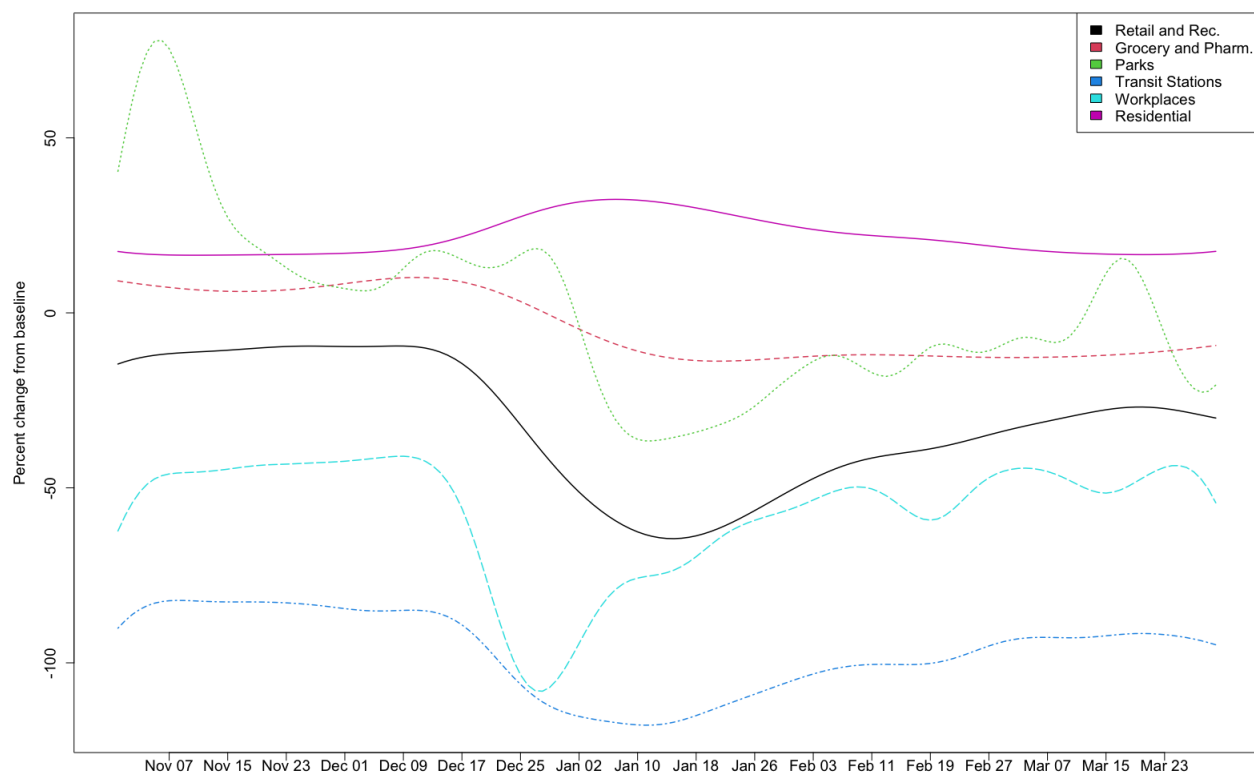


Figure 4.7: Ontario mobility trends.

The second plot presents the mobility trends in Ontario, showing changes in mobility across various categories compared to a baseline. Mobility, as presented in the plot, refers to the relative change in the frequency of visits to specific categories of locations compared to a baseline. These categories include Retail and Recreation, Grocery and Pharmacy, Parks, Transit Stations, Workplaces, and Residential areas. The baseline is typically defined as the median value of mobility for each category during a pre-pandemic reference period, which is calculated over several weeks or months. This baseline serves as a reference point to quantify changes in mobility patterns during the observed period.

The mobility data shown in the figure represent percentage changes from this baseline. For example, a value of +50% indicates a 50% increase in mobility relative to the baseline, while a value of -50% indicates a 50% decrease. These measurements are derived from anonymized and aggregated location data, often collected via GPS-enabled devices and digital services. The data are processed to ensure privacy while allowing the analysis of general trends in population movement.

In the provided plot, each curve represents the percentage change in mobility for one

category of locations over time. Positive values indicate increased mobility compared to the baseline, whereas negative values indicate decreased mobility. For instance, Residential areas might show positive mobility changes during periods when individuals spend more time at home, such as during lockdowns or holidays. Conversely, categories like Retail and Recreation or Transit Stations may exhibit negative changes during the same periods, reflecting reduced movement to these locations.

These mobility trends provide valuable insights into how public movement patterns shift over time. They are particularly useful for understanding behavioural changes driven by external factors such as public health policies, seasonal variations, or significant events. By visualizing these trends, it becomes possible to analyze how different categories of mobility respond to societal and environmental changes, helping to inform policy decisions and public health strategies.

Different lines represent various categories, including retail and recreation, grocery and pharmacy, parks, transit stations, workplaces, and residential areas. During certain periods, sharp declines in mobility are observed, especially around transit stations and workplaces, likely corresponding to lockdowns and work-from-home policies. In contrast, residential mobility increases during these periods, indicating that people stayed home more. Categories like parks show more variability, possibly due to outdoor activities being safer and subject to different restrictions. The overall trend highlights how mobility restrictions and public behaviour fluctuated in response to the pandemic waves and government policies.

These visualizations help to capture the dynamics of the pandemic, showing how public health measures, vaccination campaigns, and individual behaviours evolved over time, influencing the transmission and control of COVID-19.

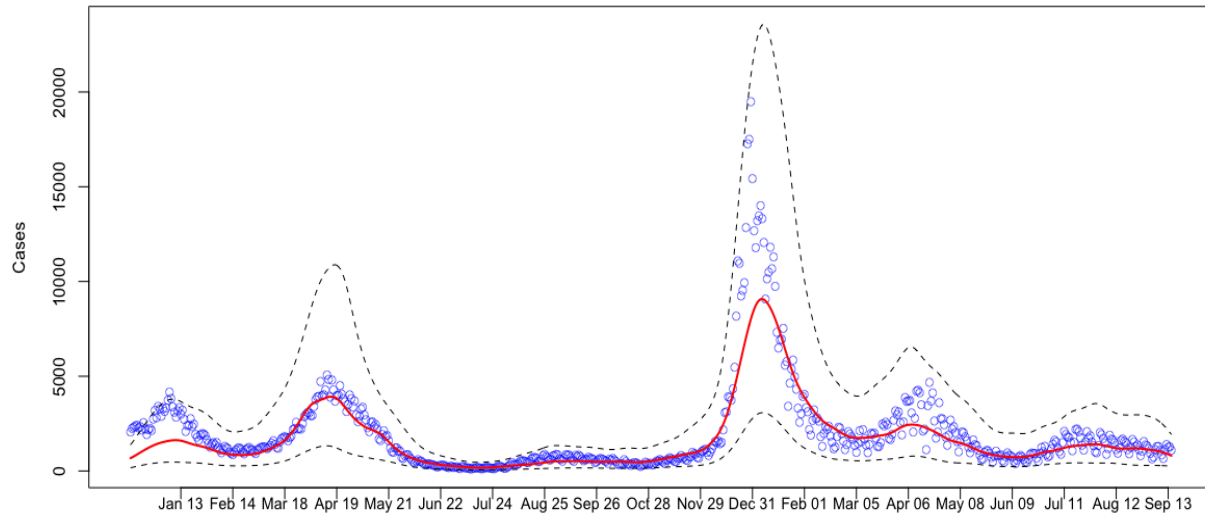


Figure 4.8: Prediction of case counts with the original Poisson model. The analysis period is taken to be 12/14/2020 to 09/16/2022.

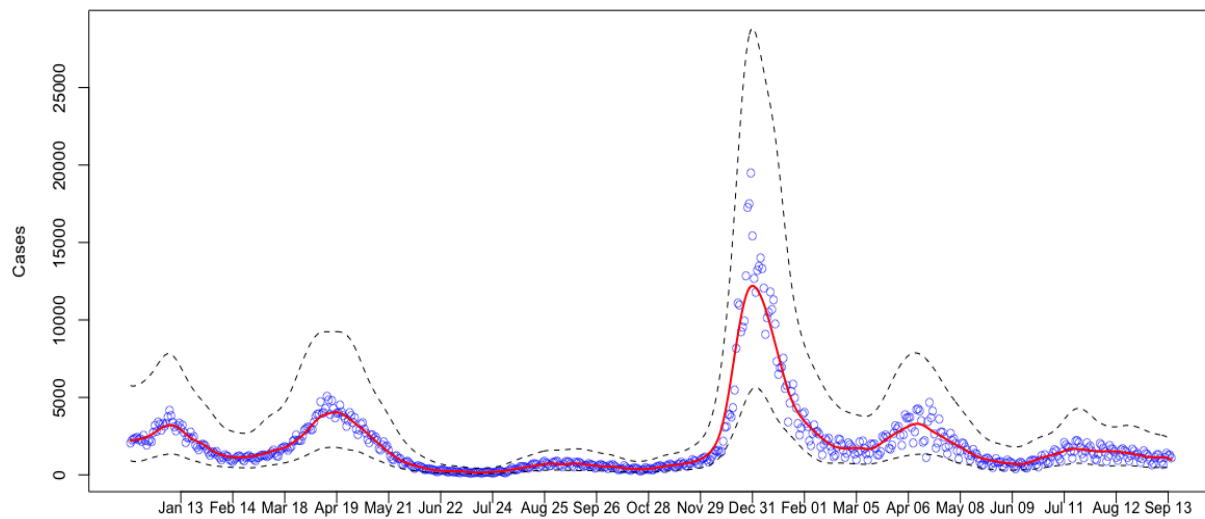


Figure 4.9: Prediction of case counts with the DLGNN model. The analysis period is taken to be 12/14/2020 to 09/16/2022.

The one-step-ahead prediction of COVID-19 case counts using the original Poisson model

is displayed in 4.8. In this plot, the blue circles represent observed daily case counts, the red line represents the predicted case counts from the model, and the dashed black lines indicate confidence intervals around the predictions. While the model captures the general trend of the observed data, it may miss some peaks and troughs, highlighting the Poisson model's limitations in capturing the full variability and nonlinearity of the data. Significant peaks around December 2021 and early 2022 suggest periods of high transmission, possibly due to new variants or changes in public health measures. This model's performance indicates that while it can follow overall trends, it might not be adequate for more complex dynamics.

Prediction of COVID-19 case counts using the DLGNB model is also illustrated in 4.9. Similar to the previous plot, the blue circles represent observed daily case counts, the red line represents the predicted case counts from the DLGNB model, and the dashed black lines show the confidence intervals around the predictions. The DLGNB model appears to provide a better fit to the observed data, capturing more of the variability and peaks in case counts. The improved fit suggests that the DLGNB model, which accounts for over-dispersion and non-Gaussian characteristics, is better suited for modeling the complex dynamics of COVID-19 transmission. Additionally, the narrower confidence intervals indicate higher confidence in the model predictions compared to the Poisson model, demonstrating its effectiveness in capturing the complexities of the pandemic's transmission dynamics.

The vaccination rate plot highlights successful vaccine rollout and coverage in Ontario, crucial for controlling COVID-19. Mobility trends show how public behaviour and restrictions varied throughout the pandemic. Comparing the original Poisson model and the DLGNB model for predicting case counts, the DLGNB model provides a better fit, capturing the complexities of COVID-19 transmission dynamics more effectively. This comparison underscores the importance of using more sophisticated models to understand and predict pandemic trends accurately.

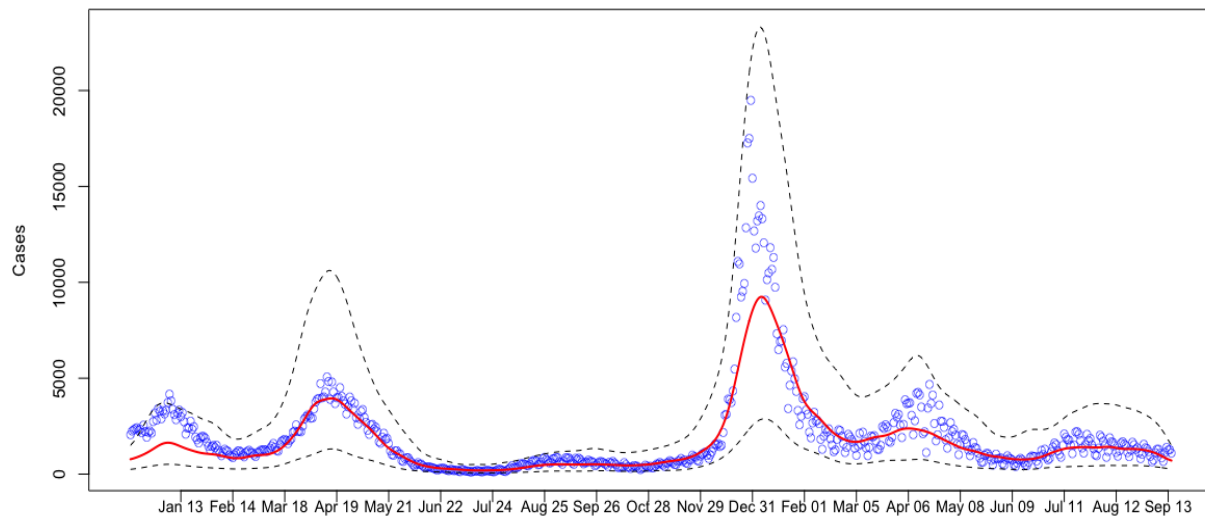


Figure 4.10: Prediction of case counts with accounted vaccination using the Poisson model.

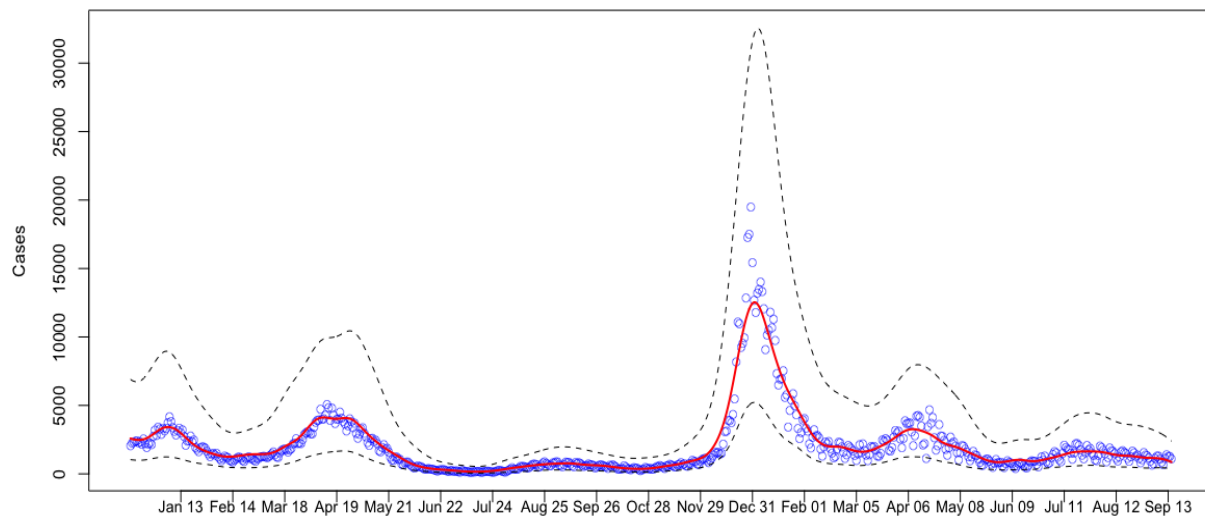


Figure 4.11: Prediction of case counts with accounted vaccination using the DLGNB model.

The above plot illustrates the predicted COVID-19 case counts using the original Poisson model in 4.8, which does not account for the effects of vaccination. The blue circles represent the observed case counts, while the solid red line indicates the model's predicted case counts.

The dashed lines surrounding the red line represent the prediction intervals, which show the range within which the true case counts are expected to fall with a certain level of confidence.

The Poisson model captures the general trends in the case counts, including the peaks and troughs, but it struggles to account for the high variability observed in the data. During periods of significant spikes, such as around early January and late December, the model's predictions tend to underestimate the actual case counts. The prediction intervals are relatively wide, indicating a high level of uncertainty in the predictions, especially during periods of high transmission.

The next plot, in [4.9](#), shows the predicted case counts using the DLGNB model. As in the previous plot, the blue circles represent the observed case counts, the solid red line indicates the model's predicted case counts, and the dashed lines are the prediction intervals.

The DLGNB model provides a more accurate fit to the observed data compared to the Poisson model, particularly in capturing the sharp increases and decreases in case counts. The prediction intervals are narrower than those of the Poisson model, suggesting that the DLGNB model offers more precise predictions with lower uncertainty. The model performs well during the peaks, providing predictions that closely follow the observed data, even during significant spikes in cases.

The third plot, in [4.10](#), depicts the predicted case counts using the Poisson model while accounting for the effects of vaccination. The observed case counts, predicted case counts, and prediction intervals are represented in the same manner as the previous plots.

Incorporating vaccination into the Poisson model improves the accuracy of the predictions. The model now better captures the decrease in case counts observed after the vaccination rollout. The model still struggles during peak periods, but the predictions are closer to the observed data compared to the original Poisson model without vaccination. The prediction intervals remain wide, indicating persistent uncertainty, but there is a noticeable improvement in the model's performance.

The fourth plot, in [4.11](#), shows the predicted case counts using the DLGNB model while accounting for the effects of vaccination. The representation of observed case counts, predicted case counts, and prediction intervals follows the same format. The DLGNB model with vaccination provides the most accurate predictions among all models analyzed. The predicted case counts closely follow the observed data, even during periods of high transmission. The prediction intervals are the narrowest, indicating high confidence in the model's predictions. The model effectively captures the impact of vaccination, showing a significant reduction in predicted case counts following the vaccination campaign.

When comparing the model results, it is clear that the DLGNB model outperforms the Poisson model in terms of prediction accuracy and precision. The DLGNB model's ability to

handle over-dispersion and further account for vaccination effects leads to narrower prediction intervals and more accurate fits to the observed data. Both models show improvements in their predictions when vaccination is accounted for. However, the DLGNB model's predictions are more responsive to changes in case counts due to vaccination, providing a closer match to the observed data.

Incorporating vaccination data into the models significantly enhances their predictive capabilities. The DLGNB model demonstrates superior performance in capturing the dynamics of COVID-19 case counts, providing the most reliable and precise predictions.

Model/Data	t_{peak}	$\hat{y}_C^{\text{max}}$ (% Error)	Smoothed $\hat{y}_C^{\text{max}}$ (% Error)
Truth	01/01/2022	19487	13645
Original	01/03/2022	13042 (33.07%)	10860 (20.41%)
DLGNB	12/31/2021	19903 (2.13%)	14124 (3.51%)

Table 4.2: Point prediction of the major epidemic peak time and intensity.

The above table provides a comparison of point predictions for the major epidemic peak time and intensity using different models, specifically contrasting the actual or true observed values with predictions made by the original Poisson model and the DLGNB model. The metrics evaluated include the peak time (t_{peak}), the maximum case count ($\hat{y}_C^{\text{max}}$) with the percentage error, and the smoothed maximum case count with the percentage error.

The observed peak time is January 1, 2022, with a maximum case count of 19,487 and a smoothed maximum case count of 13,645. In comparison, the Poisson model predicted the peak two days later on January 3, 2022, with a maximum case count of 13,042, resulting in a 33.07% error. The smoothed maximum case count predicted by the Poisson model was 10,860, with a 20.41% error. In contrast, the DLGNB model predicted the peak one day earlier, on December 31, 2021, with a maximum case count of 19,903, which is very close to the actual peak, resulting in a minimal error of 2.13%. The smoothed maximum case count predicted by the DLGNB model was 14,124, with a 3.51% error.

These results highlight the superior performance of the DLGNB model over the original Poisson model in predicting the peak time and intensity of COVID-19 cases. The DLGNB model demonstrates higher accuracy in both peak timing and peak case count predictions, with significantly lower percentage errors. This indicates that the DLGNB model is more reliable for epidemic forecasting, particularly in capturing the dynamics of the major epidemic peak. The inclusion of vaccination data and other factors likely contributes to the enhanced predictive capability of the DLGNB model, making it a more robust tool for public health

planning and response.

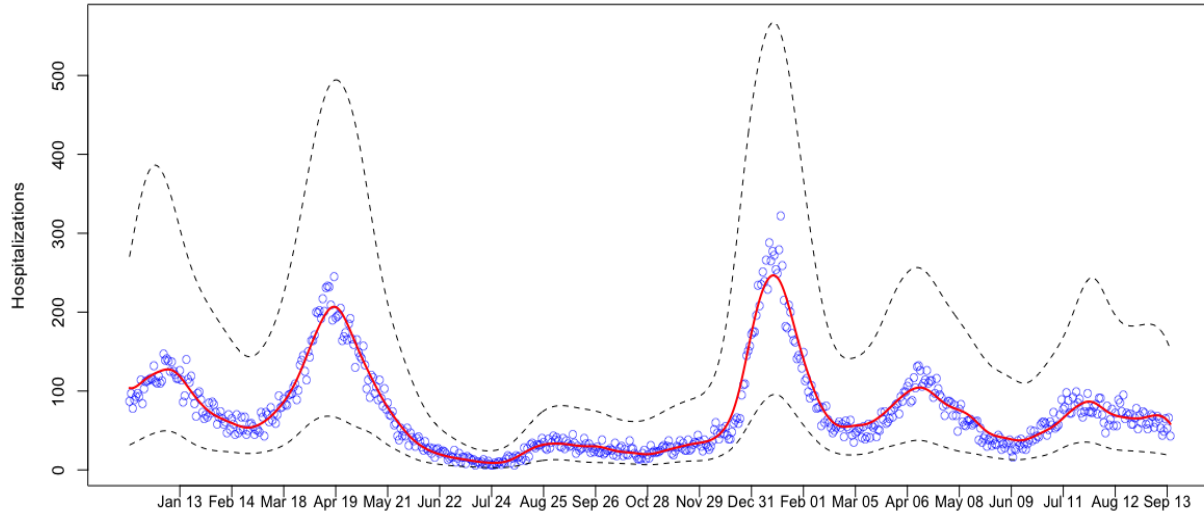


Figure 4.12: Prediction of hospitalization counts with accounted vaccination using the DL-GNB model.

The plot above illustrates the prediction of hospitalization counts using the DLGNB model, which incorporates mobility and vaccination data. The red line represents the smoothed prediction of hospitalizations over time, while the blue circles indicate the observed hospitalization counts. The dashed black lines show the prediction intervals, capturing the uncertainty in the model's predictions.

The model effectively tracks the major trends in hospitalization counts, capturing the significant peaks and troughs observed in the data. Notably, the model accurately predicts the peak hospitalization counts observed around late April and early January, demonstrating its ability to capture the dynamics of hospitalization trends during different waves of the pandemic. The prediction intervals also indicate a reasonable level of confidence in the model's forecasts, with most of the observed data points falling within these intervals. This alignment between the predicted and observed hospitalization counts validates the robustness of the DLGNB model in integrating mobility and vaccination data to enhance prediction accuracy.

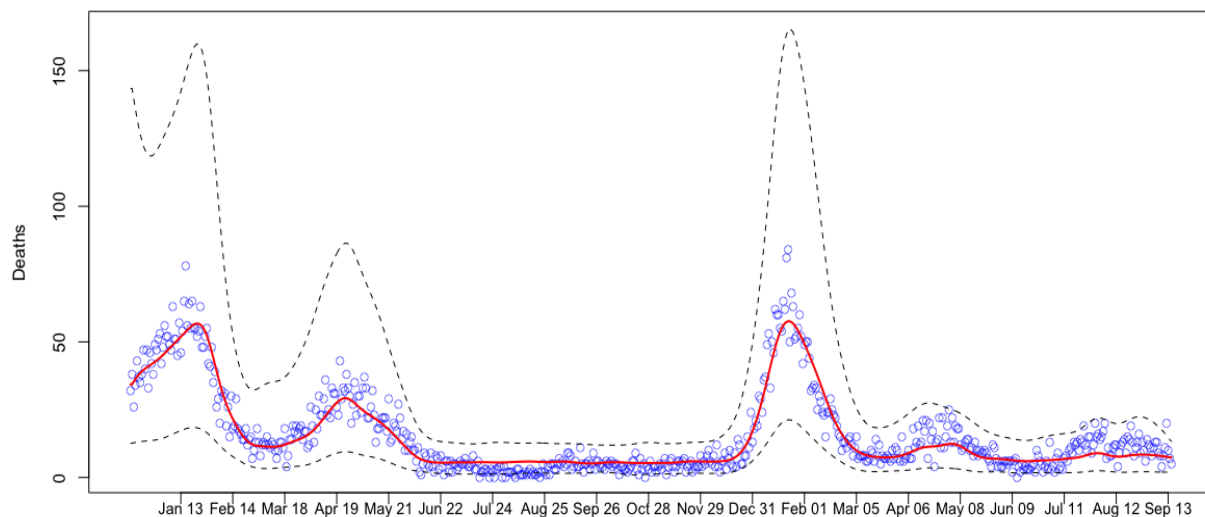


Figure 4.13: Prediction of death counts with accounted vaccination using the DLGNB model.

The plot above presents the prediction of death counts using the DLGNB model, with the inclusion of mobility and vaccination data. Similar to the hospitalization plot, the red line shows the smoothed predictions of death counts, and the blue circles represent the observed death counts. The dashed black lines depict the prediction intervals.

The model demonstrates a strong ability to predict death counts over time, capturing both the overall trend and the specific peaks observed in the data. The significant peaks around early January and late April are well-predicted by the model, indicating its effectiveness in accounting for changes in mobility and vaccination rates. The prediction intervals provide a measure of uncertainty, with the majority of observed death counts falling within these intervals, further validating the model's reliability.

The next plot displays the effective reproduction number $\mathcal{R}_0$ as published by the Ontario COVID-19 Science Advisory Table. This $\mathcal{R}_0$ value changes over time, reflecting changes in transmission dynamics due to public health measures, population behaviour, and emerging variants. Notably, $\mathcal{R}_0$ values consistently above 1 indicate periods of increasing transmission, while values below 1 indicate declining transmission.

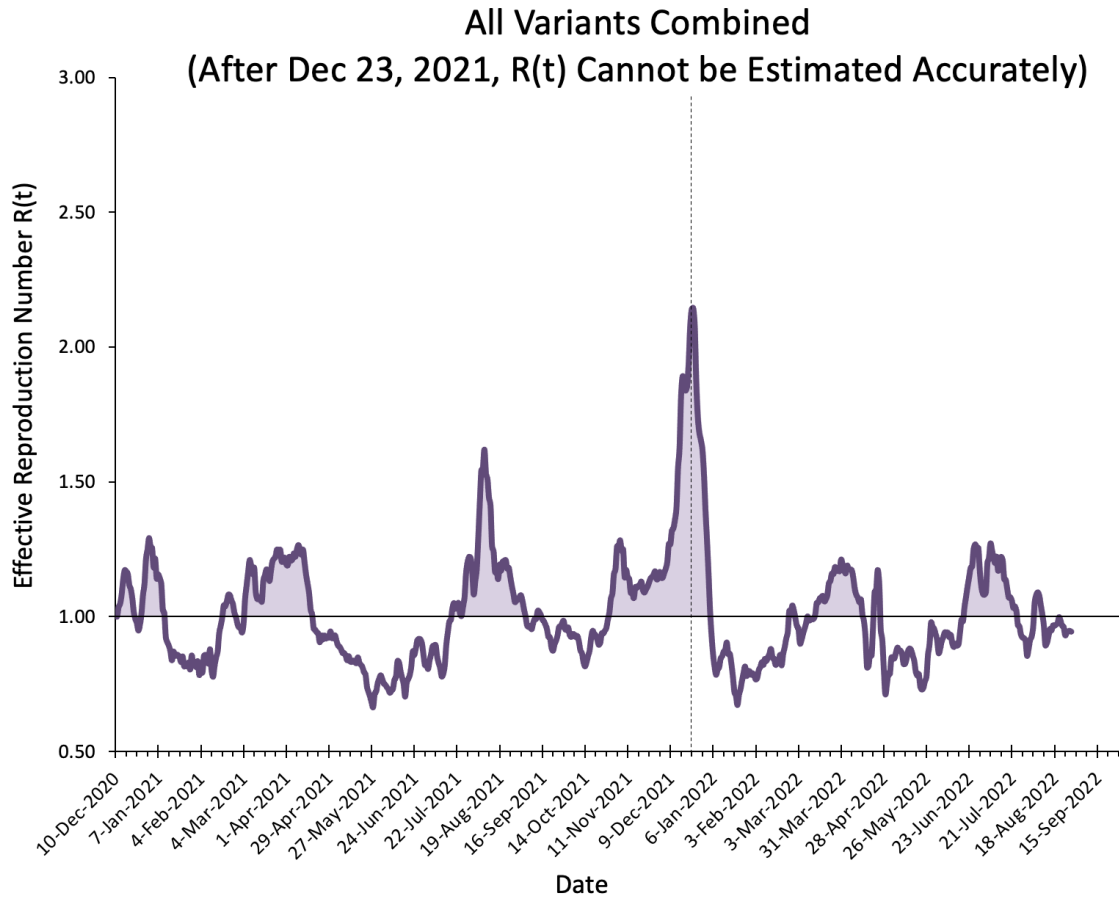


Figure 4.14: $\mathcal{R}_0$ from the Ontario COVID-19 Science Advisory Table.

The following plot shows the model's $\mathcal{R}_0$ values, incorporating mobility and vaccination data. The model's $\mathcal{R}_0$ values generally align with the published data, capturing the major trends and fluctuations. The inclusion of vaccination data results in a more refined estimation of $\mathcal{R}_0$, particularly in periods following significant vaccination efforts, where the reproduction number decreases, reflecting reduced transmission.

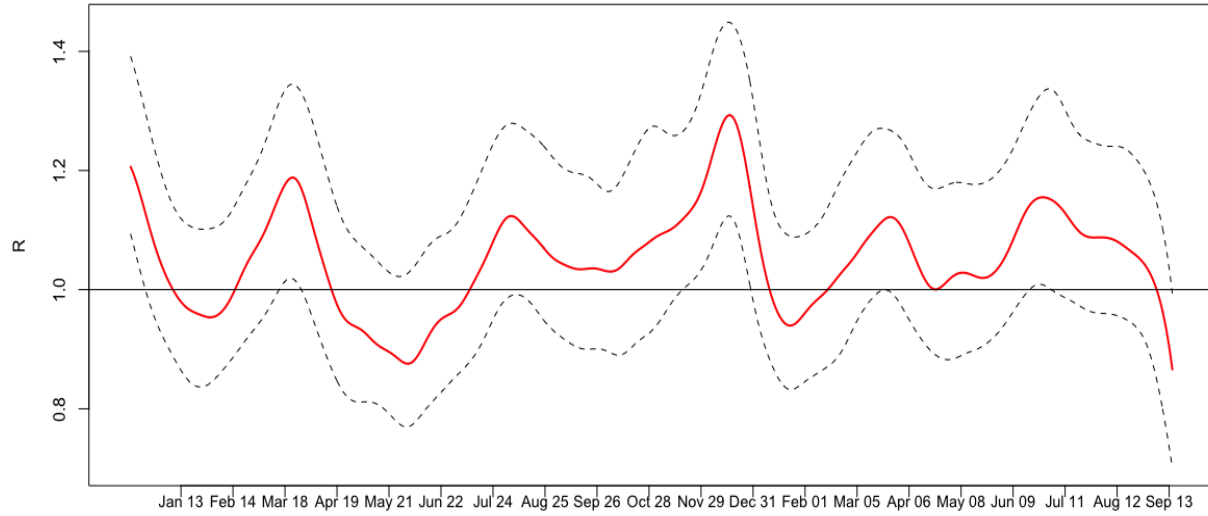


Figure 4.15: Model $\mathcal{R}_0$ including mobility and vaccination data.

The effect of vaccination on case count prediction and the reproduction number is illustrated in the following plots. In the following two figures, the DLGNB model is used to predict case counts under the assumptions of 75% and 90% vaccine effectiveness, respectively. The model shows a significant reduction in predicted case counts as vaccine effectiveness increases. With 75% effectiveness, the peak case counts are notably higher compared to the scenario with 90% effectiveness. This demonstrates the critical role of high vaccine effectiveness in controlling the spread of the virus and reducing the burden on healthcare systems. As vaccine effectiveness increases, the peaks become smaller and more manageable, suggesting that higher vaccine efficacy leads to better control of outbreaks and more stable healthcare demand.

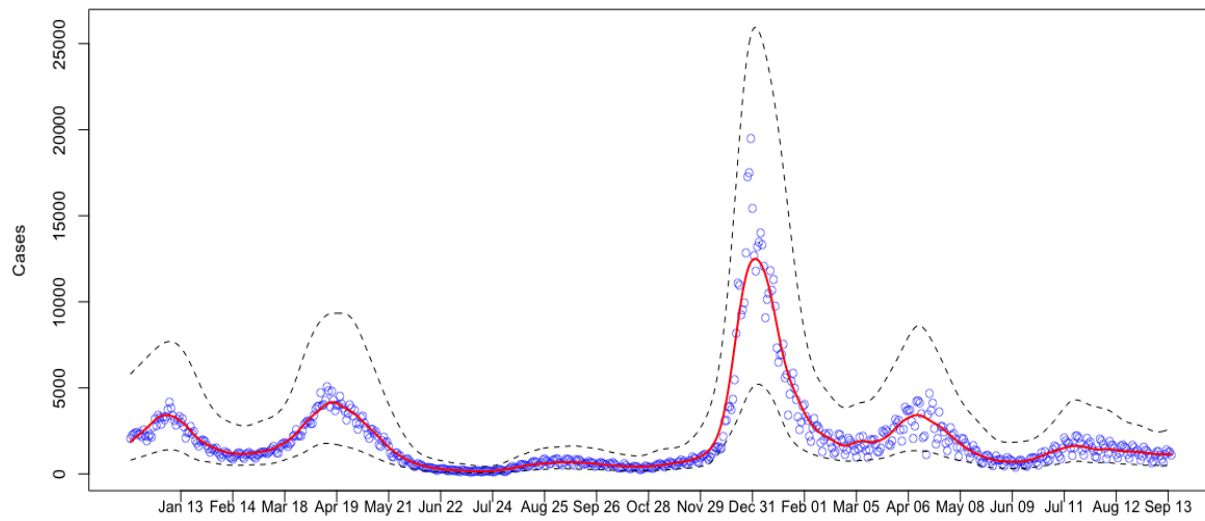


Figure 4.16: Case count prediction for the DLGNB model with a setting of 75% vaccine effectiveness.

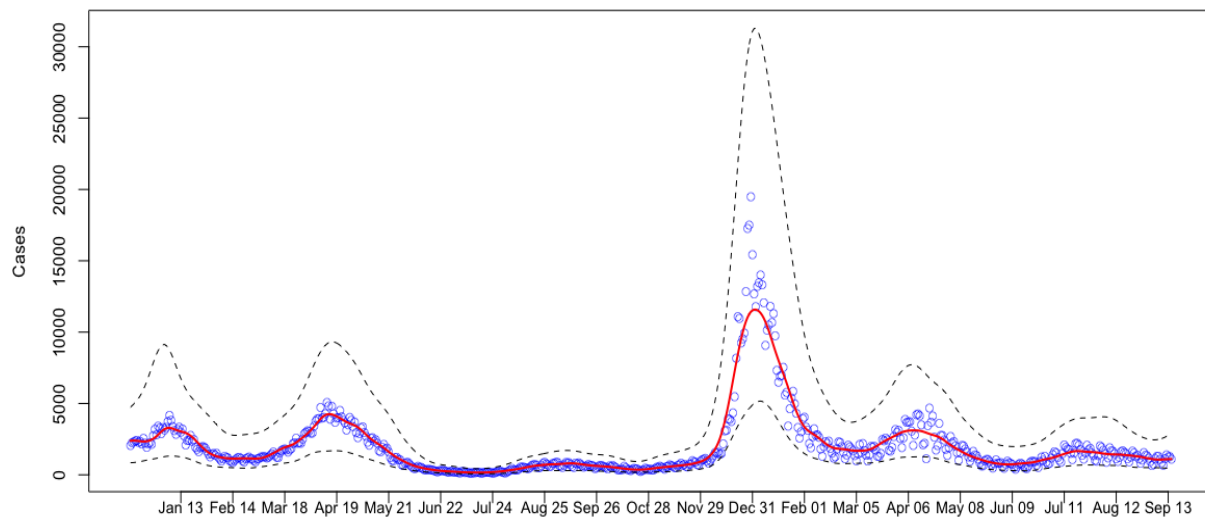


Figure 4.17: Case count prediction for the DLGNB model with a setting of 90% vaccine effectiveness.

The next two figures depict the effective reproduction number $\mathcal{R}_0$ under the same vaccine

effectiveness scenarios (75% and 90%). The $\mathcal{R}_0$ values are consistently below 1 when vaccine effectiveness is at 90%, indicating that each infected person is transmitting the virus to fewer than one other person on average, leading to a decline in new infections. However, with 75% effectiveness, $\mathcal{R}_0$ occasionally rises above 1, suggesting periods where the virus can spread more freely. This analysis highlights the importance of achieving and maintaining sufficiently high vaccine efficacy to suppress virus transmission effectively.

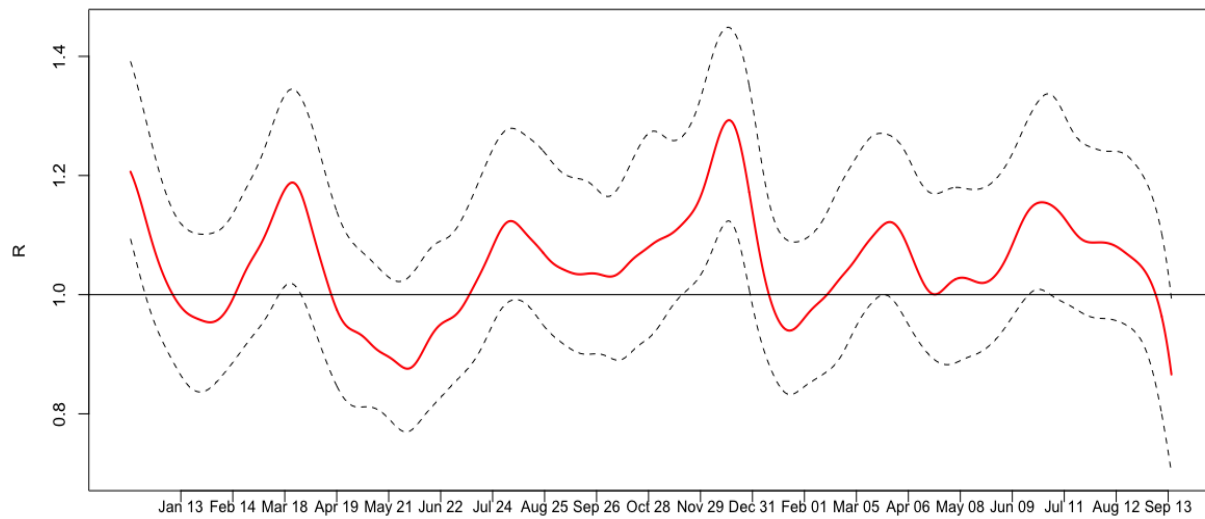


Figure 4.18: Reproduction number computed using the DLGNB model with a setting of 75% vaccine effectiveness.

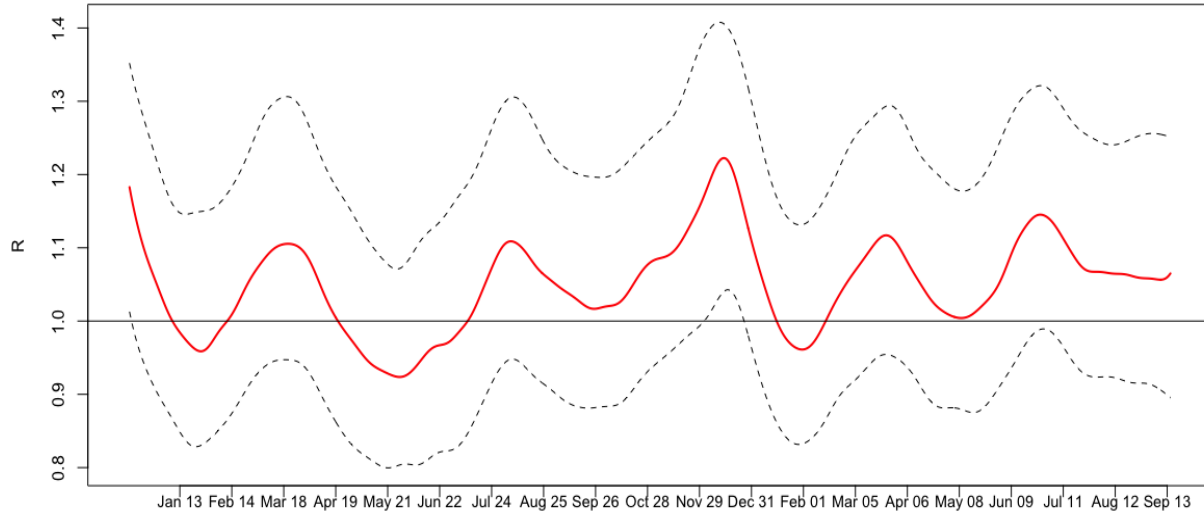


Figure 4.19: Reproduction number computed using the DLGNB model with a setting of 90% vaccine effectiveness.

These comparisons demonstrate the substantial impact of vaccination on controlling COVID-19. Higher vaccine effectiveness not only reduces the number of cases but also maintains the reproduction number below the critical threshold of 1. These findings emphasize the need for widespread vaccination campaigns and the development of highly effective vaccines. Public health strategies should focus on maximizing vaccine coverage and ensuring that vaccines administered have the highest possible effectiveness to mitigate the spread of the virus and prevent overwhelming healthcare resources.

4.4 Conclusion

In conclusion, the important features of the model are as follows. Population-level dynamics can be represented using deterministic ordinary differential equations, however, case rates vary due to local testing, reporting policies, and non-homogeneous behaviour among individuals. Community mobility impacts the spreading profile; hence, a Bayesian time-varying coefficient state-space model is implemented for infectious disease transmission. Temporal variations in epidemiological parameters are modeled using Bayesian shrinkage splines. The time-varying coefficient compartment models the dynamics among susceptible, exposed, asymptomatic, symptomatic, asymptomatic removed, hospitalized, symptomatic removed, deceased and vaccinated individuals. The infectiousness component in the model assimilates

multiple information sources, notably, population community mobility. A nonlinear and non-Gaussian process model allows for day-to-day heterogeneity in disease spread and case reporting due to reporting delays and irregularities, as well as deviations from deterministic dynamics and nonrandom mixing of the population, which is essential to recognizing the role of superspreaders [77]. The model is applied to case, hospitalization, and death observables from the province of Ontario, Canada. High correlations between mobility and transmission rates explain the complex temporal variation in infectiousness. The model presented is a rigorous, process-driven framework for modeling the impact of time-varying factors on infectious disease spread.

In terms of limitations, challenges, and extensions, the following are noted. The variable impact of mobility may be explained by differences in compliance with mask-wearing and social distancing mandates. Mobility data is obtained from individuals and then averaged at the regional level. Granular mobility patterns, further stratified by at-risk or highly active subpopulations, may reveal connections between transmission and mobility. The relation between mobility and epidemiological parameters across different locations may be used to demonstrate heterogeneous effects of different policies. Disparities in cell phone ownership due to sampling bias and usage in seniors, lower socioeconomic status, and institutionalized subpopulations, who are less likely to own a cell phone, may be poorly represented by this form of mobility data, leading to bias in the mobility metrics. In contrast to approaches that compare only time series of cases to identify marginal relationships, incorporating information on mobility into a model for disease spread allows for a more accurate representation of the potentially causal role of factors in disease spread.

Among the multiple compartments, only three contained observed data, and as several parameters determine the number of new cases produced by the model at each time point, identifiability issues arise. Modeling the detection rate for individuals transitioning from asymptomatic to symptomatic may be improved by including testing rate data. The shrinkage splines and Gaussian processes used to model the parameters are amenable to expanding the input space to include environmental factors that may influence the spread and severity of symptoms [78], [79]. The Bayesian framework may be used to fit a joint model that includes lower-level regional components and would allow for information on non-spatial parameters to be estimated using shared information. Migration patterns could further be included to account for the movement of susceptible and infected cases between regions. However, the computational complexity increases as parameters of the stochastic models must be defined for each region.

The mathematical and statistical modelling discussed provides a detailed and comprehensive approach to understanding and predicting the dynamics of infectious disease spread,

while also highlighting the importance of incorporating various data sources and addressing potential limitations and challenges in the analysis of complex real-world problems.

Chapter 5

Stochastic Threshold Analysis for SIRS model with Mean-Reverting Transmission Processes

5.1 Introduction

Epidemiological modelling plays a crucial role in understanding the spread of infectious diseases and informing public health policies. Accurate predictions of transmission rates are vital for designing effective intervention strategies and mitigating the impact of outbreaks. Traditionally, deterministic models, such as the classical Susceptible-Infectious-Recovered-Susceptible (SIRS) framework, have been used to model disease transmission [80]. While these models offer valuable insights, they fall short in capturing the inherent variability in transmission rates caused by external factors like seasonality, public health interventions, and behavioural changes within populations.

A key limitation of deterministic models is their assumption of constant or time-varying transmission rates that are often too simplistic to reflect real-world epidemic dynamics. In reality, transmission rates fluctuate unpredictably due to stochastic influences, making it essential to incorporate these random variations into modelling frameworks.

Here, we introduce a novel approach to epidemiological modelling that incorporates stochastic mean-reverting processes into the transmission dynamics by allowing the transmission rate to fluctuate according to processes such as Chan–Karolyi–Longstaff–Sanders (CKLS) [81]. This approach enables predictions of disease persistence or eradication, particularly in scenarios where external factors drive fluctuations in transmission rates.

The primary objective of this work is to present the derivation of stochastic extinction

thresholds for these mean-reverting processes and to compare their behaviour with traditional deterministic models. Through numerical simulations, we validate the performance of our stochastic models, demonstrating their ability to capture the complex, nonlinear dynamics of infectious disease transmission. Ultimately, this work aims to develop a novel framework for epidemiological modelling.

5.2 Methodology

5.2.1 Model overview

In this work, we extend the classical SIRS (Susceptible-Infectious-Recovered-Susceptible) epidemiological model by incorporating stochastic processes that allow the transmission rate to vary dynamically. Traditional deterministic models often assume constant or predefined time-varying transmission rates, limiting their ability to capture the unpredictable fluctuations observed in real-world epidemics. By introducing mean-reverting stochastic processes, we offer a framework that is more flexible and better suited to reflect the randomness inherent in disease spread.

Our model specifically incorporates the Chan–Karolyi–Longstaff–Sanders (CKLS) process, and novel mean-reverting to model transmission rate fluctuations. We also implement the Ornstein–Uhlenbeck (OU) process [82]. These stochastic processes differ in how they handle variability, volatility, and the behaviour of the transmission rate over time, particularly in maintaining its positivity. The introduction of these processes allows for a more realistic and accurate depiction of epidemic dynamics, especially when external factors, such as seasonality or interventions, cause shifts in transmission rates.

5.2.2 Key contributions

The key contributions of our approach can be summarized as follows:

- **Derivation of stochastic extinction thresholds:** One of the major theoretical contributions of this work is the derivation of extinction thresholds under various stochastic processes. These thresholds determine whether the disease will persist or die out over time, based on the dynamics of the transmission rate. Stochastic extinction thresholds are crucial because they provide a more realistic and nuanced understanding of disease eradication, considering the random fluctuations in transmission rates that deterministic models fail to capture.

- Comparison between CKLS and OU processes: We provide a comprehensive comparison between the CKLS and OU processes in the context of epidemiological modelling. The CKLS process offers certain advantages over the OU process, including the ability to maintain the positivity of the transmission rate, which is vital for epidemiological models. By comparing these two processes, we highlight the benefits and trade-offs of each in terms of capturing real-world disease dynamics. This comparison is essential for identifying the most appropriate model depending on the specific characteristics of the disease being studied.
- Introduction of novel mean-reverting processes: In addition to the CKLS and OU processes, we introduce mean-reverting Bessel processes, which provide a new perspective on transmission rate fluctuations. These processes offer an alternative for modelling mean-reverting behaviour with additional flexibility in controlling volatility. This extension represents an innovative approach to handling variability in transmission rates and contributes to the growing body of stochastic epidemiological models.
- Numerical validation through simulation: To support our theoretical findings, we present a series of numerical simulations that validate the performance of our stochastic models. These simulations demonstrate how different stochastic processes influence the long-term behaviour of the infected population and highlight the conditions under which disease outbreaks either persist or fade out. This validation is important because it provides empirical evidence that the proposed stochastic models can better capture the dynamics of infectious disease transmission compared to traditional deterministic models.
- Broader implications for epidemiological modelling: The incorporation of stochastic processes into the SIRS model has broader implications for the field of epidemiological modelling. By providing a framework that accounts for random variability in transmission rates, our work contributes to a more realistic and adaptable model that can inform public health interventions. The stochastic approach also opens up new avenues for further research, particularly in the study of diseases with high transmission rate variability or external influences.

Our work here not only advances the theoretical understanding of stochastic processes in epidemiological models but also provides practical tools for predicting and managing disease outbreaks in real-world settings. The combination of novel processes, derived extinction thresholds, and numerical validation positions this work as a significant contribution to the field.

5.2.3 Stochastic model of disease transmission

Understanding and controlling infectious diseases require accurate and dynamic modelling approaches. Traditional epidemiological models, like the SIRS model, offer valuable insights into disease spread but often overlook the inherent randomness of transmission rates influenced by environmental and social factors. To address this, stochastic models incorporate variability directly into the model structure, providing a more realistic representation of disease dynamics [83].

The classical SIRS model has been extensively studied, with various modifications to include stochastic elements, such as Brownian motion in the form of additive white noise [84], and more recently, the OU process [82], [85] and Lévy jumps [86]. However, these modifications often assume constant or linear variability in the transmission rate, which may not fully capture the complexities observed in real-world outbreaks. The CKLS process introduces non-linear stochastic variability, offering a more flexible and realistic representation of transmission dynamics. This study explores the implications of the CKLS process on the stochastic threshold and disease dynamics, filling a gap in existing literature.

In the context of modelling disease transmission, the transmission rate $\beta(t)$ is critical in determining the dynamics of an epidemic. The OU process is a notable stochastic approach, introducing stochasticity into the transmission rate and capturing random fluctuations. Alternatively, the CKLS process generalizes the OU process, providing distinct characteristics and implications for disease transmission modelling.

The CKLS process for the transmission rate $\beta(t)$ is defined as:

$$d\beta(t) = \theta(\beta_e - \beta(t))dt + \rho\beta(t)^\gamma dB(t),$$

where:

- θ is the speed of reversion,
- β_e is the long-term mean of $\beta(t)$,
- ρ is the volatility coefficient,
- $\gamma \geq 0$ determines the scaling of volatility with respect to $\beta(t)$,
- $B(t)$ represents standard Brownian motion.

The CKLS process is characterized by its mean-reverting behaviour, where the transmission rate $\beta(t)$ tends to revert to its long-term mean β_e . The volatility term $\rho\beta(t)^\gamma dB(t)$ introduces stochasticity, with γ determining the nature of the volatility. Notably, when

$\gamma = \frac{1}{2}$, the CKLS process is specialized to the Cox–Ingersoll–Ross (CIR) model which ensures that the transmission rate remains positive almost surely [87], which is crucial for modelling disease transmission since a negative transmission rate is not epidemiologically meaningful.

In comparison, the OU process for the transmission rate $\beta(t)$ is defined as:

$$d\beta(t) = \theta(\beta_e - \beta(t))dt + \rho dB(t),$$

where the terms have the same meanings as in the CKLS process.

The OU process is classified as a semimartingale [88], [89] because it can be expressed as the sum of two distinct components. The drift term, $\theta(\beta_e - \beta(t))dt$, is of finite variation and models a mean-reverting force that pulls the process toward its long-term mean β_e . Meanwhile, the stochastic term, $\rho dB(t)$, incorporates a Brownian motion component that introduces inherently unpredictable fluctuations into the system.

This characterization is important in modelling contexts, such as epidemiological dynamics. By including both a deterministic mean-reverting drift and stochastic noise from the Brownian motion, the OU process provides a flexible yet realistic framework for simulating complex changes over time. Such a framework can capture the interplay of environmental influences and random shocks that drive transmission rates, ultimately offering a richer and more accurate representation of disease processes than models relying solely on deterministic or purely random components.

Both the CKLS and OU processes share important features:

- Mean-reversion: Both processes include a reversion term $\theta(\beta_e - \beta(t))$ that drives the transmission rate toward its long-term mean β_e .
- Stochasticity: Both processes incorporate a stochastic term involving a Brownian motion $B(t)$, capturing random fluctuations in the transmission rate.

However, they differ in key aspects:

- Nature of volatility: The CKLS process features a volatility term $\rho\beta(t)^\gamma dB(t)$ that depends on the current value of $\beta(t)$, leading to multiplicative noise. In contrast, the OU process includes an additive noise term $\rho dB(t)$, which is independent of $\beta(t)$.
- Positivity constraint: The CKLS process with $\gamma = \frac{1}{2}$ ensures that $\beta(t)$ remains positive, which is essential for epidemiological modelling. The OU process does not impose this constraint, allowing the possibility of negative transmission rates.

The choice between the CKLS and OU processes for modelling the transmission rate $\beta(t)$ depends on the specific requirements of the epidemiological model. The CKLS process is suitable for models where maintaining a positive transmission rate is critical, as it can more accurately capture the proportional effects of stochastic fluctuations relative to the current transmission rate. In contrast, the OU process may be appropriate for simpler models where the possibility of negative transmission rates is not a concern, offering mathematical tractability and ease of analysis.

While both the CKLS and OU processes provide valuable frameworks for incorporating stochasticity into disease transmission models, the CKLS process offers greater flexibility and realism for epidemiological applications by ensuring positive transmission rates and capturing the proportional effects of stochastic fluctuations.

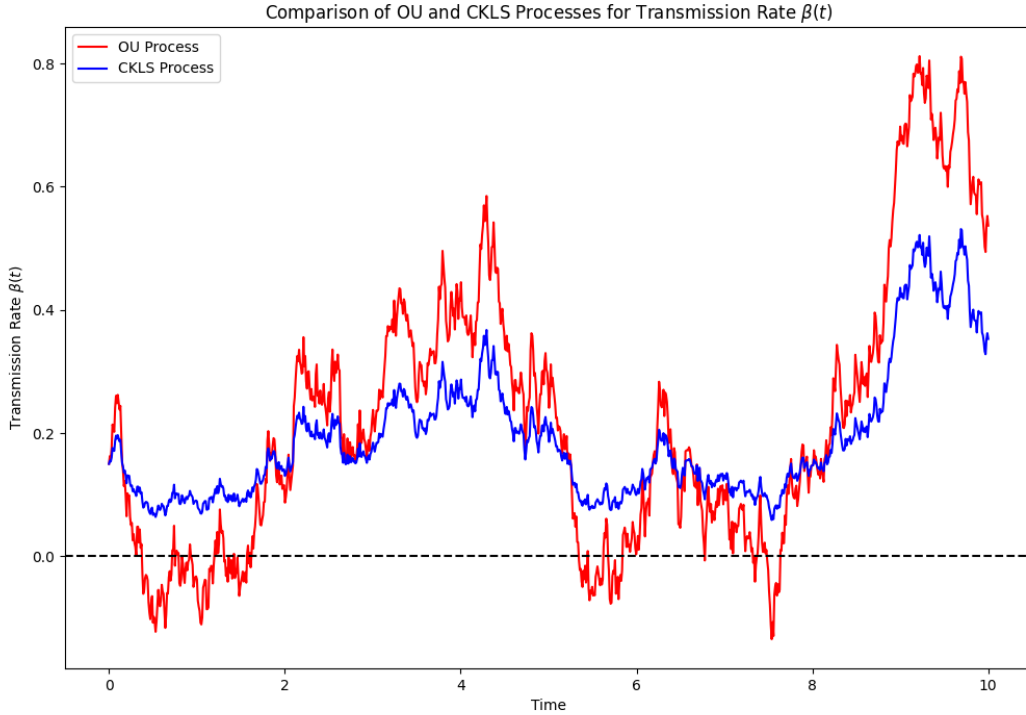


Figure 5.1: Simulation demonstrating the possibility of negative transmission rates if OU process is used. The parameter setting used are as follows: $\beta_e = 0.15$, $\theta = 0.4$, $\rho = 0.25$. The CKLS process is determined by $\gamma = \frac{1}{2}$.

5.2.4 The SIRS model

When a new disease emerges and begins to cause infections and fatalities, it is crucial for health authorities to make accurate predictions about future outbreaks. The study of infectious diseases has long been a significant area of research due to the severe impact such diseases have on humanity. In recent decades, mathematical modelling in epidemiology has garnered considerable attention within academia and public health entities. This interest aims to better understand and predict epidemic behaviours, ultimately aiding in the prevention and control of infectious diseases. Many studies have been devoted to the modelling of infectious diseases. These models aim to understand the level of manageability and the effect of prevention and control mechanisms applied to the epidemic. Such models have been shown to be essential tools in comparing the dynamics of different diseases and better understanding how the spread of a disease can be used for making vital predictions.

The Susceptible-Infectious-Recovered-Susceptible (SIRS) model is one such epidemiological model, consisting of three compartments: susceptible individuals (S), infectious individuals (I), and recovered individuals (R). Each compartment represents a distinct stage of the disease, with transitions between compartments governed by specific parameters. The SIRS model is commonly described by the following set of ordinary differential equations (ODEs):

$$\begin{cases} \frac{dS}{dt} = \Lambda - \mu S - \beta SI + \gamma_0 R, \\ \frac{dI}{dt} = \beta SI - (\mu + \lambda)I, \\ \frac{dR}{dt} = \lambda I - (\mu + \gamma_0)R, \end{cases}$$

where:

- Λ is the recruitment rate,
- μ is the natural death rate,
- β is the disease transmission rate,
- λ is the recovery rate,
- γ_0 is the rate at which recovered individuals lose immunity and return to the susceptible state.

The incidence rate, βSI , plays a critical role in the dynamics of infectious diseases. While the bilinear incidence rate is widely used, alternative non-linear incidence rates have been proposed to account for various epidemiological factors. For instance, a generalized incidence

rate of the form $\beta S f(I)$ can offer a more flexible framework, where $f(I)$ is a non-negative, twice continuously differentiable function with specific properties.

Given the stochastic nature of real-world epidemics, it is important to incorporate random perturbations into the model. One approach is to perturb the incidence rate with Gaussian white noise, resulting in a stochastic differential equation. Additionally, the disease transmission rate β can be modelled as a stochastic process, such as the CKLS process, which captures the random fluctuations in the transmission rate.

$$\begin{cases} dS = (\Lambda - \mu S - \beta S f(I) + \gamma_0 R)dt - \sigma(t) S f(I) dB(t), \\ dI = (\beta S f(I) - (\mu + \lambda) I)dt + \sigma(t) S f(I) dB(t), \\ dR = (\lambda I - (\mu + \gamma_0) R)dt, \end{cases}$$

Here, $\sigma(t)$ denotes the standard deviation of the noise.

The CKLS process ensures that $\beta(t)$ reverts to its long-term mean β_e while introducing stochasticity through the volatility term $\sigma(t) = \rho \beta(t)^\gamma$. Notably, when $\gamma = \frac{1}{2}$, the transmission rate $\beta(t)$ remains positive, which is epidemiologically relevant.

The integral form of $\beta(t)$ from the CKLS process is:

$$\beta(t) = \beta_e + (\beta_0 - \beta_e)e^{-\theta t} + \rho \int_0^t e^{-\theta(t-s)} \beta(s)^\gamma dB(s),$$

with the initial condition $\beta(0) = \beta_0$.

To incorporate the CKLS process into the SIRS model, substitute $\beta(t)$ into the infection term of the SIRS model. The stochastic SIRS model then becomes:

$$\begin{cases} dS = (\Lambda - \mu S - (\beta_e + (\beta_0 - \beta_e)e^{-\theta t}) S f(I) + \gamma_0 R)dt - \rho \beta(t)^\gamma S f(I) dB(t), \\ dI = ((\beta_e + (\beta_0 - \beta_e)e^{-\theta t}) S f(I) - (\mu + \lambda) I)dt + \rho \beta(t)^\gamma S f(I) dB(t), \\ dR = (\lambda I - (\mu + \gamma_0) R)dt, \end{cases}$$

where the noise term $\rho \beta(t)^\gamma S f(I) dB(t)$ arises from the CKLS process for $\beta(t)$.

Given that the total population $N = S + I + R$ remains constant (i.e., $N = \frac{\Lambda}{\mu}$), the model can be reduced to a 2-dimensional system of stochastic differential equations (SDEs):

$$\begin{cases} dS = \left(\Lambda - \mu S - \left(\beta_e + (\beta_0 - \beta_e)e^{-\theta t} + \rho \int_0^t e^{-\theta(t-s)} \beta(s)^\gamma dB(s) \right) S f(I) + \gamma_0 \left(\frac{\Lambda}{\mu} - S - I \right) \right) dt, \\ dI = \left(\left(\beta_e + (\beta_0 - \beta_e)e^{-\theta t} + \rho \int_0^t e^{-\theta(t-s)} \beta(s)^\gamma dB(s) \right) S f(I) - (\mu + \lambda) I \right) dt. \end{cases}$$

This reduced system explicitly captures the influences of births and deaths, infection dynamics, immunity waning, and the time-dependent transmission function governed by an Ornstein–Uhlenbeck process. Each parameter plays a specific role: Λ represents the birth rate, μ the natural death rate, λ the disease-induced death rate, γ the rate of immunity loss, and β_e, β_0, θ characterize the time variation in the transmission rate. The term $\sigma(t)$ introduces stochastic fluctuations into the infection process, and the nonlinear function $f(I)$ models how infected individuals affect the transmission dynamics.

We define two important sets to contextualize the state space of the system. Let

$$\mathbb{R}_+^2 = \{(x_1, x_2) \mid x_1 > 0, x_2 > 0\}$$

denote the positive quadrant of $\mathbb{R}^2$, and let

$$\Delta = \{(S, I) \in \mathbb{R}_+^2 : S + I \leq \frac{\Lambda}{\mu}\}$$

represent the biologically relevant domain where the total population size (susceptible plus infected) does not exceed the carrying capacity Λ/μ . This ensures that the model remains interpretable in an epidemiological sense and that the state variables $S(t)$ and $I(t)$ represent realistic population fractions over time.

In subsequent sections, we will investigate how the intensity of stochastic perturbations and the rate of mean reversion in the Ornstein–Uhlenbeck process influence the long-term dynamics of this SIRS model. Specifically, we will review some results pertaining to the existence and uniqueness of a global positive solution to the above system. With these foundational results in hand, we will then derive sufficient conditions for the disease to either go extinct or persist in mean. Finally, we will provide numerical examples to illustrate these theoretical findings and to demonstrate the practical implications of the parameters on the system’s qualitative behaviour.

5.2.5 Reproduction number of the deterministic model

The main goal is to establish sufficient conditions for extinction of the disease. To this end, the basic reproduction number [90]–[92] of the corresponding deterministic system of the SIRS model (i.e., $\rho = 0$) is given by:

$$\mathcal{R}_0 = \frac{\beta_e \Lambda f'(0)}{\mu(\mu + \lambda)}.$$

$\mathcal{R}_0$ represents the average number of secondary infections produced by one infected in-

dividual in a fully susceptible population, and it indicates the condition for disease spread in the deterministic model:

- If $\mathcal{R}_0 > 1$, the disease can spread in the population.
- If $\mathcal{R}_0 \leq 1$, the disease will eventually die out.

For the calculation of $\mathcal{R}_0$, consider the following. At the disease-free equilibrium (DFE), the infected population $I = 0$, which implies $f(I) = 0$ and the recovered population $R = 0$. The system is in equilibrium when the entire population is susceptible such that $S = \frac{\Lambda}{\mu}$.

Linearize the system around the DFE and focus on the equation describing the dynamics of I :

$$\frac{dI}{dt} = \beta S f(I) - (\mu + \lambda)I.$$

At the DFE, $S = \frac{\Lambda}{\mu}$ and consider the approximation $f(I) \approx f'(0)I$:

$$\frac{dI}{dt} = \beta \left(\frac{\Lambda}{\mu} \right) f'(0)I - (\mu + \lambda)I.$$

To ensure $\frac{dI}{dt} > 0$, set the expression equal to zero to solve for $\mathcal{R}_0$:

$$\mathcal{R}_0 = \frac{\beta \Lambda f'(0)}{\mu(\mu + \lambda)}.$$

5.2.6 Theoretical foundations

Following the approach of [82], this subsection provides the core theoretical results that guarantee the well-posedness of the stochastic SIRS model, as well as conditions for the extinction or persistence of the disease. We begin by establishing the global existence and uniqueness of positive solutions, ensuring that the model is both mathematically and biologically meaningful. With this foundational result in hand, we then consider under what circumstances the disease will almost surely die out or persist. These outcomes are governed by a stochastic threshold that extends the classical deterministic threshold, accounting for environmental noise and mean-reverting effects.

Existence and Uniqueness of Global Positive Solution

A fundamental requirement for any epidemic model is that its solutions be global, unique, and biologically interpretable. In the context of a two-dimensional stochastic SIRS system, this means showing that, starting from any non-negative initial conditions, there exists a unique

solution that remains positive and finite for all future times. Such a guarantee ensures that no explosions occur and that the state variables, representing populations, never become negative or nonphysical.

The following theorem ensures the solution remains in Δ , a set that represents the biologically meaningful state space, and thereby confirms that the model is well-posed.

Theorem 5.2.1. For any given initial value $(S(0), I(0)) \in \Delta$, there is a unique global positive solution $(S(t), I(t))$ to the 2-dimensional SDE on $t \geq 0$, and the solution remains in Δ almost surely.

This result implies that our stochastic model does not produce pathological outcomes such as negative populations or infinite values in finite time. Consequently, the model can be meaningfully interpreted within an epidemiological framework.

Extinction of the Disease

Once we have established that the model is well-defined, the next question concerns the long-term fate of the disease. Will the infection die out, or will it persist indefinitely in the population? In deterministic settings, this question is answered by examining the basic reproduction number $\mathcal{R}_0$. If $\mathcal{R}_0 \leq 1$, the disease vanishes, whereas if $\mathcal{R}_0 > 1$, it persists. In a stochastic setting, environmental noise and mean-reverting dynamics alter this threshold, resulting in a modified stochastic threshold $\tilde{\mathcal{T}}_0^S$.

If the stochastic threshold $\tilde{\mathcal{T}}_0^S \leq 1$, the disease will almost surely go extinct, potentially even when the deterministic threshold $\mathcal{R}_0$ suggests persistence. Under certain conditions, the influence of noise and slow mean reversion can drive the infection to die out, defying the intuition drawn from the deterministic model.

Theorem 5.2.2. Let $X(t) = (S(t), I(t))$ be the solution of the SDE system with initial value $X(0) \in \Delta$. If

$$\tilde{\mathcal{T}}_0^S < 1 \quad \text{and} \quad \rho^2 \frac{\Lambda}{\mu} f'(0) \leq 2\beta_e \theta,$$

or

$$\frac{\beta_e^2 \theta}{\mu + \lambda} < \rho^2,$$

then

$$\mathbb{P} \left(\limsup_{t \rightarrow \infty} \frac{\log I(t)}{t} < 0 \right) = 1.$$

Namely, the disease goes extinct exponentially almost surely.

This theorem shows that if the stochastic threshold is not greater than one, the disease eventually fades out. The terms involving ρ^2 , θ , and other parameters highlight the balance between environmental fluctuations and the model's intrinsic stabilizing forces.

For the critical case $\tilde{\mathcal{T}}_0^S = 1$, a more careful analysis is required. Although classical methods fail in this borderline scenario, it is still possible, under certain parameter conditions, to show that the disease will die out with probability one.

Theorem 5.2.3. For any initial value $(S(0), I(0)) \in \Delta$, if the conditions

$$\tilde{\mathcal{T}}_0^S = 1 \quad \text{and} \quad \rho^2 \frac{\Lambda}{\mu} f'(0) \leq 2\theta(\beta_e - \beta_0),$$

hold, then the solution obeys

$$\lim_{t \rightarrow \infty} I(t) = 0 \quad \text{a.s.},$$

which means that the disease dies out with probability 1.

This result ensures that even at the threshold, certain parameter configurations guarantee eventual extinction.

Persistence in Mean

On the other hand, if the stochastic threshold $\tilde{\mathcal{T}}_0^S > 1$, the disease persists in the population. This does not mean it settles into a fixed equilibrium, but rather that the time-averaged population of infected individuals remains strictly positive. Hence, environmental noise and mean-reversion effects are not strong enough to eradicate the disease; instead, they coexist with the infection dynamics.

Theorem 5.2.4. For any given initial value $(S(0), I(0)) \in \Delta$, if $\tilde{\mathcal{T}}_0^S > 1$, then all the following hold:

$$\liminf_{t \rightarrow +\infty} \frac{1}{t} \int_0^t S(s) ds \geq \frac{\Lambda}{\left(\mu + (\beta_e \vee \beta_0) \frac{\Lambda}{\mu} f'(0) \right)} \quad \text{a.s.},$$

$$\liminf_{t \rightarrow +\infty} \frac{1}{t} \int_0^t I(s) ds \geq m_1(\mu + \lambda) \left(\tilde{\mathcal{T}}_0^S - 1 \right) \quad \text{a.s.},$$

for some positive constant m_1 .

This result implies that once the stochastic threshold surpasses one, both susceptible and infected compartments maintain a strictly positive average presence, guaranteeing that the disease remains endemic in the long run.

These theoretical foundations detail how the relationship between deterministic factors and stochastic influences shapes the fate of an epidemic. The existence and uniqueness

theorem ensures that the model operates in a biologically sensible realm. The extinction theorems show that under certain conditions, noise and mean-reversion can help eradicate the disease. Conversely, if the stochastic threshold rises above unity, the disease's persistence in mean highlights its resilience despite random fluctuations. Together, these results form a rigorous basis for understanding the complex long-term behaviour of epidemics in a stochastic environment.

5.3 Derivation of the stochastic threshold for mean-reverting processes

5.3.1 Stochastic integral and variance analysis

The stochastic threshold $\tilde{\mathcal{T}}_0^S$ depends on the stochastic process assumed for $\beta(t)$. Below, we derive the threshold for both the OU process and the general CKLS process.

Ornstein–Uhlenbeck (OU) process

The OU process is given by the stochastic differential equation (SDE):

$$d\beta(t) = \theta(\beta_e - \beta(t)) dt + \rho dB(t),$$

where:

- $\theta > 0$ is the mean reversion rate,
- β_e is the long-term mean,
- $\rho > 0$ is the volatility coefficient,
- $dB(t)$ represents a standard Brownian motion.

The solution to this SDE using the integrating factor method is:

$$\beta(t) = \beta(0)e^{-\theta t} + \beta_e(1 - e^{-\theta t}) + \rho e^{-\theta t} \int_0^t e^{\theta s} dB(s).$$

The stochastic integral term is:

$$J(t) = \rho e^{-\theta t} \int_0^t e^{\theta s} dB(s).$$

The variance of this integral, using Itô's isometry, is:

$$\text{Var}(J(t)) = \rho^2 e^{-2\theta t} \int_0^t e^{2\theta s} ds = \frac{\rho^2}{2\theta} (1 - e^{-2\theta t}).$$

As $t \rightarrow \infty$, the variance approaches:

$$\lim_{t \rightarrow \infty} \text{Var}(J(t)) = \frac{\rho^2}{2\theta}.$$

This represents the steady-state variance of the OU process, demonstrating its mean-reverting property.

Chan–Karolyi–Longstaff–Sanders (CKLS) process

The CKLS process is described by the SDE:

$$d\beta(t) = \theta(\beta_e - \beta(t)) dt + \rho\beta(t)^\gamma dB(t),$$

where:

- $\theta > 0$ is the mean reversion rate,
- β_e is the long-term mean,
- $\rho > 0$ is the volatility coefficient,
- $\gamma \geq 0$ determines the scaling of volatility with respect to $\beta(t)$,
- $dB(t)$ represents a standard Brownian motion.

The integral form of $\beta(t)$ under the CKLS process is:

$$\beta(t) = \beta(0)e^{-\theta t} + \beta_e(1 - e^{-\theta t}) + \rho e^{-\theta t} \int_0^t e^{\theta s} \beta(s)^\gamma dB(s).$$

The stochastic integral term is:

$$J_\gamma(t) = \rho e^{-\theta t} \int_0^t e^{\theta s} \beta(s)^\gamma dB(s).$$

The variance of this integral, again using Itô's isometry, is:

$$\text{Var}(J_\gamma(t)) = \rho^2 e^{-2\theta t} \int_0^t e^{2\theta s} \mathbb{E}[\beta(s)^{2\gamma}] ds.$$

In the long-term steady state, assuming stationarity, this variance can be expressed as:

$$\lim_{t \rightarrow \infty} \text{Var}(J_\gamma(t)) = \frac{\rho^2}{2\theta} \lim_{t \rightarrow \infty} \mathbb{E}[\beta(t)^{2\gamma}].$$

This expression indicates that the variance depends on the expectation, $\mathbb{E}[\beta(t)^{2\gamma}]$, which may be derived from the stationary distribution of $\beta(t)$ in the CKLS process. Accurate parameter estimation in the CKLS model is crucial for effective disease transmission modelling [93].

5.3.2 Derivation strategy for extinction threshold

To derive the extinction threshold, we focus on the infected population $I(t)$. Applying Itô's lemma to the logarithm of $I(t)$ helps in analyzing the growth rate of $I(t)$ over time.

Taking the differential of $\log I(t)$:

$$d(\log I(t)) = \frac{1}{I(t)} dI(t) - \frac{1}{2I(t)^2} (dI(t))^2.$$

Substituting the expression for $dI(t)$ from the SDE:

$$\begin{aligned} d(\log I(t)) &= \frac{1}{I(t)} \left[\left((\beta_e + (\beta_0 - \beta_e)e^{-\theta t}) S(t) f(I(t)) - (\mu + \lambda) I(t) \right) dt + \sigma(t) S(t) f(I(t)) dB(t) \right] \\ &\quad - \frac{1}{2I(t)^2} [\sigma(t) S(t) f(I(t))]^2 dt \\ &= \left(\frac{(\beta_e + (\beta_0 - \beta_e)e^{-\theta t}) S(t) f(I(t))}{I(t)} - (\mu + \lambda) - \frac{1}{2} \sigma(t)^2 S(t)^2 \frac{f(I(t))^2}{I(t)^2} \right) dt \\ &\quad + \sigma(t) \frac{S(t) f(I(t))}{I(t)} dB(t). \end{aligned}$$

5.3.3 Long-term behaviour

Interest lies in the long-term behaviour of $\log I(t)$. Taking the expectation and dividing by t :

$$\frac{\mathbb{E}[\log I(t)]}{t} = \frac{\log I(0)}{t} + \frac{1}{t} \int_0^t \mathbb{E} \left[\frac{(\beta_e + (\beta_0 - \beta_e)e^{-\theta s}) S(s) f(I(s))}{I(s)} - (\mu + \lambda) - \frac{1}{2} \sigma(s)^2 S(s)^2 \frac{f(I(s))^2}{I(s)^2} \right] ds.$$

As t becomes large, the first term $\frac{\log I(0)}{t}$ tends to 0, so:

$$\lim_{t \rightarrow \infty} \frac{\mathbb{E}[\log I(t)]}{t} = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \mathbb{E} \left[\frac{(\beta_e + (\beta_0 - \beta_e)e^{-\theta s}) S(s) f(I(s))}{I(s)} - (\mu + \lambda) - \frac{1}{2} \sigma(s)^2 S(s)^2 \frac{f(I(s))^2}{I(s)^2} \right] ds.$$

5.3.4 Simplification using equilibrium values

Assuming $S(t)$ and $I(t)$ are near their equilibrium values, approximate $S(t) \approx \frac{\Lambda}{\mu}$ and $f(I) \approx f'(0)I$. Substituting these into the expression:

$$\frac{\mathbb{E}[\log I(t)]}{t} \approx \frac{\beta_e \Lambda f'(0)}{\mu} - (\mu + \lambda) - \frac{1}{2} \sigma(t)^2 \left(\frac{\Lambda}{\mu} f'(0) \right)^2.$$

This represents the average growth rate of the logarithm of the infected population over time, determining whether the infected population $I(t)$ will grow or decay.

5.3.5 Deriving the extinction threshold

The critical threshold $\tilde{\mathcal{T}}_0^S$ is derived by setting:

$$\frac{\mathbb{E}[\log I(t)]}{t} = 0,$$

which defines the point at which the infected population neither grows nor shrinks on average.

- If $\lim_{t \rightarrow \infty} \frac{\mathbb{E}[\log I(t)]}{t} < 0$, the infected population $I(t)$ decreases, leading to disease extinction.
- If $\lim_{t \rightarrow \infty} \frac{\mathbb{E}[\log I(t)]}{t} > 0$, the infected population $I(t)$ increases, leading to disease persistence.

The expression for the expected growth rate of the logarithm of the infected population $I(t)$ over time is:

$$\frac{\mathbb{E}[\log I(t)]}{t} \approx \frac{\beta_e \Lambda f'(0)}{\mu} - (\mu + \lambda) - \frac{1}{2} \sigma(t)^2 \left(\frac{\Lambda}{\mu} f'(0) \right)^2.$$

To find the threshold condition, we set the expected growth rate to zero:

$$\frac{\beta_e \Lambda f'(0)}{\mu} - (\mu + \lambda) - \frac{1}{2} \sigma(t)^2 \left(\frac{\Lambda}{\mu} f'(0) \right)^2 = 0.$$

Solving for $\tilde{\mathcal{T}}_0^S$ by dividing by $(\mu + \lambda)$, we obtain:

$$\boxed{\tilde{\mathcal{T}}_0^S = \frac{\beta_e \Lambda f'(0)}{\mu(\mu + \lambda)} - \frac{\sigma^2(t)}{2(\mu + \lambda)} \left(\frac{\Lambda}{\mu} f'(0) \right)^2}.$$

Here, $\tilde{\mathcal{T}}_0^S$ is a generic formula where the function stochastic component $\sigma^2(t)$ is determined by the stochastic process chosen for $\beta(t)$.

From the variance analysis section for the Ornstein–Uhlenbeck process, $\beta(t)$ has constant variance $\sigma_{\text{OU}} = \frac{\rho^2}{2\theta}$. The stochastic threshold becomes:

$$\tilde{\mathcal{T}}_0^{\text{OU}} = \mathcal{R}_0 - \frac{\rho^2}{4\theta(\mu + \lambda)} \left(\frac{\Lambda}{\mu} f'(0) \right)^2,$$

where $\mathcal{R}_0 = \frac{\beta_e \Lambda f'(0)}{\mu(\mu + \lambda)}$.

For the CKLS process, $\sigma_{\text{CKLS}}(t) = \frac{\rho^2}{2\theta} \times \mathbb{E}[\beta(t)^{2\gamma}]$ depends on $\beta(t)$, and the expectation $\mathbb{E}[\beta(t)^{2\gamma}]$ is used:

$$\tilde{\mathcal{T}}_0^{\text{CKLS}} = \mathcal{R}_0 - \frac{\rho^2}{4\theta(\mu + \lambda)} \left(\frac{\Lambda}{\mu} f'(0) \right)^2 \times \mathbb{E}[\beta(t)^{2\gamma}].$$

The value of $\mathbb{E}[\beta(t)^{2\gamma}]$ is later derived from the stationary distribution of the CKLS process.

Here, the form of $\tilde{\mathcal{T}}_0^S$ depends on the chosen stochastic process for $\beta(t)$, which in turn determines the function $\sigma^2(t)$. This flexibility allows for a comprehensive analysis of disease extinction conditions based on different modelling assumptions on $\beta(t)$, for example, mean-reversion and nature of volatility.

5.4 Preliminary results for CKLS process

Here, we work toward an explicit formula for the stochastic threshold under a transmission rate which follows the CKLS process. To this end, the goal is to calculate the required expectation in the $\sigma_{\text{CKLS}}(t)$ term.

5.4.1 CKLS process and Fokker–Planck equation

The Fokker–Planck equation for the probability density function $p(\beta, t)$ of the CKLS process is:

$$\frac{\partial p(\beta, t)}{\partial t} = -\frac{\partial}{\partial \beta} [\theta(\beta_e - \beta)p(\beta, t)] + \frac{1}{2} \frac{\partial^2}{\partial \beta^2} [\rho^2 \beta^{2\gamma} p(\beta, t)]. \quad (5.4.1)$$

The Fokker–Planck equation describes the time evolution of the probability density function of a stochastic process. The first term on the right-hand side represents the drift term, which causes the process to revert to the long-term mean β_e . The second term represents the diffusion term, which adds randomness based on the current value of $\beta(t)$.

5.4.2 Stationary distribution of the CKLS process

To find the stationary distribution $p_{st}(\beta)$, set $\frac{\partial p(\beta, t)}{\partial t} = 0$:

$$0 = -\frac{\partial}{\partial \beta}[\theta(\beta_e - \beta)p_{st}(\beta)] + \frac{1}{2}\frac{\partial^2}{\partial \beta^2}[\rho^2 \beta^{2\gamma} p_{st}(\beta)]. \quad (5.4.2)$$

This simplifies to:

$$\frac{\partial}{\partial \beta}[\theta(\beta_e - \beta)p_{st}(\beta)] = \frac{1}{2}\frac{\partial^2}{\partial \beta^2}[\rho^2 \beta^{2\gamma} p_{st}(\beta)]. \quad (5.4.3)$$

Assuming a solution of the form $p_{st}(\beta) = C\beta^a e^{-b\beta}$, where C , a , and b are constants, we determine a and b by equating the coefficients of the highest-order terms.

While the above description may indicate an approximation, it is actually part of a standard method for solving such differential equations. By proposing a solution of the form $p_{st}(\beta) = C\beta^a e^{-b\beta}$, we leverage the fact that the Fokker–Planck equation is a linear second-order ordinary differential equation in β . Substituting this ansatz back into the equation and matching terms at each order of β yields exact conditions for a and b . This is not merely a high-order approximation step, but a direct technique for deriving the exact stationary solution.

First, consider the left-hand side:

$$\frac{\partial}{\partial \beta}[\theta(\beta_e - \beta)C\beta^a e^{-b\beta}] = \frac{\partial}{\partial \beta}[\theta C\beta^a(\beta_e - \beta)e^{-b\beta}]. \quad (5.4.4)$$

Expanding and differentiating:

$$\theta C \left[\beta_e \frac{\partial}{\partial \beta}(\beta^a e^{-b\beta}) - \frac{\partial}{\partial \beta}(\beta^{a+1} e^{-b\beta}) \right]. \quad (5.4.5)$$

For the right-hand side:

$$\frac{1}{2}\frac{\partial^2}{\partial \beta^2}[\rho^2 \beta^{2\gamma} C\beta^a e^{-b\beta}] = \frac{1}{2}\rho^2 C \frac{\partial^2}{\partial \beta^2}[\beta^{a+2\gamma} e^{-b\beta}]. \quad (5.4.6)$$

By equating the coefficients of the highest-order terms:

$$a + 2\gamma - 1 = 0 \implies a = 2\gamma - 1, \quad (5.4.7)$$

and

$$b = \frac{2\theta}{\rho^2}. \quad (5.4.8)$$

Thus, the stationary distribution is:

$$p_{st}(\beta) = C\beta^{2\gamma-1}e^{-\frac{2\theta}{\rho^2}\beta}. \quad (5.4.9)$$

The normalization constant C is determined by ensuring the total probability integrates to 1:

$$\int_0^\infty p_{st}(\beta)d\beta = 1. \quad (5.4.10)$$

Using the Gamma function to normalize $p_{st}(\beta)$:

$$C \int_0^\infty \beta^{2\gamma-1}e^{-\frac{2\theta}{\rho^2}\beta}d\beta = 1. \quad (5.4.11)$$

Recognizing this as the integral of a Gamma distribution:

$$C \left(\frac{\rho^2}{2\theta}\right)^{2\gamma} \Gamma(2\gamma) = 1 \implies C = \frac{\left(\frac{2\theta}{\rho^2}\right)^{2\gamma}}{\Gamma(2\gamma)}. \quad (5.4.12)$$

Thus, the stationary distribution is:

$$p_{st}(\beta) = \frac{\left(\frac{2\theta}{\rho^2}\right)^{2\gamma}}{\Gamma(2\gamma)}\beta^{2\gamma-1}e^{-\frac{2\theta}{\rho^2}\beta}. \quad (5.4.13)$$

The stationary distribution $p_{st}(\beta)$ represents the long-term behaviour of the process $\beta(t)$. By solving the differential equation and normalizing the distribution using the Gamma function, we ensure that $p_{st}(\beta)$ integrates to 1, making it a valid probability density function.

5.4.3 Distribution from Girsanov transform

This section provides a detailed derivation to show the equivalence between the stationary distribution of the CKLS process derived using the Fokker–Planck equation and the distribution obtained through the Girsanov transform [94]. The goal is to demonstrate that both approaches describe the same underlying process's long-term behaviour.

Following the methodology and results in [94], under the Girsanov transform, the CKLS process is converted into a CIR-like process. The transformed process $Y_t = f(\beta_t)$ given by:

$$f(\beta) = \frac{C^2}{4(1-\gamma)^2}\beta^{2(1-\gamma)}, \quad (5.4.14)$$

satisfies:

$$dY_t = \left(\frac{\sigma^2 C^2}{4} + 2b(1 - \gamma)Y_t \right) dt + \sigma C \sqrt{Y_t} d\bar{B}_t, \quad (5.4.15)$$

where $\bar{B}_t$ is a Brownian motion under the transformed measure Q , with $C = \frac{2(1-\gamma)}{\rho}$.

The density function $g_t(x)$ for the transformed process Y_t (a non-central chi-square distribution) is:

$$g_t(x) = g_{\delta, \zeta} \left(\frac{f(x)}{L} \right) \left| \frac{f'(x)}{L} \right|, \quad (5.4.16)$$

where $g_{\delta, \zeta}$ is the density of the non-central chi-square distribution with parameters δ and ζ .

To show equivalence, we relate the stationary distribution obtained via the Fokker–Planck equation to the distribution derived from the Girsanov transform.

Given the transformation:

$$Y = f(\beta) = \frac{C^2}{4(1 - \gamma)^2} \beta^{2(1-\gamma)}, \quad (5.4.17)$$

we find the inverse:

$$\beta = f^{-1}(Y) = \left(\frac{4(1 - \gamma)^2}{C^2} Y \right)^{\frac{1}{2(1-\gamma)}}. \quad (5.4.18)$$

The density transformation rule between β and Y is:

$$p_{st}(\beta) = g_{st}(Y) \left| \frac{dY}{d\beta} \right|^{-1}. \quad (5.4.19)$$

We compute $\frac{dY}{d\beta}$:

$$\frac{dY}{d\beta} = \frac{C^2}{4(1 - \gamma)^2} \cdot 2(1 - \gamma) \beta^{2(1-\gamma)-1} = \frac{C^2}{2(1 - \gamma)} \beta^{2(1-\gamma)-1}. \quad (5.4.20)$$

The inverse derivative is:

$$\left| \frac{dY}{d\beta} \right|^{-1} = \frac{2(1 - \gamma)}{C^2} \beta^{1-2(1-\gamma)} = \frac{2(1 - \gamma)}{C^2} \beta^{2\gamma-1}. \quad (5.4.21)$$

Substitute Y in terms of β and the derivative:

$$g_{st}(Y) \left| \frac{dY}{d\beta} \right|^{-1} = \left(g_{\delta, \zeta} \left(\frac{\frac{C^2}{4(1-\gamma)^2} \beta^{2(1-\gamma)}}{L} \right) \cdot \left| \frac{\frac{C^2}{2(1-\gamma)} \beta^{2(1-\gamma)-1}}{L} \right| \right) \cdot \frac{2(1-\gamma)}{C^2} \beta^{2\gamma-1}. \quad (5.4.22)$$

Given $g_{\delta, \zeta}(x)$ is the density of a non-central chi-square, and as $Y \rightarrow \infty$:

$$g_{st}(Y) \sim \frac{1}{2} e^{-Y/2}. \quad (5.4.23)$$

Converting back to β :

$$p_{st}(\beta) = \frac{1}{2} e^{-\frac{\frac{C^2}{4(1-\gamma)^2} \beta^{2(1-\gamma)}}{2}} \cdot \frac{2(1-\gamma)}{C^2} \beta^{2\gamma-1}. \quad (5.4.24)$$

We recognize that

$$e^{-\frac{\beta^{2(1-\gamma)} C^2}{8(1-\gamma)^2}} = e^{-\frac{2\theta\beta}{\rho^2}}, \quad (5.4.25)$$

since,

$$\frac{C^2}{8(1-\gamma)^2} = \frac{2\theta}{\rho^2}. \quad (5.4.26)$$

With $C = \frac{2(1-\gamma)}{\rho}$,

$$\frac{\left(\frac{2(1-\gamma)}{\rho} \right)^2}{8(1-\gamma)^2} = \frac{2}{\rho^2} = \frac{2\theta}{\rho^2}. \quad (5.4.27)$$

Thus,

$$p_{st}(\beta) = C \beta^{2\gamma-1} e^{-\frac{2\theta\beta}{\rho^2}}, \quad (5.4.28)$$

confirming the equivalence.

5.4.4 Expectation calculation for the CKLS process

Given the stationary distribution, compute the required expectation:

$$\mathbb{E}[\beta(t)^{2\gamma}] = \int_0^\infty \beta^{2\gamma} p_{st}(\beta) d\beta = \int_0^\infty \beta^{2\gamma} \frac{\left(\frac{2\theta}{\rho^2} \right)^{2\gamma}}{\Gamma(2\gamma)} \beta^{2\gamma-1} e^{-\frac{2\theta}{\rho^2} \beta} d\beta. \quad (5.4.29)$$

This simplifies to:

$$\mathbb{E}[\beta(t)^{2\gamma}] = \frac{\left(\frac{2\theta}{\rho^2}\right)^{2\gamma}}{\Gamma(2\gamma)} \int_0^\infty \beta^{4\gamma-1} e^{-\frac{2\theta}{\rho^2}\beta} d\beta. \quad (5.4.30)$$

Recognizing this as a gamma distribution integral:

$$\mathbb{E}[\beta(t)^{2\gamma}] = \frac{\left(\frac{2\theta}{\rho^2}\right)^{2\gamma} \Gamma(4\gamma)}{\left(\frac{2\theta}{\rho^2}\right)^{4\gamma} \Gamma(2\gamma)} = \frac{\Gamma(4\gamma)}{\Gamma(2\gamma)} \left(\frac{\rho^2}{2\theta}\right)^{2\gamma}. \quad (5.4.31)$$

The expectation $\mathbb{E}[\beta(t)^{2\gamma}]$ is computed using the stationary distribution. By recognizing the integral as a gamma distribution, properties of the gamma function can be directly used to simplify the expectation.

5.5 Stochastic threshold for the CKLS process

The formula for the stochastic threshold $\tilde{\mathcal{T}}_0^{\text{CKLS}}$ for the CKLS process is given by:

$$\tilde{\mathcal{T}}_0^{\text{CKLS}} = \mathcal{R}_0 - \frac{\rho^2}{4\theta(\mu + \lambda)} \left(\frac{\Lambda}{\mu} f'(0)\right)^2 \times \mathbb{E}[\beta(t)^{2\gamma}],$$

where the expectation $\mathbb{E}[\beta(t)^{2\gamma}]$ is:

$$\mathbb{E}[\beta(t)^{2\gamma}] = \frac{\Gamma(4\gamma)}{\Gamma(2\gamma)} \left(\frac{\rho^2}{2\theta}\right)^{2\gamma}. \quad (5.5.1)$$

Substituting this expression into the formula for $\tilde{\mathcal{T}}_0^{\text{CKLS}}$, we get:

$$\tilde{\mathcal{T}}_0^{\text{CKLS}} = \mathcal{R}_0 - \frac{\rho^2}{4\theta(\mu + \lambda)} \left(\frac{\Lambda}{\mu} f'(0)\right)^2 \frac{\Gamma(4\gamma)}{\Gamma(2\gamma)} \left(\frac{\rho^2}{2\theta}\right)^{2\gamma}.$$

Simplifying this expression further:

$$\tilde{\mathcal{T}}_0^{\text{CKLS}} = \mathcal{R}_0 - \frac{\Gamma(4\gamma)}{\Gamma(2\gamma)} \frac{\rho^{2(1+2\gamma)}}{4\theta^{1+2\gamma}(\mu + \lambda)} \left(\frac{\Lambda}{\mu} f'(0)\right)^2.$$

Thus, the final result for the stochastic threshold $\tilde{\mathcal{T}}_0^{\text{CKLS}}$ for the CKLS process is:

$$\boxed{\tilde{\mathcal{T}}_0^{\text{CKLS}} = \mathcal{R}_0 - \frac{\Gamma(4\gamma)}{\Gamma(2\gamma)} \frac{\rho^{2(1+2\gamma)}}{4\theta^{1+2\gamma}(\mu + \lambda)} \left(\frac{\Lambda}{\mu} f'(0)\right)^2}.$$

5.5.1 Stochastic threshold for the CIR process

When $\gamma = \frac{1}{2}$, the CKLS process is specialized to the Cox–Ingersoll–Ross (CIR) process [87]. Substituting $\gamma = \frac{1}{2}$ into the general formula:

$$\tilde{\mathcal{T}}_0^{\text{CIR}} = \mathcal{R}_0 - \frac{\rho^4}{4\theta^2(\mu + \lambda)} \left(\frac{\Lambda}{\mu} f'(0) \right)^2.$$

The stochastic threshold $\tilde{\mathcal{T}}_0^S$ takes different forms depending on the stochastic process chosen for $\beta(t)$. For the OU process, it depends on the constant variance $\frac{\rho^2}{2\theta}$. For the more general CKLS process, it depends on the expectation $\mathbb{E}[\beta(t)^{2\gamma}]$.

5.6 Additional mean-reverting diffusion processes

The processes which are introduced in subsequent section are based on well-known classes of stochastic diffusions and transformations thoroughly documented in the literature. For example, the Cox–Ingersoll–Ross (CIR) process [87] is a classical mean-reverting diffusion closely related to the squared Bessel process [95], and Doob’s h -transform is a standard technique for constructing new Markov processes from well-studied ones [96]. Thus, the approach taken by combining existing results and known transformations is firmly grounded in established theory.

5.6.1 Mean-reverting Bessel process

The stochastic differential equation (SDE) for the mean-reverting Bessel process is:

$$d\beta(t) = \left(\theta(\beta_e - \beta(t)) + \frac{\nu - 1}{2\beta(t)} \right) dt + \rho dB(t),$$

where:

- $\theta > 0$ is the mean reversion rate,
- β_e is the long-term mean (corrected for the Bessel drift),
- $\rho > 0$ is the volatility coefficient,
- $\nu > 1$ is the dimension parameter of the Bessel process,
- $dB(t)$ represents a standard Brownian motion.

Fokker–Planck equation

The Fokker–Planck equation corresponding to this SDE is:

$$\frac{\partial p(\beta, t)}{\partial t} = -\frac{\partial}{\partial \beta} \left[\left(\theta(\beta_e - \beta) + \frac{\nu - 1}{2\beta} \right) p(\beta, t) \right] + \frac{1}{2} \frac{\partial^2}{\partial \beta^2} (\rho^2 p(\beta, t)).$$

At steady state, $\frac{\partial p(\beta, t)}{\partial t} = 0$:

$$0 = -\frac{\partial}{\partial \beta} \left[\left(\theta(\beta_e - \beta) + \frac{\nu - 1}{2\beta} \right) p_{st}(\beta) \right] + \frac{1}{2} \frac{\partial^2}{\partial \beta^2} (\rho^2 p_{st}(\beta)).$$

Solving for the stationary distribution

To find the stationary distribution $p_{st}(\beta)$, we solve:

$$\frac{\partial}{\partial \beta} \left[\left(\theta(\beta_e - \beta) + \frac{\nu - 1}{2\beta} \right) p_{st}(\beta) \right] = \frac{\rho^2}{2} \frac{\partial^2}{\partial \beta^2} p_{st}(\beta).$$

Assume a solution of the form $p_{st}(\beta) = C\beta^a e^{-b\beta}$. Substitute into the differential equation and match coefficients:

$$a - 1 = \frac{\nu - 1}{2}, \quad b = \frac{2\theta}{\rho^2}.$$

Thus:

$$a = \frac{\nu + 1}{2}, \quad b = \frac{2\theta}{\rho^2}.$$

Stationary distribution

The stationary distribution is:

$$p_{st}(\beta) = C\beta^{\frac{\nu+1}{2}} e^{-\frac{2\theta\beta}{\rho^2}}.$$

To normalize $p_{st}(\beta)$:

$$C \int_0^\infty \beta^{\frac{\nu+1}{2}} e^{-\frac{2\theta\beta}{\rho^2}} d\beta = 1.$$

This integral is a Gamma distribution form:

$$C \left(\frac{\rho^2}{2\theta} \right)^{\frac{\nu+1}{2}} \Gamma \left(\frac{\nu+1}{2} \right) = 1,$$

giving:

$$C = \frac{\left(\frac{2\theta}{\rho^2}\right)^{\frac{\nu+1}{2}}}{\Gamma\left(\frac{\nu+1}{2}\right)}.$$

So:

$$p_{st}(\beta) = \frac{\left(\frac{2\theta}{\rho^2}\right)^{\frac{\nu+1}{2}}}{\Gamma\left(\frac{\nu+1}{2}\right)} \beta^{\frac{\nu-1}{2}} e^{-\frac{2\theta\beta}{\rho^2}}.$$

Stochastic threshold for the mean-reverting Bessel process

Since the noise perturbation for the mean-reverting Bessel (MRB) process, $\rho dB(t)$, is identical to that of the OU process, the stochastic threshold is also the same. From the variance analysis section, the OU process, $\beta(t)$, has constant variance $\sigma_{OU} = \frac{\rho^2}{2\theta}$. The stochastic threshold is:

$$\tilde{\mathcal{T}}_0^{\text{MRB}} = \mathcal{R}_0 - \frac{\rho^2}{4\theta(\mu + \lambda)} \left(\frac{\Lambda}{\mu} f'(0) \right)^2,$$

where $\mathcal{R}_0 = \frac{\beta_e \Lambda f'(0)}{\mu(\mu + \lambda)}$.

Derivation via Doob's h -transform

We aim to derive a mean-reverting Bessel process that stays positive and has a target long-run mean by applying Doob's h -transform [96]–[98] to the OU process.

The OU process $\beta(t)$ is defined by the SDE:

$$d\beta(t) = \theta(\beta_e - \beta(t)) dt + \rho dB(t),$$

where $\theta > 0$, β_e is the long-term mean, $\rho > 0$ is the volatility, and $dB(t)$ represents standard Brownian motion.

The generator $\mathcal{L}$ of the OU process is:

$$\mathcal{L}f(x) = \theta(\beta_e - x)f'(x) + \frac{\rho^2}{2}f''(x).$$

To transform the OU process into a positive mean-reverting process, we choose $h(x) = x^{\nu-1}$ for some $\nu > 1$, which ensures positivity.

The transformed generator $\mathcal{L}^h$ is:

$$\mathcal{L}^h f(x) = \frac{1}{h(x)} \mathcal{L}(h(x)f(x)) - \frac{f(x)}{h(x)} \mathcal{L}h(x).$$

First, compute $\mathcal{L}(x^{\nu-1})$:

$$\mathcal{L}(x^{\nu-1}) = \theta(\beta_e - x)(\nu - 1)x^{\nu-2} + \frac{\rho^2}{2}(\nu - 1)(\nu - 2)x^{\nu-3}.$$

Next, compute $\mathcal{L}(x^{\nu-1}f(x))$:

$$\begin{aligned} \mathcal{L}(x^{\nu-1}f(x)) &= \theta(\beta_e - x) \left((\nu - 1)x^{\nu-2}f(x) + x^{\nu-1}f'(x) \right) \\ &\quad + \frac{\rho^2}{2} \left((\nu - 1)(\nu - 2)x^{\nu-3}f(x) + 2(\nu - 1)x^{\nu-2}f'(x) + x^{\nu-1}f''(x) \right). \end{aligned}$$

Substitute these into $\mathcal{L}^h f(x)$:

$$\mathcal{L}^h f(x) = \theta(\beta_e - x)f'(x) + \frac{\rho^2}{2}f''(x) + \left(\theta \frac{(\nu - 1)(\beta_e - x)}{x} + \frac{\rho^2(\nu - 1)}{x} - \frac{\rho^2(\nu - 1)(\nu - 2)}{2x^2} \right) f(x).$$

The corresponding SDE for the transformed process $Y(t)$ is derived by matching the form of the generator $\mathcal{L}^h$ to the form of an SDE. The first two terms, $\theta(\beta_e - x)f'(x)$ and $\frac{\rho^2}{2}f''(x)$, correspond to the drift and diffusion terms of the original OU process. The additional terms arise from the h -transform, and these terms adjust the drift to enforce positivity.

Among these additional terms, we retain the term $\frac{\nu-1}{2Y(t)}$, as it ensures that the process stays positive. This term becomes large as $Y(t)$ approaches zero, providing a repelling force that prevents the process from becoming negative, similar to a Bessel process. The other terms in the transformed generator, $\frac{\rho^2(\nu-1)}{x}$ and $-\frac{\rho^2(\nu-1)(\nu-2)}{2x^2}$, introduce higher-order corrections, but their contribution is less significant for ensuring positivity. These terms tend to be negligible as x moves away from zero, and they do not substantially affect the mean-reverting behaviour of the process.

Thus, the corresponding SDE is:

$$dY(t) = \left(\theta(\mu - Y(t)) + \frac{\nu - 1}{2Y(t)} \right) dt + \rho dB(t),$$

where $\mu = \beta_e - \frac{\nu-1}{2\theta\beta_e}$.

The term $\frac{\nu-1}{2Y(t)}$ ensures that the process stays positive, while the $\theta(\mu - Y(t))$ term ensures mean reversion towards the long-run mean μ . The process retains the same volatility ρ as the original OU process.

To ensure the mean-reverting Bessel process has the desired long-run mean μ_{target} , we adjust μ based on the drift term. The drift term in the SDE is:

$$\theta(\mu - Y(t)) + \frac{\nu - 1}{2Y(t)}.$$

At steady state, the expected drift should be zero, i.e., when $Y(t) = \mu_{\text{target}}$, the drift term vanishes:

$$\theta(\mu - \mu_{\text{target}}) + \frac{\nu - 1}{2\mu_{\text{target}}} = 0.$$

Solving for μ , we get:

$$\mu = \mu_{\text{target}} - \frac{\nu - 1}{2\theta\mu_{\text{target}}}.$$

This ensures the mean-reverting behaviour of the process aligns with the desired long-run mean μ_{target} .

5.6.2 Modified Bessel process with power-law drift

The stochastic differential equation (SDE) for the Modified Bessel Process with Power-Law Drift is:

$$d\beta(t) = \left(\theta(\beta_e - \beta(t)) + \frac{2k(\nu - 1)}{\beta(t)^r} \right) dt + \rho dB(t),$$

where:

- $\theta > 0$ is the mean reversion rate,
- β_e is the long-term mean,
- $r \geq 0$ determines the strength of the repelling force,
- $k > 0$ is a constant that scales the power-law repelling force,
- $\nu > 1$ is a parameter associated with the harmonic function used in the transformation,
- $\rho > 0$ is the volatility coefficient,
- $dB(t)$ represents a standard Brownian motion.

Fokker–Planck equation

The Fokker–Planck equation corresponding to this SDE is:

$$\frac{\partial p(\beta, t)}{\partial t} = -\frac{\partial}{\partial \beta} \left[\left(\theta(\beta_e - \beta) + \frac{2k(\nu - 1)}{\beta^r} \right) p(\beta, t) \right] + \frac{1}{2} \frac{\partial^2}{\partial \beta^2} (\rho^2 p(\beta, t)).$$

At steady state, $\frac{\partial p(\beta, t)}{\partial t} = 0$, this simplifies to:

$$0 = -\frac{\partial}{\partial \beta} \left[\left(\theta(\beta_e - \beta) + \frac{2k(\nu - 1)}{\beta^r} \right) p_{st}(\beta) \right] + \frac{1}{2} \frac{\partial^2}{\partial \beta^2} (\rho^2 p_{st}(\beta)).$$

Solving for the stationary distribution

Assume a solution of the form:

$$p_{st}(\beta) = C_1 \beta^a e^{-b\beta}.$$

Substituting this into the Fokker–Planck equation and matching coefficients for the drift and diffusion terms, we get the following conditions:

$$a - 1 = 2k(\nu - 1), \quad b = \frac{2\theta}{\rho^2}.$$

Thus, the parameter a is influenced by the additional drift term due to the repelling force, and b remains the same as in the traditional mean-reverting process, arising from the mean-reverting component.

Stationary distribution

The stationary distribution $p_{st}(\beta)$ is now determined by both the power-law drift and the mean-reverting component:

$$p_{st}(\beta) = C_1 \beta^{2k(\nu-1)} e^{-\frac{2\theta\beta}{\rho^2}}.$$

To normalize $p_{st}(\beta)$, we need to compute the following integral:

$$C_1 \int_0^\infty \beta^{2k(\nu-1)} e^{-\frac{2\theta\beta}{\rho^2}} d\beta.$$

This integral is of the Gamma distribution form:

$$C_1 \left(\frac{\rho^2}{2\theta} \right)^{2k(\nu-1)+1} \Gamma(2k(\nu - 1) + 1) = 1.$$

Solving for C_1 , we get:

$$C_1 = \frac{\left(\frac{2\theta}{\rho^2}\right)^{2k(\nu-1)+1}}{\Gamma(2k(\nu-1)+1)}.$$

Thus, the stationary distribution becomes:

$$p_{st}(\beta) = \frac{\left(\frac{2\theta}{\rho^2}\right)^{2k(\nu-1)+1}}{\Gamma(2k(\nu-1)+1)} \beta^{2k(\nu-1)} e^{-\frac{2\theta\beta}{\rho^2}}.$$

Stochastic threshold for the mean-reverting Bessel process with power-law drift

Again, since the noise perturbation for the mean-reverting Bessel (MRBP) process, $\rho dB(t)$, is identical to that of the OU process, the stochastic threshold is also the same. From the variance analysis section, the OU process, $\beta(t)$, has constant variance $\sigma_{\text{MRBP}} = \frac{\rho^2}{2\theta}$. The stochastic threshold is:

$$\tilde{\mathcal{T}}_0^{\text{MRBP}} = \mathcal{R}_0 - \frac{\rho^2}{4\theta(\mu + \lambda)} \left(\frac{\Lambda}{\mu} f'(0) \right)^2,$$

where $\mathcal{R}_0 = \frac{\beta_e \Lambda f'(0)}{\mu(\mu + \lambda)}$.

Derivation via Doob's h -transform

We aim to derive the stochastic differential equation (SDE) for the Modified Bessel Process with Power-Law Drift by applying Doob's h -transform to the OU process. The goal is to introduce a power-law repelling force term $\frac{2k(\nu-1)}{\beta(t)^r}$, while ensuring that the process remains positive and mean-reverting.

The OU process is defined by the SDE:

$$d\beta(t) = \theta(\beta_e - \beta(t)) dt + \rho dB(t),$$

where:

- $\theta > 0$ is the mean reversion rate,
- β_e is the long-term mean,
- $\rho > 0$ is the volatility coefficient,
- $dB(t)$ represents a standard Brownian motion.

The generator $\mathcal{L}$ of the OU process is:

$$\mathcal{L}f(x) = \theta(\beta_e - x)f'(x) + \frac{\rho^2}{2}f''(x),$$

where $f(x)$ is a smooth function.

We apply Doob's h -transform using a harmonic function $h(x)$ that introduces the desired power-law repelling force and ensures positivity. The harmonic function we choose is:

$$h(x) = x^{\nu-1}e^{-\frac{k}{x^r}},$$

where:

- $\nu > 1$ ensures the process remains positive,
- $r \geq 0$ controls the strength of the repelling force,
- $k > 0$ is a constant that scales the power-law repelling force.

This choice of $h(x)$ allows us to create a process that is both mean-reverting and exhibits a strong repelling force near zero (due to the term $e^{-\frac{k}{x^r}}$), ensuring that the process stays positive.

Using Doob's h -transform, the new generator $\mathcal{L}^h$ is given by:

$$\mathcal{L}^h f(x) = \frac{1}{h(x)}\mathcal{L}(h(x)f(x)) - \frac{f(x)}{h(x)}\mathcal{L}h(x).$$

We first compute $\mathcal{L}h(x)$ for $h(x) = x^{\nu-1}e^{-\frac{k}{x^r}}$. Applying the generator $\mathcal{L}$ to $h(x)$, we get two terms: one from $x^{\nu-1}$, which is responsible for the positivity constraint, and another from $e^{-\frac{k}{x^r}}$, which introduces the power-law repelling force. We compute:

$$\mathcal{L}(h(x)) = \theta(\beta_e - x) \left((\nu - 1)x^{\nu-2}e^{-\frac{k}{x^r}} - \frac{kr}{x^{r+1}}x^{\nu-1}e^{-\frac{k}{x^r}} \right) + \frac{\rho^2}{2}(\text{second derivatives}).$$

The term $\frac{kr}{x^{r+1}}$ in the generator is what gives rise to the power-law drift force that prevents the process from reaching zero.

Next, we compute $\mathcal{L}(h(x)f(x))$ using the product rule:

$$\mathcal{L}(h(x)f(x)) = \theta(\beta_e - x) \left((\nu - 1)x^{\nu-2}e^{-\frac{k}{x^r}}f(x) + x^{\nu-1}e^{-\frac{k}{x^r}}f'(x) \right) + \frac{\rho^2}{2}(\text{second derivatives of } f(x)).$$

From the transformed generator $\mathcal{L}^h f(x)$, we derive the corresponding stochastic differential equation (SDE) for the process $\beta(t)$.

The key term in this derivation is $\frac{kr}{x^{r+1}}$, which appears due to the $e^{-\frac{k}{x^r}}$ component in $h(x)$. This term generates a power-law repelling force in the drift, which acts to repel the process from zero. When x (or $\beta(t)$) becomes small, this force becomes very large, preventing the process from hitting zero.

Thus, the term $\frac{2k(\nu-1)}{\beta(t)^r}$ emerges in the drift, introducing a strong repelling force near zero. This force, combined with the original mean-reverting term $\theta(\beta_e - \beta(t))$, gives us the desired dynamics for a positive, mean-reverting process.

The resulting SDE for the Modified Bessel Process with Power-Law Drift is:

$$d\beta(t) = \left(\theta(\beta_e - \beta(t)) + \frac{2k(\nu - 1)}{\beta(t)^r} \right) dt + \rho dB(t),$$

where:

- $\theta > 0$ is the mean reversion rate,
- β_e is the long-term mean,
- $r \geq 0$ determines the strength of the repelling force,
- $k > 0$ scales the power-law repelling force,
- $\nu > 1$ ensures positivity,
- $\rho > 0$ is the volatility coefficient,
- $dB(t)$ represents a standard Brownian motion.

To ensure that the process $\beta(t)$ has the desired long-run mean μ_{target} , we adjust μ based on the drift term. The drift term is:

$$\theta(\mu - Y(t)) + \frac{2k(\nu - 1)}{Y(t)^r}.$$

At steady state, we set the expected drift to zero:

$$E \left[\theta(\mu - Y(t)) + \frac{2k(\nu - 1)}{Y(t)^r} \right] = 0.$$

This simplifies to:

$$\theta(\mu - \mu_{\text{target}}) + \frac{2k(\nu - 1)}{\mu_{\text{target}}^r} = 0.$$

Solving for μ , we obtain:

$$\mu = \mu_{\text{target}} - \frac{2k(\nu - 1)}{\theta \mu_{\text{target}}^r}.$$

By applying Doob's h -transform with the harmonic function $h(x) = x^{\nu-1}e^{-\frac{k}{x^r}}$, we derived the Modified Bessel Process with Power-Law Drift. This process exhibits both mean-reverting behaviour and a power-law repelling force, ensuring positivity and introducing a flexible repelling force near zero.

The final stochastic differential equation (SDE) is:

$$d\beta(t) = \left(\theta(\beta_e - \beta(t)) + \frac{2k(\nu - 1)}{\beta(t)^r} \right) dt + \rho dB(t),$$

with the long-run mean μ_{target} adjusted as needed.

5.6.3 Logistic GBM process

In this section, we derive the stochastic threshold for the Logistic Geometric Brownian Motion (Logistic GBM) process [99], [100], similar to the procedure of the previously analyzed processes.

The stochastic differential equation (SDE) for the Logistic Geometric Brownian Motion is:

$$d\beta(t) = \theta(\beta_e - \beta(t)) dt + \rho \beta(t)(1 - \beta(t)) dB(t),$$

where:

- $\theta > 0$ is the mean reversion rate,
- β_e is the long-term mean,
- $\rho > 0$ is the volatility coefficient,
- $dB(t)$ represents a standard Brownian motion.

Fokker–Planck Equation

The Fokker–Planck equation corresponding to this SDE describes the time evolution of the probability density function $p(\beta, t)$ of the Logistic GBM:

$$\frac{\partial p(\beta, t)}{\partial t} = -\frac{\partial}{\partial \beta} \left[\theta \left(\frac{\beta_e - \beta}{1 - \beta} \right) \beta(1 - \beta)p(\beta, t) \right] + \frac{1}{2} \frac{\partial^2}{\partial \beta^2} [\rho^2 \beta^2 (1 - \beta)^2 p(\beta, t)].$$

At steady state, $\frac{\partial p(\beta, t)}{\partial t} = 0$:

$$0 = -\frac{\partial}{\partial \beta} \left[\theta \left(\frac{\beta_e - \beta}{1 - \beta} \right) \beta(1 - \beta) p_{st}(\beta) \right] + \frac{1}{2} \frac{\partial^2}{\partial \beta^2} [\rho^2 \beta^2 (1 - \beta)^2 p_{st}(\beta)] .$$

Solving for the stationary distribution

To find the stationary distribution $p_{st}(\beta)$, we solve:

$$\frac{\partial}{\partial \beta} \left[\theta \left(\frac{\beta_e - \beta}{1 - \beta} \right) \beta(1 - \beta) p_{st}(\beta) \right] = \frac{\rho^2}{2} \frac{\partial^2}{\partial \beta^2} [\beta^2 (1 - \beta)^2 p_{st}(\beta)] .$$

Assuming a solution of the form $p_{st}(\beta) = C\beta^a(1 - \beta)^b$, we substitute into the differential equation and match coefficients.

Solving the differential equation gives us:

$$a = \frac{2\theta\beta_e}{\rho^2} - 1, \quad b = \frac{2\theta(1 - \beta_e)}{\rho^2} - 1.$$

Stationary distribution

The stationary distribution is:

$$p_{st}(\beta) = C\beta^{\frac{2\theta\beta_e}{\rho^2}-1}(1 - \beta)^{\frac{2\theta(1-\beta_e)}{\rho^2}-1}.$$

To normalize $p_{st}(\beta)$:

$$C \int_0^1 \beta^{\frac{2\theta\beta_e}{\rho^2}-1} (1 - \beta)^{\frac{2\theta(1-\beta_e)}{\rho^2}-1} d\beta = 1.$$

This integral is a Beta distribution form:

$$C \text{Beta} \left(\frac{2\theta\beta_e}{\rho^2}, \frac{2\theta(1 - \beta_e)}{\rho^2} \right) = 1,$$

giving:

$$C = \frac{1}{\text{Beta} \left(\frac{2\theta\beta_e}{\rho^2}, \frac{2\theta(1-\beta_e)}{\rho^2} \right)}.$$

So:

$$p_{st}(\beta) = \frac{\beta^{\frac{2\theta\beta_e}{\rho^2}-1} (1 - \beta)^{\frac{2\theta(1-\beta_e)}{\rho^2}-1}}{\text{Beta} \left(\frac{2\theta\beta_e}{\rho^2}, \frac{2\theta(1-\beta_e)}{\rho^2} \right)}.$$

Expectation calculation

To compute the expectation $\mathbb{E}[\beta(t)]$, we use the stationary distribution of the Logistic GBM process, which follows a Beta distribution-like form due to the bounded nature of $\beta(t)$ between 0 and 1:

$$p_{st}(\beta) = \frac{\beta^{\frac{2\theta\beta_e}{\rho^2}-1}(1-\beta)^{\frac{2\theta(1-\beta_e)}{\rho^2}-1}}{\text{Beta}\left(\frac{2\theta\beta_e}{\rho^2}, \frac{2\theta(1-\beta_e)}{\rho^2}\right)},$$

where $\text{Beta}(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ is the Beta function.

For $\mathbb{E}[\beta(t)]$, the expectation is given by:

$$\mathbb{E}[\beta(t)] = \int_0^1 \beta p_{st}(\beta) d\beta.$$

Using the properties of the Beta distribution:

$$\mathbb{E}[\beta(t)] = \frac{\frac{2\theta\beta_e}{\rho^2}}{\frac{2\theta\beta_e}{\rho^2} + \frac{2\theta(1-\beta_e)}{\rho^2}}.$$

Simplifying the expression:

$$\mathbb{E}[\beta(t)] = \frac{\frac{2\theta\beta_e}{\rho^2}}{\frac{2\theta}{\rho^2}} = \beta_e.$$

Hence, the expectation $\mathbb{E}[\beta(t)]$ confirms that β_e is the mean of the stationary distribution, reflecting the long-term average behaviour of $\beta(t)$.

Next, we calculate $\mathbb{E}[(\beta(t)(1-\beta(t)))^2]$ using the stationary distribution:

$$\mathbb{E}[(\beta(t)(1-\beta(t)))^2] = \int_0^1 \beta^2(1-\beta)^2 p_{st}(\beta) d\beta.$$

Substitute the expression for $p_{st}(\beta)$:

$$\mathbb{E}[(\beta(t)(1-\beta(t)))^2] = \frac{1}{\text{Beta}\left(\frac{2\theta\beta_e}{\rho^2}, \frac{2\theta(1-\beta_e)}{\rho^2}\right)} \int_0^1 \beta^{\frac{2\theta\beta_e}{\rho^2}+1}(1-\beta)^{\frac{2\theta(1-\beta_e)}{\rho^2}+1} d\beta.$$

Recognizing this integral as a Beta function form, we have:

$$\int_0^1 \beta^{a-1}(1-\beta)^{b-1} d\beta = \text{Beta}(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

with $a = \frac{2\theta\beta_e}{\rho^2} + 2$ and $b = \frac{2\theta(1-\beta_e)}{\rho^2} + 2$.

Thus, the expectation simplifies to:

$$\mathbb{E}[(\beta(t)(1 - \beta(t)))^2] = \frac{\text{Beta}\left(\frac{2\theta\beta_e}{\rho^2} + 2, \frac{2\theta(1-\beta_e)}{\rho^2} + 2\right)}{\text{Beta}\left(\frac{2\theta\beta_e}{\rho^2}, \frac{2\theta(1-\beta_e)}{\rho^2}\right)}.$$

Substitute the Beta function definition:

$$\mathbb{E}[(\beta(t)(1 - \beta(t)))^2] = \frac{\frac{\Gamma\left(\frac{2\theta\beta_e}{\rho^2} + 2\right)\Gamma\left(\frac{2\theta(1-\beta_e)}{\rho^2} + 2\right)}{\Gamma\left(\frac{2\theta}{\rho^2} + 4\right)}}{\frac{\Gamma\left(\frac{2\theta\beta_e}{\rho^2}\right)\Gamma\left(\frac{2\theta(1-\beta_e)}{\rho^2}\right)}{\Gamma\left(\frac{2\theta}{\rho^2}\right)}}.$$

This simplifies further to:

$$\mathbb{E}[(\beta(t)(1 - \beta(t)))^2] = \frac{\Gamma\left(\frac{2\theta\beta_e}{\rho^2} + 2\right)\Gamma\left(\frac{2\theta(1-\beta_e)}{\rho^2} + 2\right)\Gamma\left(\frac{2\theta}{\rho^2}\right)}{\Gamma\left(\frac{2\theta\beta_e}{\rho^2}\right)\Gamma\left(\frac{2\theta(1-\beta_e)}{\rho^2}\right)\Gamma\left(\frac{2\theta}{\rho^2} + 4\right)}.$$

Stochastic threshold for Logistic GBM

Substitute $\mathbb{E}[(\beta(t)(1 - \beta(t)))^2]$ into the stochastic threshold formula:

$$\tilde{\mathcal{T}}_0^{\text{LGBM}} = \mathcal{R}_0 - \frac{\rho^2}{2(\mu + \lambda)} \left(\frac{\Lambda}{\mu} f'(0)\right)^2 \mathbb{E}[(\beta(t)(1 - \beta(t)))^2],$$

which yields:

$$\tilde{\mathcal{T}}_0^{\text{LGBM}} = \mathcal{R}_0 - \frac{\rho^2}{2(\mu + \lambda)} \left(\frac{\Lambda}{\mu} f'(0)\right)^2 \frac{\Gamma\left(\frac{2\theta\beta_e}{\rho^2} + 2\right)\Gamma\left(\frac{2\theta(1-\beta_e)}{\rho^2} + 2\right)\Gamma\left(\frac{2\theta}{\rho^2}\right)}{\Gamma\left(\frac{2\theta\beta_e}{\rho^2}\right)\Gamma\left(\frac{2\theta(1-\beta_e)}{\rho^2}\right)\Gamma\left(\frac{2\theta}{\rho^2} + 4\right)}.$$

5.6.4 Jacobi process

In this section, we derive the stochastic threshold for the Jacobi process [101], similar to the procedure of the previously analyzed processes.

The stochastic differential equation for the Jacobi process is:

$$d\beta(t) = \theta(\beta_e - \beta(t)) dt + \rho\sqrt{\beta(t)(1 - \beta(t))} dB(t),$$

where:

- $\theta > 0$ is the mean reversion rate,

- β_e is the long-term mean,
- $\rho > 0$ is the volatility coefficient,
- $dB(t)$ represents a standard Brownian motion.

Fokker–Planck equation

The Fokker–Planck equation corresponding to this SDE describes the time evolution of the probability density function $p(\beta, t)$ of the Jacobi process:

$$\frac{\partial p(\beta, t)}{\partial t} = -\frac{\partial}{\partial \beta} [\theta(\beta_e - \beta)p(\beta, t)] + \frac{1}{2} \frac{\partial^2}{\partial \beta^2} [\rho^2 \beta(1 - \beta)p(\beta, t)].$$

At steady state, $\frac{\partial p(\beta, t)}{\partial t} = 0$:

$$0 = -\frac{\partial}{\partial \beta} [\theta(\beta_e - \beta)p_{st}(\beta)] + \frac{1}{2} \frac{\partial^2}{\partial \beta^2} [\rho^2 \beta(1 - \beta)p_{st}(\beta)].$$

Solving for the stationary distribution

To find the stationary distribution $p_{st}(\beta)$, we solve:

$$\frac{\partial}{\partial \beta} [\theta(\beta_e - \beta)p_{st}(\beta)] = \frac{\rho^2}{2} \frac{\partial^2}{\partial \beta^2} [\beta(1 - \beta)p_{st}(\beta)].$$

Assuming a solution of the form $p_{st}(\beta) = C\beta^a(1 - \beta)^b$, we substitute into the differential equation and match coefficients.

Solving gives us:

$$a = \frac{2\theta\beta_e}{\rho^2} - 1, \quad b = \frac{2\theta(1 - \beta_e)}{\rho^2} - 1.$$

Stationary distribution

The stationary distribution is:

$$p_{st}(\beta) = C\beta^{\frac{2\theta\beta_e}{\rho^2}-1}(1 - \beta)^{\frac{2\theta(1-\beta_e)}{\rho^2}-1}.$$

To normalize $p_{st}(\beta)$, we need to satisfy the condition:

$$C \int_0^1 \beta^{\frac{2\theta\beta_e}{\rho^2}-1} (1 - \beta)^{\frac{2\theta(1-\beta_e)}{\rho^2}-1} d\beta = 1.$$

This integral is a Beta function:

$$C \text{Beta} \left(\frac{2\theta\beta_e}{\rho^2}, \frac{2\theta(1-\beta_e)}{\rho^2} \right) = 1,$$

giving us the normalization constant:

$$C = \frac{1}{\text{Beta} \left(\frac{2\theta\beta_e}{\rho^2}, \frac{2\theta(1-\beta_e)}{\rho^2} \right)}.$$

Thus, the stationary distribution becomes:

$$p_{st}(\beta) = \frac{\beta^{\frac{2\theta\beta_e}{\rho^2}-1} (1-\beta)^{\frac{2\theta(1-\beta_e)}{\rho^2}-1}}{\text{Beta} \left(\frac{2\theta\beta_e}{\rho^2}, \frac{2\theta(1-\beta_e)}{\rho^2} \right)}.$$

Expectation calculation

Now, we calculate $\mathbb{E}[\beta(t)(1-\beta(t))]$. This expectation is important for analyzing the average behaviour of the diffusion term in the Jacobi process.

The expectation is given by:

$$\mathbb{E}[\beta(t)(1-\beta(t))] = \int_0^1 \beta(1-\beta)p_{st}(\beta) d\beta.$$

Substitute the stationary distribution $p_{st}(\beta)$:

$$\mathbb{E}[\beta(t)(1-\beta(t))] = \frac{1}{\text{Beta} \left(\frac{2\theta\beta_e}{\rho^2}, \frac{2\theta(1-\beta_e)}{\rho^2} \right)} \int_0^1 \beta^{\frac{2\theta\beta_e}{\rho^2}} (1-\beta)^{\frac{2\theta(1-\beta_e)}{\rho^2}} d\beta.$$

This is recognized as another Beta function:

$$\mathbb{E}[\beta(t)(1-\beta(t))] = \frac{\text{Beta} \left(\frac{2\theta\beta_e}{\rho^2} + 1, \frac{2\theta(1-\beta_e)}{\rho^2} + 1 \right)}{\text{Beta} \left(\frac{2\theta\beta_e}{\rho^2}, \frac{2\theta(1-\beta_e)}{\rho^2} \right)}.$$

Using the fact that the Beta function can be expressed in terms of the Gamma function:

$$\text{Beta}(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

we rewrite the expectation as:

$$\mathbb{E}[\beta(t)(1 - \beta(t))] = \frac{\frac{\Gamma\left(\frac{2\theta\beta_e}{\rho^2} + 1\right)\Gamma\left(\frac{2\theta(1-\beta_e)}{\rho^2} + 1\right)}{\Gamma\left(\frac{2\theta}{\rho^2} + 2\right)}}{\frac{\Gamma\left(\frac{2\theta\beta_e}{\rho^2}\right)\Gamma\left(\frac{2\theta(1-\beta_e)}{\rho^2}\right)}{\Gamma\left(\frac{2\theta}{\rho^2}\right)}}.$$

This simplifies to:

$$\mathbb{E}[\beta(t)(1 - \beta(t))] = \frac{\Gamma\left(\frac{2\theta\beta_e}{\rho^2} + 1\right)\Gamma\left(\frac{2\theta(1-\beta_e)}{\rho^2} + 1\right)\Gamma\left(\frac{2\theta}{\rho^2}\right)}{\Gamma\left(\frac{2\theta\beta_e}{\rho^2}\right)\Gamma\left(\frac{2\theta(1-\beta_e)}{\rho^2}\right)\Gamma\left(\frac{2\theta}{\rho^2} + 2\right)}.$$

Stochastic threshold for the Jacobi process

The stochastic threshold $\tilde{\mathcal{T}}_0^{\text{Jacobi}}$ is calculated by substituting $\mathbb{E}[\beta(t)(1 - \beta(t))]$ into the threshold formula:

$$\tilde{\mathcal{T}}_0^{\text{Jacobi}} = \mathcal{R}_0 - \frac{\rho^2}{2(\mu + \lambda)} \left(\frac{\Lambda}{\mu} f'(0) \right)^2 \mathbb{E}[\beta(t)(1 - \beta(t))].$$

Substituting the expectation:

$$\tilde{\mathcal{T}}_0^{\text{Jacobi}} = \mathcal{R}_0 - \frac{\rho^2}{2(\mu + \lambda)} \left(\frac{\Lambda}{\mu} f'(0) \right)^2 \cdot \frac{\Gamma\left(\frac{2\theta\beta_e}{\rho^2} + 1\right)\Gamma\left(\frac{2\theta(1-\beta_e)}{\rho^2} + 1\right)\Gamma\left(\frac{2\theta}{\rho^2}\right)}{\Gamma\left(\frac{2\theta\beta_e}{\rho^2}\right)\Gamma\left(\frac{2\theta(1-\beta_e)}{\rho^2}\right)\Gamma\left(\frac{2\theta}{\rho^2} + 2\right)}.$$

Thus, the stochastic threshold for the Jacobi process is:

$$\tilde{\mathcal{T}}_0^{\text{Jacobi}} = \mathcal{R}_0 - \frac{\rho^2}{2(\mu + \lambda)} \left(\frac{\Lambda}{\mu} f'(0) \right)^2 \cdot \frac{\Gamma\left(\frac{2\theta\beta_e}{\rho^2} + 1\right)\Gamma\left(\frac{2\theta(1-\beta_e)}{\rho^2} + 1\right)\Gamma\left(\frac{2\theta}{\rho^2}\right)}{\Gamma\left(\frac{2\theta\beta_e}{\rho^2}\right)\Gamma\left(\frac{2\theta(1-\beta_e)}{\rho^2}\right)\Gamma\left(\frac{2\theta}{\rho^2} + 2\right)}.$$

5.7 Numerical results

5.7.1 Parameter sets for numerical simulations

The following sections describe the parameter settings used for simulating and analyzing the dynamics of the SIRS (Susceptible-Infectious-Recovered-Susceptible) epidemic model. Two primary parameter sets, **Set A** and **Set B**, are used in the simulations, the latter based on different scenarios and theoretical conditions, as outlined by Laaribi (2023). These sets include key epidemiological parameters that govern infection transmission, recovery, and

removal in populations, along with stochastic components that influence the behaviour of the transmission rate over time [102].

Parameter set A: Baseline epidemic dynamics

Set A represents a baseline scenario designed to explore deterministic and simple epidemic dynamics without extensive stochastic influences. The following categorizations highlight the role of the various parameters within the different cases:

- β_e : Fixed at 0.22 in all scenarios, representing the baseline transmission rate.
- γ : The immunity loss rate, which varies across scenarios:
 - **Case 1:** $\gamma = 0.00$ – *No reversion*: No immunity loss, meaning once recovered, individuals stay immune.
 - **Case 2:** $\gamma = 0.01$ – *Oscillating solutions*: Low immunity loss allows for oscillations in infection numbers over time.
 - **Case 3:** $\gamma = 0.10$ – *Rapid waning immunity*: High immunity loss rate causes individuals to quickly return to the susceptible population.
 - **Case 4:** $\gamma = 0.03$ – *Slow waning immunity*: Moderate immunity loss rate leads to slower transitions back to susceptibility.
- λ : Disease-induced removal or death rate, fixed at 0.1 across all cases.
- Λ, μ : Birth rate and natural death rate, both set to a small value (1×10^{-4}) to maintain a stable population.
- S_0, I_0, R_0 : Initial population conditions for each case:
 - $S_0 = 0.99$ (99% susceptible),
 - $I_0 = 0.01$ (1% infected),
 - $R_0 = 0$ (no recovered individuals).
- T : Simulation time is set to 1000 days, providing a sufficient period to observe the dynamics.
- N : Number of time steps, also set to 1000, ensuring detailed observation of the epidemic progression.

Similar to Set B, **Set A** also incorporates mean reversion and volatility parameters to capture some level of stochastic behaviour:

- θ : The rate of mean reversion, which is fixed at 0.4 across all cases.
- ρ : Volatility coefficient, fixed at 0.25.

The merged parameter settings for **Set A** are as follows:

- **Case 1 (No reversion)**: $\beta_e = 0.22$, $\gamma = 0.00$, $\lambda = 0.1$, $\Lambda = 1 \times 10^{-4}$, $\mu = 1 \times 10^{-4}$, $S_0 = 0.99$, $I_0 = 0.01$, $R_0 = 0$, $T = 1000$, $N = 1000$, $\theta = 0.4$, $\rho = 0.25$
- **Case 2 (Oscillating solutions)**: $\beta_e = 0.22$, $\gamma = 0.01$, $\lambda = 0.1$, $\Lambda = 1 \times 10^{-4}$, $\mu = 1 \times 10^{-4}$, $S_0 = 0.99$, $I_0 = 0.01$, $R_0 = 0$, $T = 1000$, $N = 1000$, $\theta = 0.4$, $\rho = 0.25$
- **Case 3 (Rapid waning immunity)**: $\beta_e = 0.22$, $\gamma = 0.10$, $\lambda = 0.1$, $\Lambda = 1 \times 10^{-4}$, $\mu = 1 \times 10^{-4}$, $S_0 = 0.99$, $I_0 = 0.01$, $R_0 = 0$, $T = 1000$, $N = 1000$, $\theta = 0.4$, $\rho = 0.25$
- **Case 4 (Slow waning immunity)**: $\beta_e = 0.22$, $\gamma = 0.03$, $\lambda = 0.1$, $\Lambda = 1 \times 10^{-4}$, $\mu = 1 \times 10^{-4}$, $S_0 = 0.99$, $I_0 = 0.01$, $R_0 = 0$, $T = 1000$, $N = 1000$, $\theta = 0.4$, $\rho = 0.25$

Parameter set B: Laaribi, 2023 scenarios

Set B explores different scenarios with stochastic dynamics and varying transmission rates. Each configuration in Set B controls essential aspects of disease spread, including transmission, recovery, immunity loss, and disease-induced removal. The following are the key parameters for each scenario in **Set B**:

- β_e : The long-term mean transmission rate, which directly influences the basic reproduction number R_0 .
- Λ : The birth rate or rate of entry into the susceptible population, maintaining the population balance.
- μ : The natural death or recovery rate, accounting for individuals exiting the population or recovering naturally.
- γ : The rate of immunity loss, which describes how quickly recovered individuals return to the susceptible state.
- λ : The disease-induced death or removal rate, representing how quickly infected individuals are removed from the population due to the disease.

- S_0, I_0, R_0 : Initial conditions representing the proportions of the population that are susceptible (S_0), infected (I_0), and recovered (R_0).
- T : The total simulation time, over which the epidemic dynamics will be observed.
- N : The number of discrete time steps, also set to 1000, into which the time interval $[0, T]$ is divided. Thus, each time step is $\Delta t = T/N$. Increasing N yields a finer temporal resolution, resulting in a smaller Δt and a more detailed approximation of the underlying continuous-time process.

In addition to these core parameters, **Set B** also includes stochastic parameters to account for randomness in the transmission dynamics:

- θ : The rate of mean reversion, which governs how quickly the transmission rate ($\beta(t)$) returns to its long-term mean (β_e).
- ρ : The volatility coefficient, representing the randomness or variability in the transmission rate over time.

The merged parameters for **Set B**, incorporating mean reversion and volatility, are as follows:

- **Scenario 1:** $\beta_e = 0.2, \Lambda = 0.09, \mu = 0.09, \gamma = 0.01, \lambda = 0.1, S_0 = 0.5, I_0 = 0.15, R_0 = 0.35, T = 1000, N = 1000, \theta = 0.4, \rho = 0.25$
- **Scenario 2:** $\beta_e = 0.22, \Lambda = 0.09, \mu = 0.09, \gamma = 0.01, \lambda = 0.1, S_0 = 0.5, I_0 = 0.15, R_0 = 0.35, T = 1000, N = 1000, \theta = 0.6, \rho = 0.29$
- **Scenario 3:** $\beta_e = 0.35, \Lambda = 0.09, \mu = 0.09, \gamma = 0.01, \lambda = 0.1, S_0 = 0.5, I_0 = 0.15, R_0 = 0.35, T = 1000, N = 1000, \theta = 1.3, \rho = 0.4$
- **Scenario 4:** $\beta_e = 0.188, \Lambda = 0.09, \mu = 0.09, \gamma = 0.01, \lambda = 0.1, S_0 = 0.5, I_0 = 0.15, R_0 = 0.35, T = 1000, N = 1000, \theta = 0.4, \rho = 0.25$

These parameter configurations allow us to explore various stochastic processes in the epidemic, accounting for different transmission rates and how quickly the transmission returns to a baseline level after fluctuations.

Merging of parameter sets

For both **Set A** and **Set B**, the mean reversion and volatility parameters were integrated into the corresponding scenarios, ensuring that both deterministic and stochastic elements are properly captured in the simulations. This provides a comprehensive view of how both baseline and fluctuating dynamics impact the spread of an epidemic.

These parameter sets form the foundation for the upcoming numerical simulations and theoretical validation. The different combinations of transmission rates, recovery dynamics, and stochastic processes will provide insight into the behaviour of the epidemic under various conditions. The subsequent sections will demonstrate the numerical results, plots, and an in-depth analysis based on these configurations.

5.7.2 Deterministic and noisy synthetic data generation

The deterministic SIRS model was applied to both **Set A** and **Set B** parameter configurations to simulate the spread of infection over time. For each set, simulations were run both without noise (deterministic) and with noise (stochastic), providing a comprehensive view of how the epidemic dynamics evolve under idealized and realistic conditions.

Initially, the deterministic SIRS model was used to generate clean, theoretical infection trajectories. These simulations provided a baseline representation of the epidemic, focusing on key factors such as transmission rate, immunity loss, and recovery. The infection time series for each scenario in **Set A** and **Set B** was plotted to visualize how the infection evolves over time across different parameter settings.

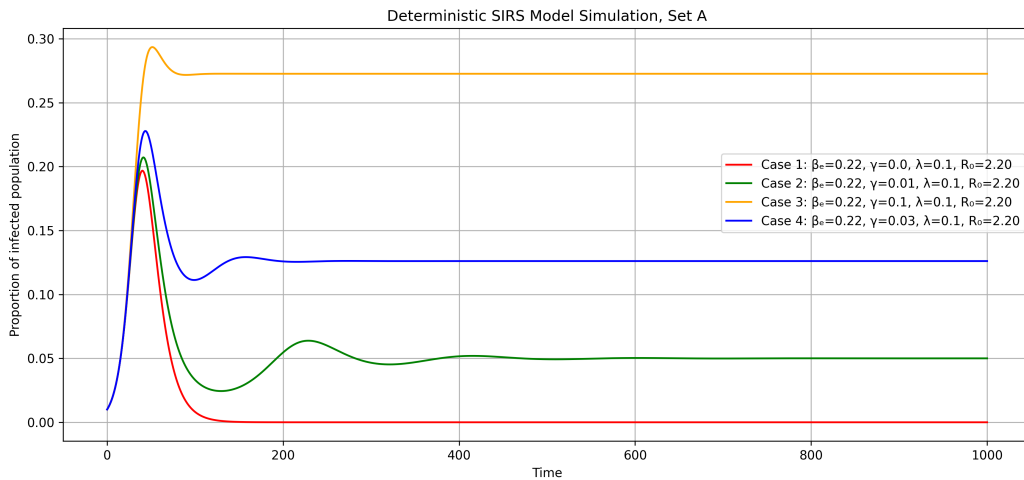


Figure 5.2: A plot of the deterministic SIRS model simulation for Set A.

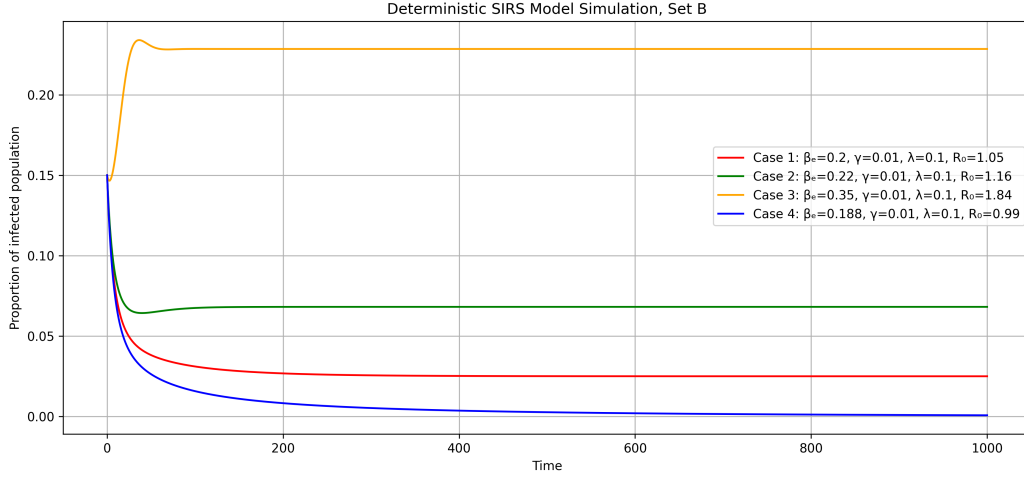


Figure 5.3: A plot of the deterministic SIRS model simulation for Set B.

Following the deterministic simulations, synthetic noise was introduced to emulate real-world uncertainties, such as measurement errors or random fluctuations in disease transmission. This noisy dataset offers a more realistic depiction of epidemic dynamics, reflecting the variability seen in real-world outbreaks. Adding noise is crucial for evaluating the robustness of the model and ensuring that it is applicable to real-world scenarios. The noisy infection trajectories were also visualized, offering a comparison between the idealized and noisy outcomes.

For both **Set A** and **Set B**, the noisy infection time series data was saved as CSV files. These datasets serve as a foundation for further analysis, including model calibration, sensitivity testing, and validation. The combination of deterministic and noisy simulations provides a robust framework for understanding how various factors, both intrinsic and extrinsic, affect the spread of infectious diseases.

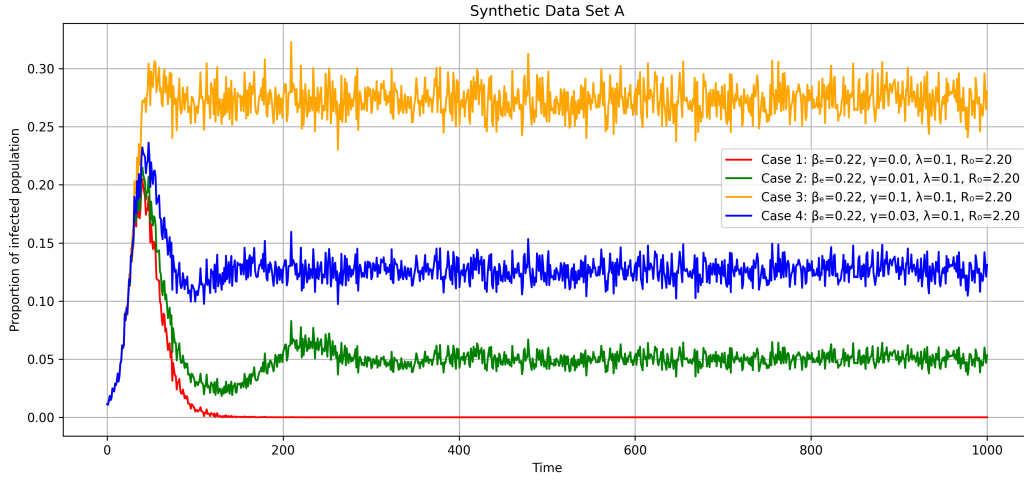


Figure 5.4: A plot of the synthetic data for parameter Set A.

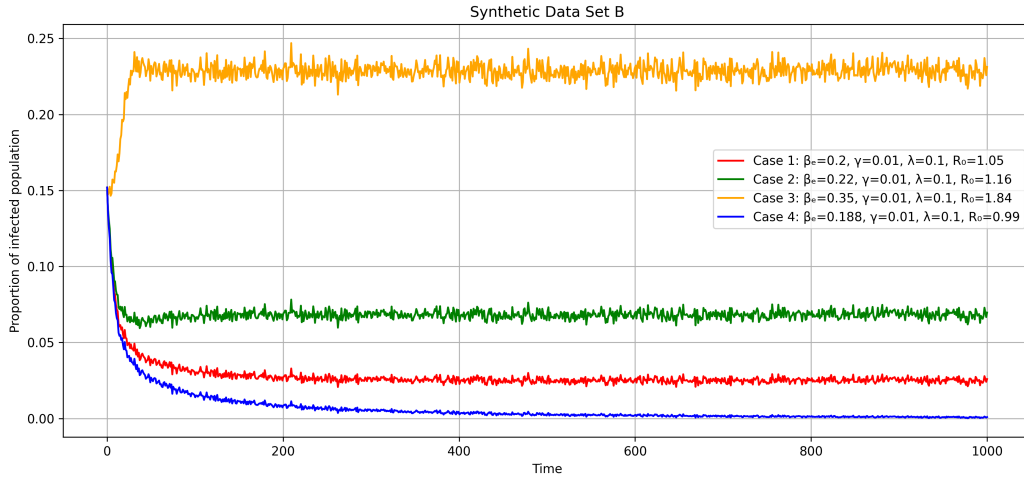


Figure 5.5: A plot of the synthetic data for parameter Set B.

5.7.3 Simulation and plotting of model fit

This section describes the process of simulating and plotting the fit of various stochastic processes against the observed infection data. The goal is to evaluate how well different models, such as the OU process, fit the observed infection time series using estimated parameters. Below is an explanation of the key steps involved in the process.

The primary function used in this section is designed to simulate the infection dynamics based on a chosen stochastic process and compare the results to the original infection data.

For each process, a specific set of parameters is required. The number of parameters varies depending on the complexity of the process:

- For the OU process, there is a requirement of three parameters: β_e (long-term mean transmission rate), θ (rate of mean reversion), and ρ (volatility).
- The Bessel process requires an additional parameter, ν , to account for the added complexity.
- The power-law Bessel process requires six parameters, adding k (drift parameter), ν , and r (power-law exponent).

Generating the transmission rate (β_t) is accomplished as follows. The function generates a time series for the transmission rate, $\beta(t)$, based on the selected stochastic process (e.g., OU, CKLS, etc.). This rate drives the infection dynamics in the SIRS model. The stochastic component of $\beta(t)$ is generated using random numbers, ensuring reproducibility.

Simulating the infection dynamics is as follows. The SIRS model is run using the simulated β_t , initial conditions (susceptible S_0 , infected I_0 , recovered R_0), and parameters such as birth rate (Λ) and death rate (μ). The Milstein method is employed to numerically solve the SIRS model under stochastic variability [103], [104].

The observed infection data (I_{observed}) is compared to the simulated infection data (I). The plotting includes the original infection data and the simulated infection data, based on the chosen stochastic process.

Running the simulation for different processes

Several stochastic processes are evaluated, including:

- OU
- CKLS with $\gamma = \frac{1}{2}$
- Mean-reverting Bessel process
- Power-law mean-reverting Bessel process
- Logistic geometric Brownian motion
- Jacobi process

Each process has specific parameters that influence its behaviour. Initial parameter guesses are provided, which are later optimized to fit the observed infection data.

Fitting the model and estimating parameters

We implement a procedure that iterates over all cases in the dataset (Set A or Set B) and fits the model to estimate optimal parameters for each process.

The parameter estimation involves several key steps to ensure that the model-generated infection trajectory aligns as closely as possible with the observed data. First, the model simulates the SIRS dynamics over a prescribed time horizon T , divided into N discrete steps, and uses a chosen stochastic process for $\beta(t)$. To numerically solve the underlying stochastic differential equations, a Milstein discretization scheme [103], [104] is employed, providing an accurate and stable approximation of the continuous-time process.

A loss function is then defined as the sum of squared differences between the observed infection counts I_{observed} and the model's simulated infection trajectory I . This loss quantifies how effectively a given set of parameters reproduces the observed data. To maintain positivity and improve numerical stability, the parameters are estimated in logarithmic form. Inside the loss function, the parameters are exponentiated before the simulation runs, thus ensuring that all parameter values remain positive.

The Nelder–Mead simplex method is used to minimize the loss function. Starting from an initial guess provided in logarithmic space, the algorithm iteratively adjusts the parameters in an attempt to reduce the discrepancy between the observed and simulated trajectories. Each iteration updates the parameter values, and the process continues until a suitable convergence criterion is met or a specified iteration limit is reached.

Upon convergence, the optimized parameters are those that yield the best fit according to the chosen criterion. These parameters are then exponentiated and, if required, adjusted to satisfy any model-specific constraints such as $\nu > 1$. The estimated parameters represent the set of values that, within the given modelling framework, most faithfully reproduce the observed infection data.

For each process, the results are presented as plots comparing the observed and simulated infection data. The plots are placed after the explanation of each process and case.

After each plot, the estimated parameters for the corresponding process should be reported, including β_e , θ , ρ for the Ornstein–Uhlenbeck process and additional parameters ν for the Bessel process, and k , ν , r for the power-law Bessel process. These estimated parameters provide insight into how the transmission rate evolves over time under each model.

The simulations serve two primary purposes. The first pertains to model validation. By comparing simulated infection curves with observed synthetic data, the suitability of each stochastic process can be assessed. This helps identify the most appropriate model for describing infection dynamics. Secondly, optimization yielding the parameter estimates

provide valuable insights into the underlying dynamics of transmission rates and epidemic behaviour.

5.7.4 Illustrations of the determination of eradication/extinction or endemicity/persistence based on model type

Here, we show the plots of the original synthetic infected data along with the estimated SIRS model with OU and CKLS transmission processes for data set A. The remaining scenarios for the other processes considered and data set B are provided in the Appendix.

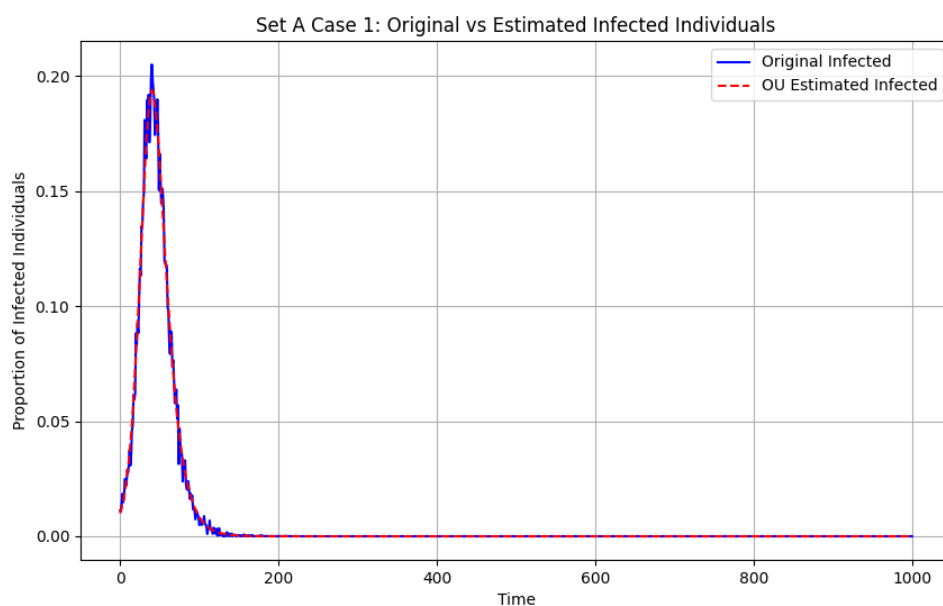


Figure 5.6: A plot of the original synthetic infected data along with the estimated SIRS model with OU transmission process. The data set and case number are A and 1, respectively.

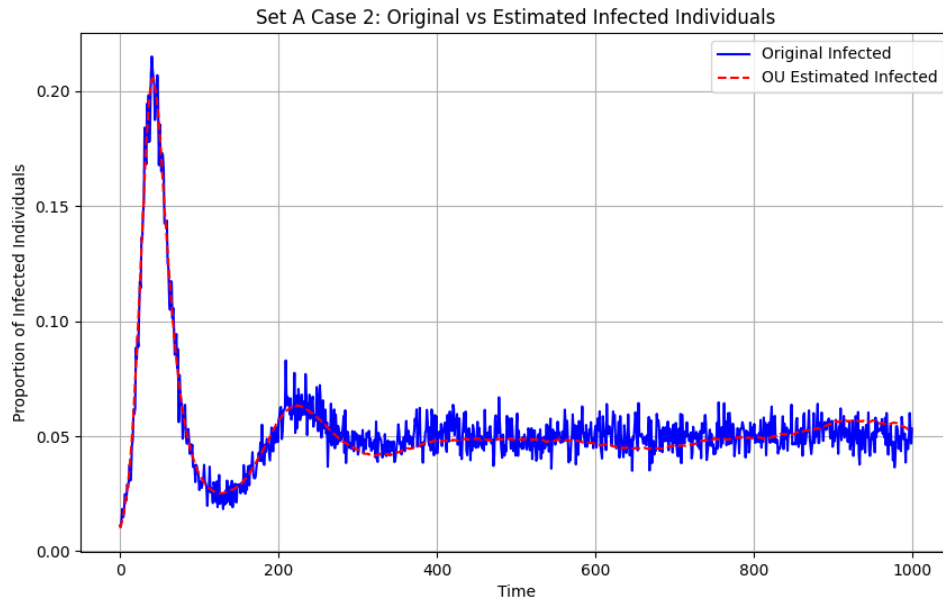


Figure 5.7: A plot of the original synthetic infected data along with the estimated SIRS model with OU transmission process. The data set and case number are A and 2, respectively.

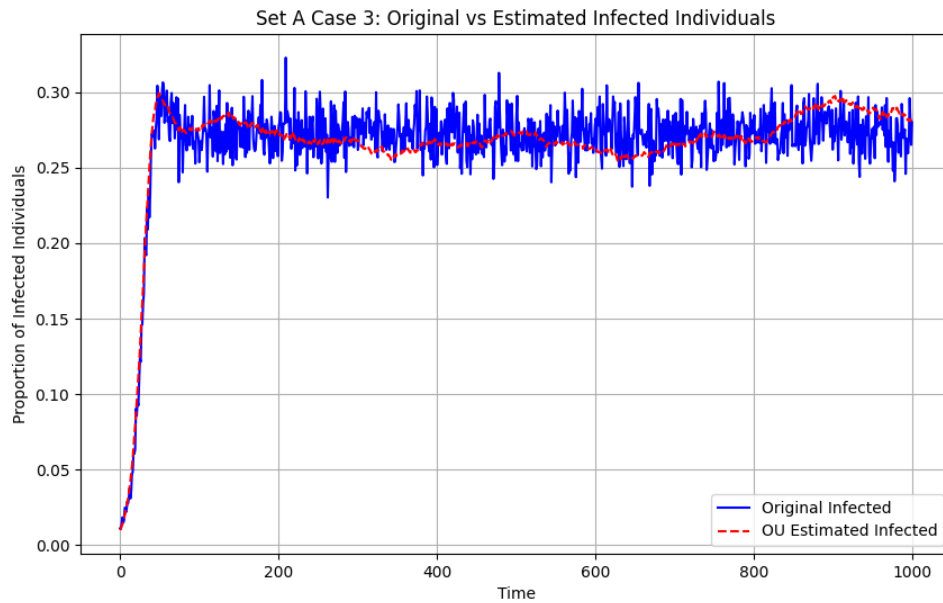


Figure 5.8: A plot of the original synthetic infected data along with the estimated SIRS model with OU transmission process. The data set and case number are A and 3, respectively.

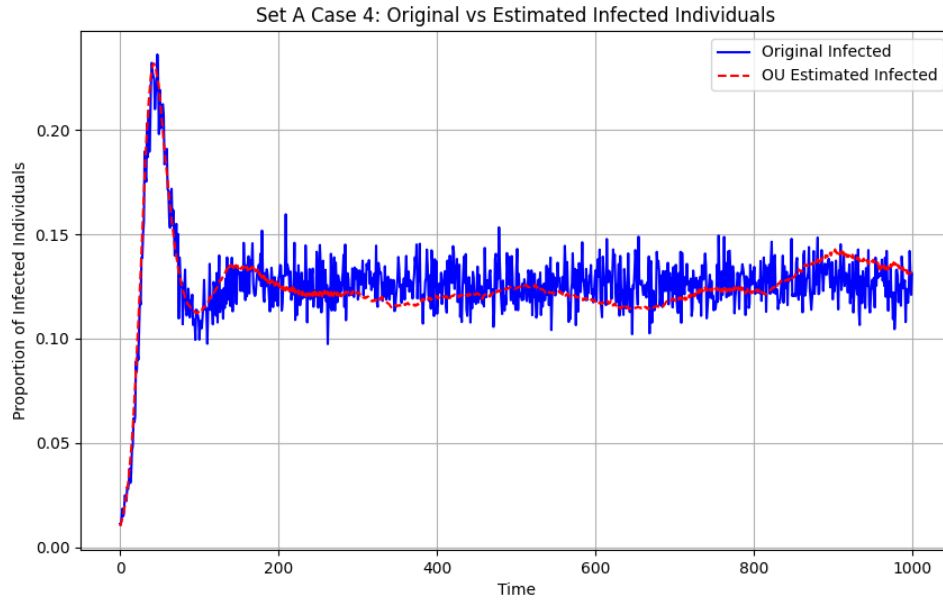


Figure 5.9: A plot of the original synthetic infected data along with the estimated SIRS model with OU transmission process. The data set and case number are A and 4, respectively.

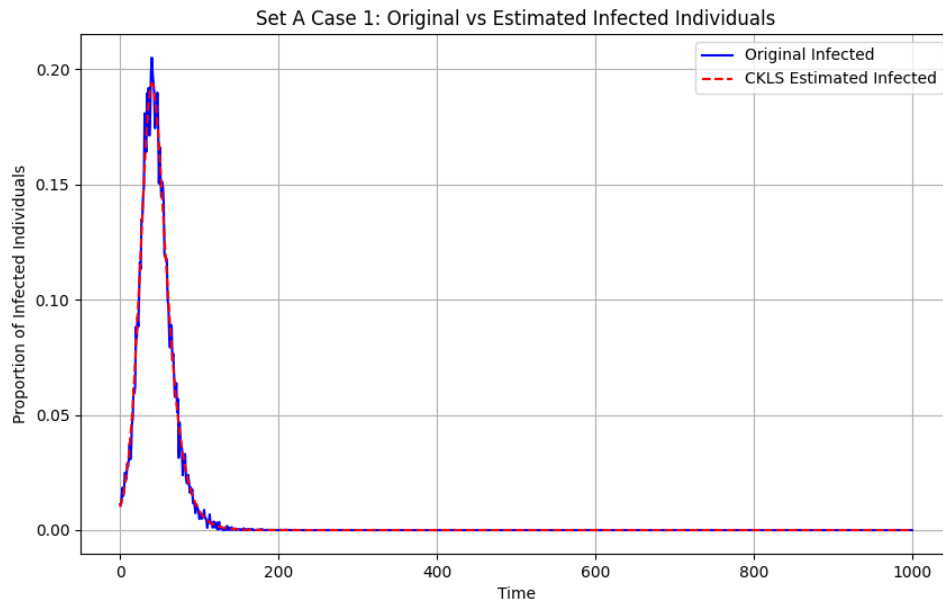


Figure 5.10: A plot of the original synthetic infected data along with the estimated SIRS model with CKLS transmission process. The data set and case number are A and 1, respectively.

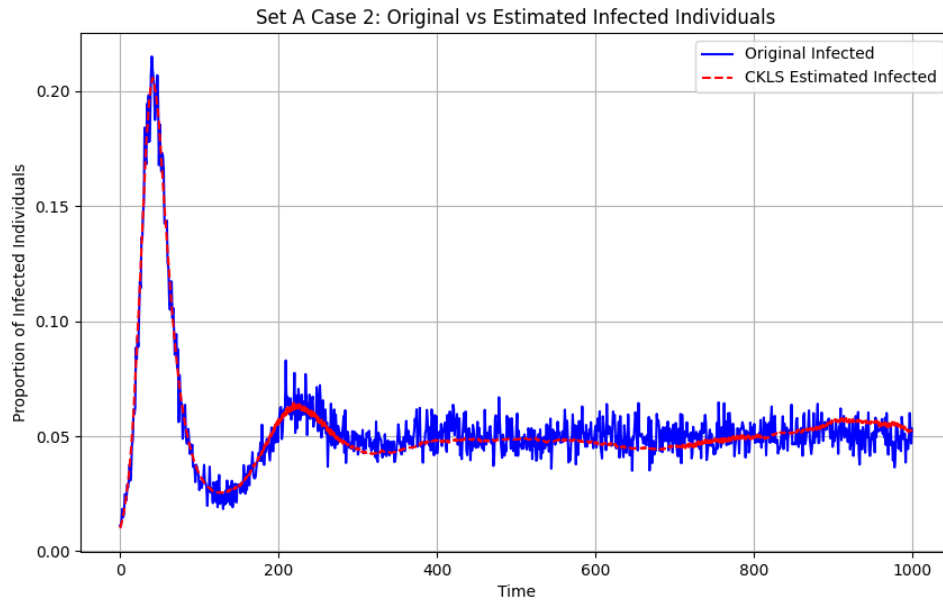


Figure 5.11: A plot of the original synthetic infected data along with the estimated SIRS model with CKLS transmission process. The data set and case number are A and 2, respectively.

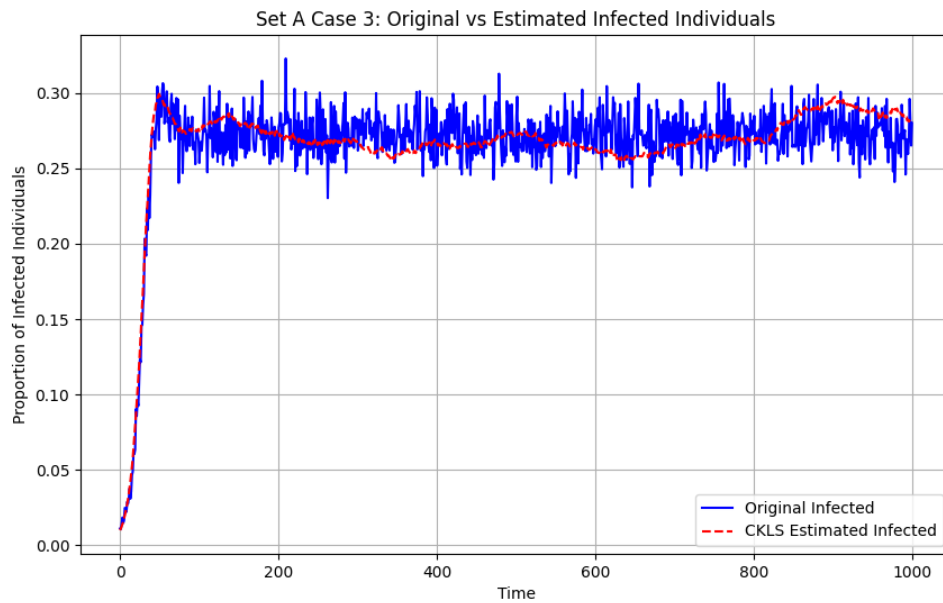


Figure 5.12: A plot of the original synthetic infected data along with the estimated SIRS model with CKLS transmission process. The data set and case number are A and 3, respectively.

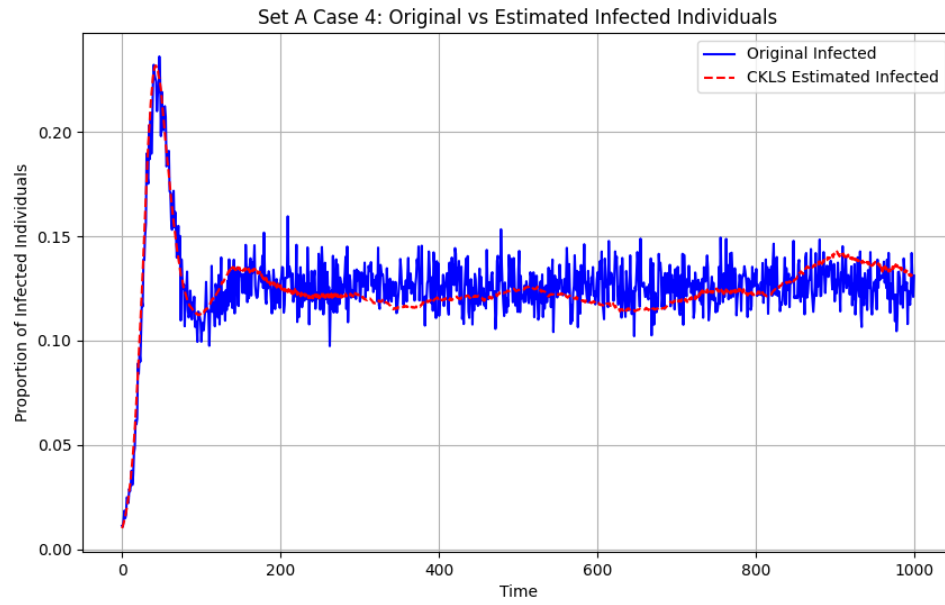


Figure 5.13: A plot of the original synthetic infected data along with the estimated SIRS model with CKLS transmission process. The data set and case number are A and 4, respectively.

Case	Process	β_e	θ	ρ
Set A				
1	OU	0.2176	2.0941	9.1319e-10
2	OU	0.2184	1.6826	1.7616e-14
3	OU	0.2223	1.9985	6.2463e-04
4	OU	0.2220	1.9987	1.4955e-03
1	CKLS	0.2176	2.0571	6.9417e-05
2	CKLS	0.2189	1.9990	1.1000e-02
3	CKLS	0.2223	1.9986	1.3000e-03
4	CKLS	0.2220	1.9987	3.1000e-03
1	Bessel	0.2176	5.4700	1.0455e-17
2	Bessel	0.2192	0.4433	2.7000e-03
3	Bessel	0.2223	5.3379	6.1954e-11
4	Bessel	0.2222	1.1437	7.2797e-04
1	PowerLawBessel	0.2176	0.4929	9.4660e-02
2	PowerLawBessel	0.2184	0.4913	9.4607e-02
3	PowerLawBessel	0.2223	0.4868	9.4429e-02
4	PowerLawBessel	0.2219	0.4857	9.4605e-02
1	LogisticGBM	0.2175	10.1890	1.5680e-06
2	LogisticGBM	0.2184	9.1976	7.4233e-06
3	LogisticGBM	0.2223	9.0585	1.2544e-06
4	LogisticGBM	0.2219	9.0540	9.7708e-06
1	Jacobi	0.2176	2.0569	1.5916e-04
2	Jacobi	0.2187	1.9999	7.9660e-03
3	Jacobi	0.2223	1.9995	9.7333e-05
4	Jacobi	0.2220	1.9987	3.5810e-03
Set B				
1	OU	0.2001	2.0025	5.2543e-04
2	OU	0.2199	1.0437	1.2386e-04
3	OU	0.3500	0.1171	7.6907e-17
4	OU	0.1881	1.8604	7.1042e-03
1	CKLS	0.1999	0.2787	1.5733e-17
2	CKLS	0.2199	1.9980	4.8877e-04
3	CKLS	0.3500	0.0000	2.9827e-17
4	CKLS	0.1881	1.8554	1.6180e-02
1	Bessel	0.1999	0.0252	6.4356e-16
2	Bessel	0.2204	1.9596	6.7961e-04
3	Bessel	0.3506	0.1803	1.3221e-03
4	Bessel	0.1881	0.5129	5.2611e-04
1	PowerLawBessel	0.1999	0.4869	9.3462e-02
2	PowerLawBessel	0.2199	0.4901	9.3690e-02
3	PowerLawBessel	0.3500	0.4907	8.5768e-02
4	PowerLawBessel	0.1880	0.4932	9.9474e-02
1	LogisticGBM	0.1999	10.0440	6.3741e-05
2	LogisticGBM	0.2199	8.8112	3.4295e-03
3	LogisticGBM	0.3500	0.0013	6.2869e-18
4	LogisticGBM	0.1881	5.4338	2.4085e-02
1	Jacobi	0.1999	1.7325	1.3014e-15
2	Jacobi	0.2199	1.9980	5.5363e-04
3	Jacobi	0.3500	1.9388	1.8267e-03
4	Jacobi	0.1881	1.8568	1.8023e-02

Table 5.1: Estimated parameters for different processes in Sets A and B across cases.

The parameter values listed in the above table differ from those presented in the pa-

parameter sets because they were derived from a separate set of simulation experiments and estimation procedures. The discrepancies arise for several reasons. First, it is important to recognize that the parameters displayed in the table above correspond to fitted or estimated values from particular numerical simulations conducted under specific modelling scenarios. These estimation routines, which often involve optimization algorithms tailored to the chosen stochastic processes, can produce parameter values that differ from the canonical values reported earlier in the text.

In contrast, the parameter settings in Set A and Set B reflect the baseline or nominal parameter settings initially prescribed for the simulations. Those earlier values serve as theoretical starting points or intended configurations for the epidemic scenarios. They were chosen to illustrate general epidemic behaviours and to provide a reference for how key epidemiological parameters, such as the baseline transmission rate β_e , the immunity loss rate γ , and the mean reversion rate θ , influence the resulting dynamics.

However, once one moves beyond these idealized starting conditions and applies more complex stochastic processes such as OU, CKLS, Bessel-type, Power-Law Bessel, Logistic GBM, or Jacobi frameworks, the effective parameter values that best fit the observed or simulated data may shift. This shift is a natural consequence of the model calibration process. During calibration, fitting algorithms adjust parameters to minimize discrepancies between model predictions and data, or to achieve certain statistical criteria. As a result, parameters like θ and ρ may end up being estimated at values quite different from the baseline inputs.

Moreover, each stochastic process imposes distinct constraints and interactions among parameters. For instance, ensuring positivity in CKLS or Bessel-type processes or achieving stable mean-reverting behaviour in OU or Logistic GBM models can lead to parameter estimates that do not closely resemble the originally set baseline values. The introduction of additional complexity and the requirement to satisfy certain model properties often necessitates such adjustments.

The difference in values are the outcome of model fitting or parameter estimation under more complex stochastic frameworks, while the earlier parameter definitions present nominal, theoretically chosen parameters used as initial reference points. The differences illustrate how model complexity, data fitting, and stochastic constraints influence parameter estimation, resulting in parameter sets that are well-suited to the chosen modelling scenario and process definitions.

Case	Process	β_e	θ	ρ	ν	k	r
Set A							
Case 1	Bessel	0.2176	5.4700	1.0455×10^{-17}	1.3649	-	-
Case 2	Bessel	0.2192	0.4433	2.7110×10^{-03}	2.4510	-	-
Case 3	Bessel	0.2223	5.3379	6.1954×10^{-11}	2.0325	-	-
Case 4	Bessel	0.2222	1.1437	7.2797×10^{-04}	2.0056	-	-
Case 1	PowerLawBessel	0.2176	0.4929	9.4660×10^{-02}	3.0047	1.0001	1.1979
Case 2	PowerLawBessel	0.2184	0.4913	9.4607×10^{-02}	3.0057	1.0001	1.1984
Case 3	PowerLawBessel	0.2223	0.4868	9.4429×10^{-02}	2.9982	1.0001	1.1981
Case 4	PowerLawBessel	0.2219	0.4857	9.4605×10^{-02}	2.9991	1.0001	1.1982
Set B							
Case 1	Bessel	0.1999	0.0252	6.4356×10^{-16}	2.1946	-	-
Case 2	Bessel	0.2204	1.9596	6.7961×10^{-04}	1.8479	-	-
Case 3	Bessel	0.3506	0.1803	1.3221×10^{-03}	2.1253	-	-
Case 4	Bessel	0.1881	0.5129	5.2611×10^{-04}	10.9570	-	-
Case 1	PowerLawBessel	0.1999	0.4869	9.3462×10^{-02}	2.9455	1.0000	1.1961
Case 2	PowerLawBessel	0.2199	0.4901	9.3690×10^{-02}	3.0099	1.0001	1.1985
Case 3	PowerLawBessel	0.3500	0.4907	8.5768×10^{-02}	3.0183	1.0004	1.2099
Case 4	PowerLawBessel	0.1880	0.4932	9.9474×10^{-02}	2.9977	0.9999	1.1931

Table 5.2: Estimated extended parameters for the Bessel and modified Bessel processes in Sets A and B.

5.7.5 Comparison of infected behaviour of deterministic and stochastic models

Here, we show the plots of the deterministic solution along with the simulated SIRS model with the OU and CKLS transmission processes. The remaining scenarios for the other processes considered and data set B are provided in the Appendix.

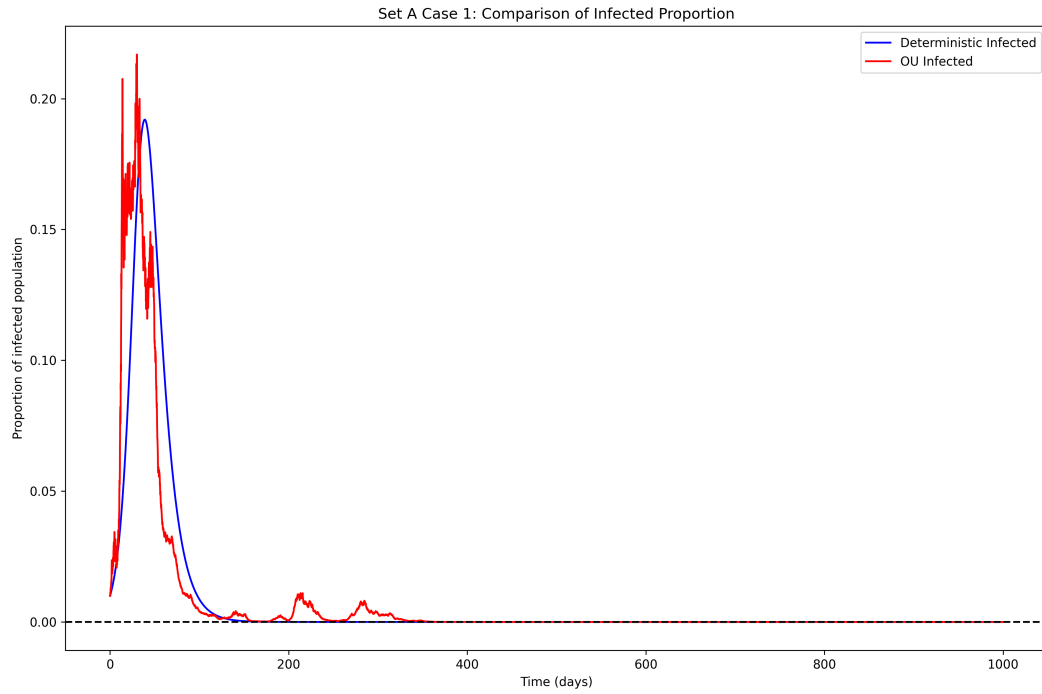


Figure 5.14: A plot of the deterministic solution along with the simulated SIRS model with the OU transmission process. The data set is A and the case number is 1.

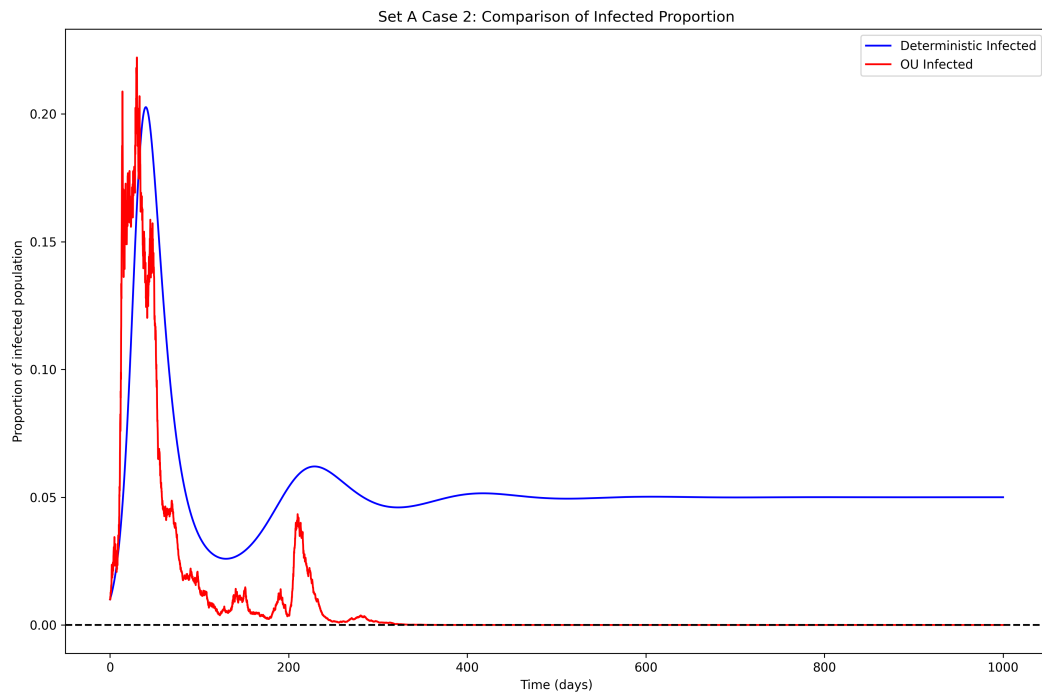


Figure 5.15: A plot of the deterministic solution along with the simulated SIRS model with the OU transmission process. The data set is A and the case number is 2.

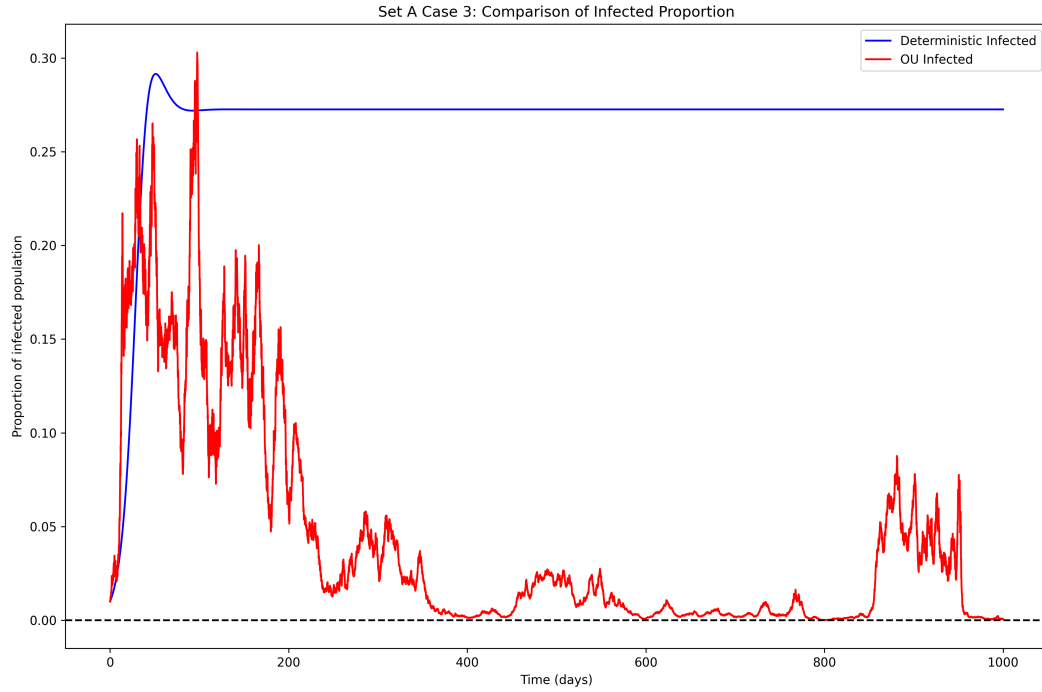


Figure 5.16: A plot of the deterministic solution along with the simulated SIRS model with the OU transmission process. The data set is A and the case number is 3.

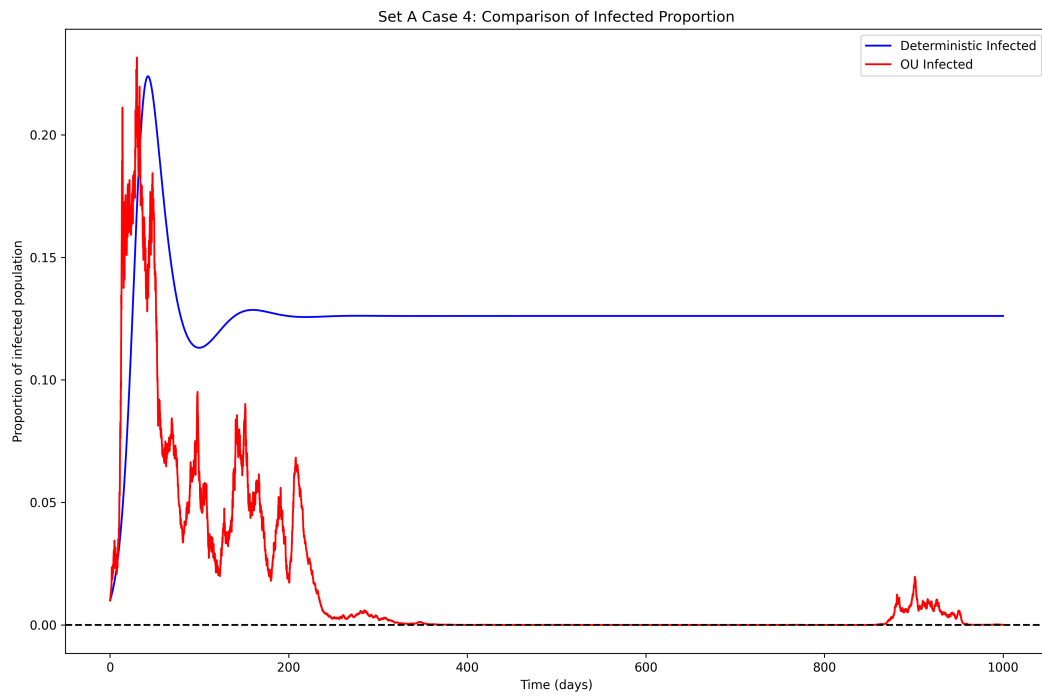


Figure 5.17: A plot of the deterministic solution along with the simulated SIRS model with the OU transmission process. The data set is A and the case number is 4.

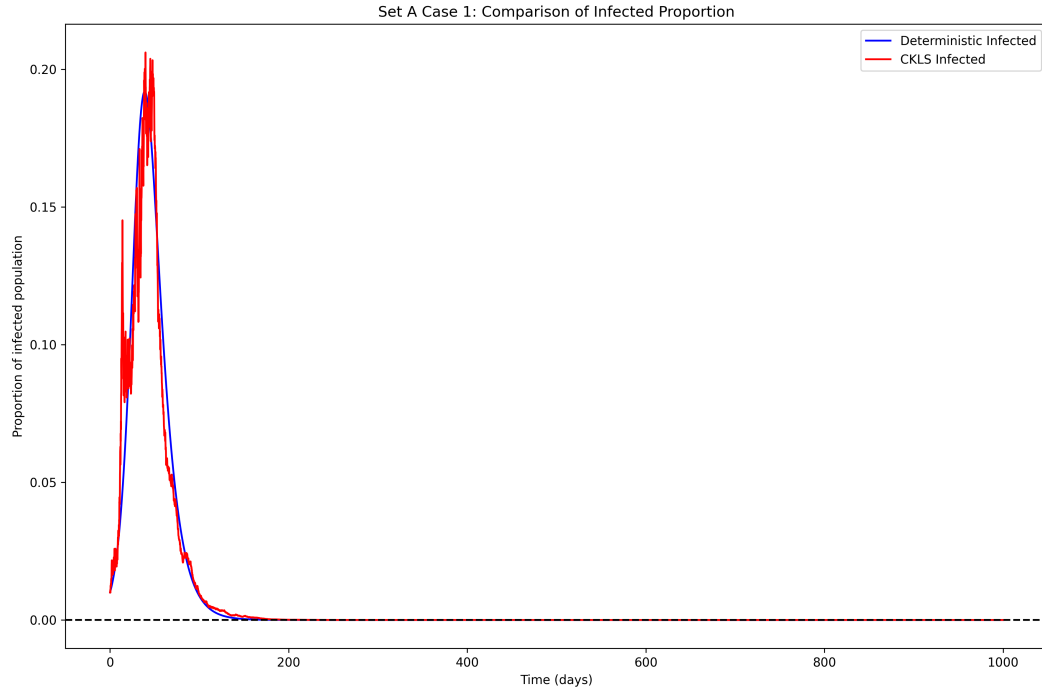


Figure 5.18: A plot of the deterministic solution along with the simulated SIRS model with the CKLS transmission process. The data set is A and the case number is 1.

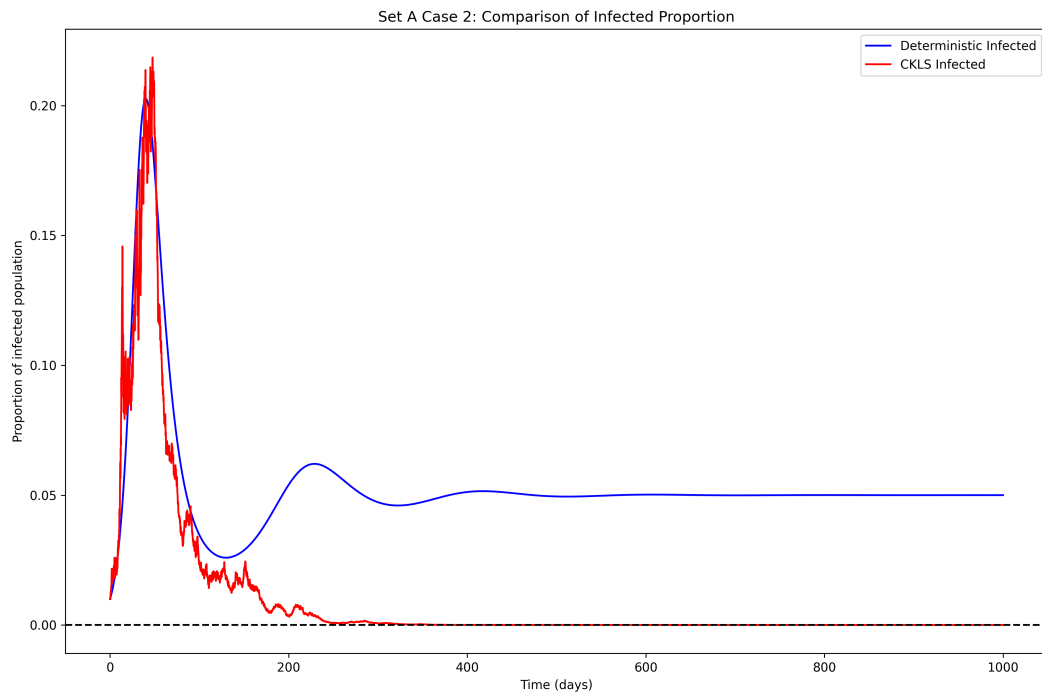


Figure 5.19: A plot of the deterministic solution along with the simulated SIRS model with the CKLS transmission process. The data set is A and the case number is 2.

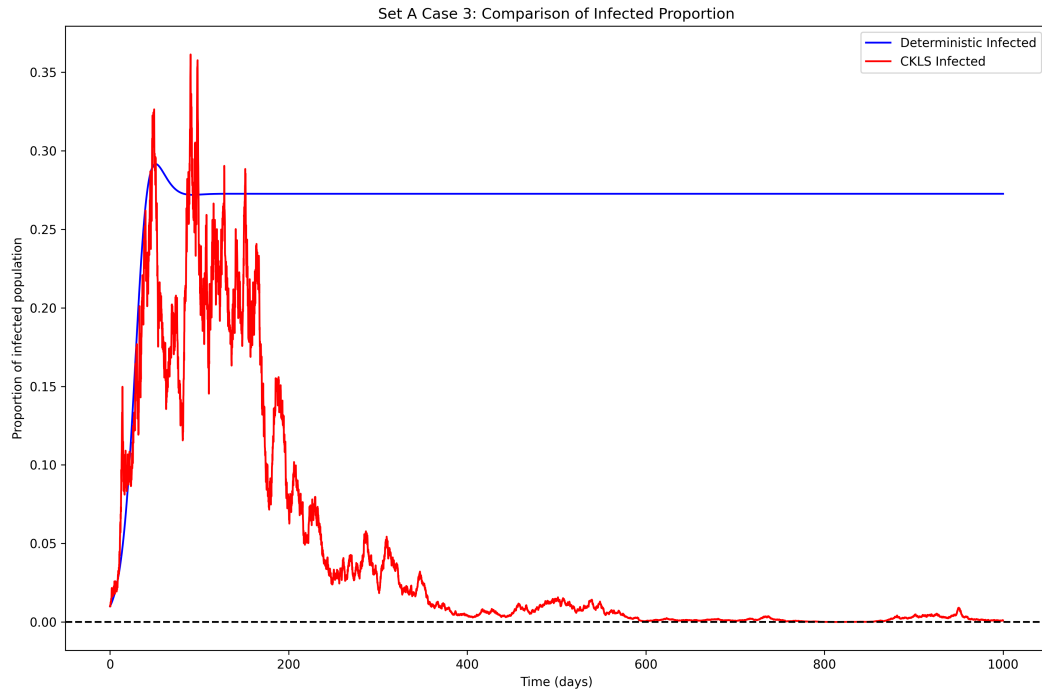


Figure 5.20: A plot of the deterministic solution along with the simulated SIRS model with the CKLS transmission process. The data set is A and the case number is 3.

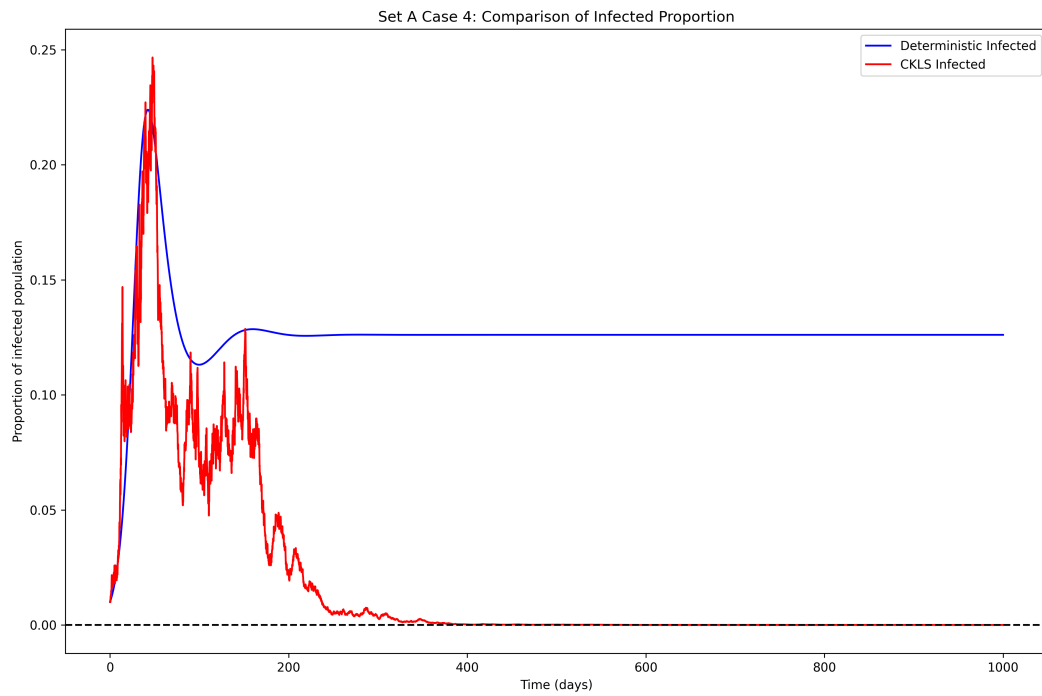


Figure 5.21: A plot of the deterministic solution along with the simulated SIRS model with the CKLS transmission process. The data set is A and the case number is 4.

Case	β_e	γ	θ	ρ	$\mathcal{R}_0$	$\mathcal{T}_0^{\text{OU}}$	$\mathcal{T}_0^{\text{CKLS}}$	$\mathcal{T}_0^{\text{LGBM}}$	$\mathcal{T}_0^{\text{Jacobi}}$
Set A									
Case 1	0.22	0.00	0.4	0.25	2.1978	1.8076	1.8396	1.9964	2.1322
Case 2	0.22	0.01	0.4	0.25	2.1978	1.8076	1.8396	1.9964	2.1322
Case 3	0.22	0.10	0.4	0.25	2.1978	1.8076	1.8396	1.9964	2.1322
Case 4	0.22	0.03	0.4	0.25	2.1978	1.8076	1.8396	1.9964	2.1322
Set B									
Case 1	0.20	0.01	0.4	0.25	1.0526	0.8470	0.8650	0.9361	1.0211
Case 2	0.22	0.01	0.6	0.29	1.1579	0.9735	0.9884	1.0054	1.1114
Case 3	0.35	0.01	1.3	0.40	1.8421	1.6802	1.6892	1.5667	1.7030
Case 4	0.188	0.01	0.4	0.25	0.9895	0.7839	0.8025	0.8730	0.9598

Table 5.3: Merged parameter settings for Set A and Set B with calculated $\mathcal{R}_0$ and stochastic thresholds for OU, CKLS, Logistic GBM, and Jacobi processes.

In the following discussion, please refer to the remaining plots for the other processes in the Appendix. The theoretical thresholds derived for the various stochastic models, such as $\mathcal{T}_0^{\text{OU}}$, $\mathcal{T}_0^{\text{CKLS}}$, $\mathcal{T}_0^{\text{LGBM}}$, and $\mathcal{T}_0^{\text{Jacobi}}$, are asymptotic indicators that describe the disease's long-term tendency to persist or fade out under certain probabilistic frameworks. While these thresholds are invaluable for understanding the average or expected behaviour of an epidemic process over an infinite horizon, they do not constitute deterministic guarantees of persistence in every individual realization, particularly in finite-time simulations.

In the context of Set B, Case 2, it is observed that the LGBM and Jacobi models yield theoretical thresholds slightly above one, suggesting that, under stationary or long-run conditions, there may be a propensity for the disease to persist. Nevertheless, the actual simulated trajectories show that the proportion of infected individuals eventually declines to zero. The presence of non-zero infection levels around $T = 550$ days in the Jacobi model is entirely consistent with early transient fluctuations. Over time, these infected populations diminish, culminating in extinction by $T = 1000$ days.

This apparent discrepancy between theoretical thresholds and observed outcomes arises from several factors inherent to stochastic systems. Firstly, the thresholds reflect expected tendencies rather than certainties. While a threshold above one indicates a possible bias towards persistence, individual stochastic paths can deviate substantially from this expectation, especially over finite durations. Random perturbations, small population sizes, and chance events can all facilitate the infected fraction's descent to zero.

Secondly, the distinction between long-term asymptotic behaviour and finite-time realizations is crucial. As time progresses, there may be numerous opportunities for infection chains

to terminate, particularly when the number of infected individuals becomes very small. Once the process drifts downward due to random fluctuations, it can settle into extinction. Even if, over an infinite horizon, there might have been non-negligible probabilities of persistence, the finite time frame observed does not preclude the disease from dying out.

Lastly, practical considerations, such as parameter estimation and environmental noise, may also influence the trajectory. The slight theoretical advantage implied by a threshold just above one can be outweighed by random shocks in the actual simulated environment, leading to the infected fraction reaching zero.

The extinction observed in the simulations is not contradictory to the threshold analysis. Rather, it demonstrates that these thresholds are probabilistic indicators of long-term tendencies. Actual epidemic realizations, especially over finite time intervals, can and do produce paths that end in extinction despite theoretical conditions that would allow for persistence on average. This outcome shows the inherently stochastic nature of epidemics and the importance of interpreting thresholds as guidelines rather than absolute rules.

5.7.6 Transmission process simulations

In this section, we present the results of simulations for different stochastic processes used to model transmission rates in an epidemic. The transmission processes are evaluated across two distinct parameter sets, referred to as **Set A** and **Set B**, each consisting of multiple cases with varying parameter values for the processes.

The following figures illustrate plots of these transmission processes for both **Set A** and **Set B**, with the processes displayed from top to bottom in the order specified. Each set is evaluated across four cases, where different parameter settings for the transmission processes are used.

The empirical long-run statistics for these transmission processes are summarized in the accompanying table. The table provides insights into the mean and standard deviation of the transmission rates for each process, allowing for a comparison of how each process behaves under different scenarios and parameter settings.

The figures below present plots for the transmission processes in **Set A** and **Set B, Case 1**; the remaining plots for the other Set B settings can be found in the Appendix.

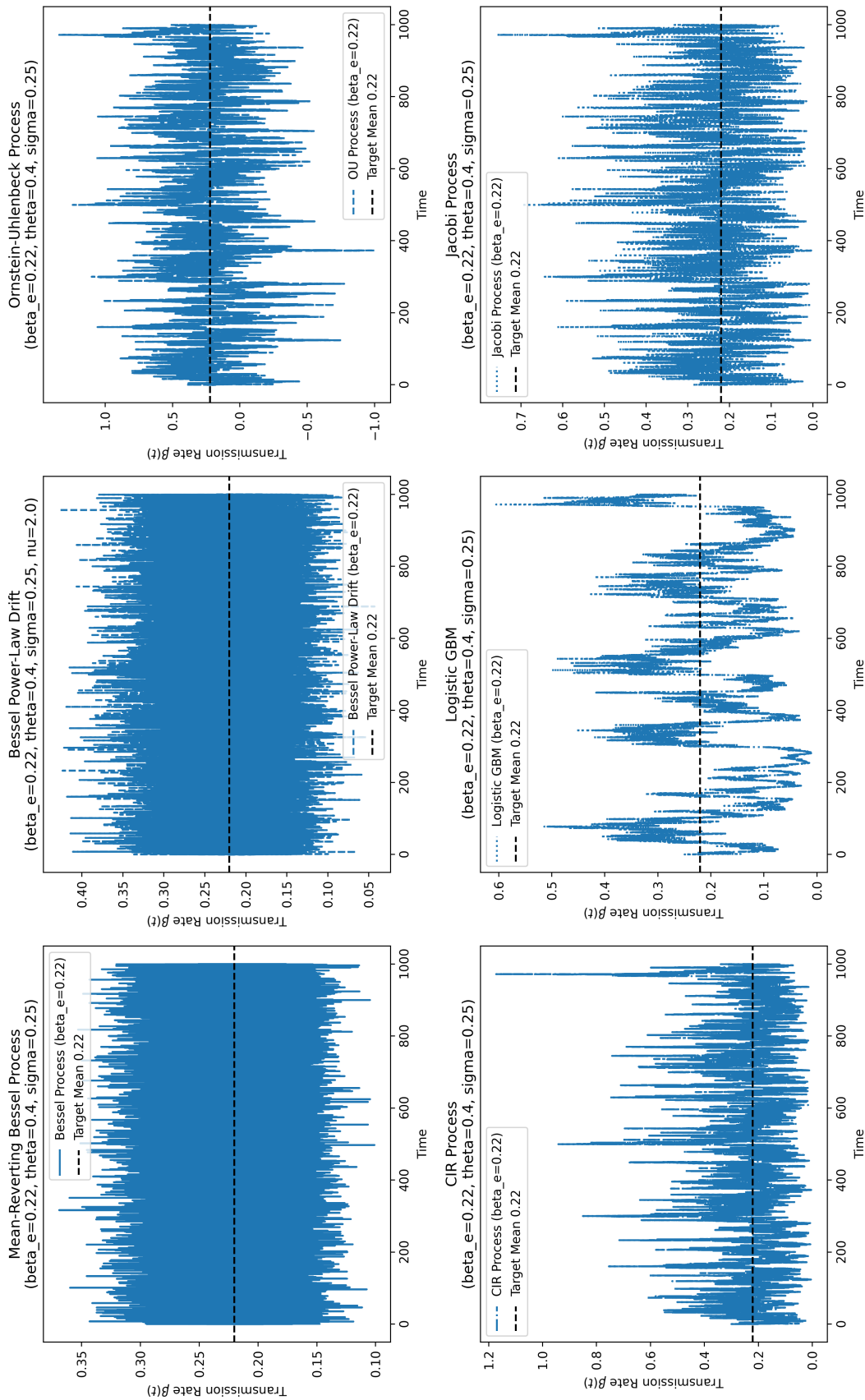


Figure 5.22: Plots of the transmission processes for Set A. Top row, from left to right: mean-reverting Bessel process, mean-reverting power law Bessel process, OU process; Bottom row, from left to right: CKLS process, Logistic GBM process, Jacobi process.

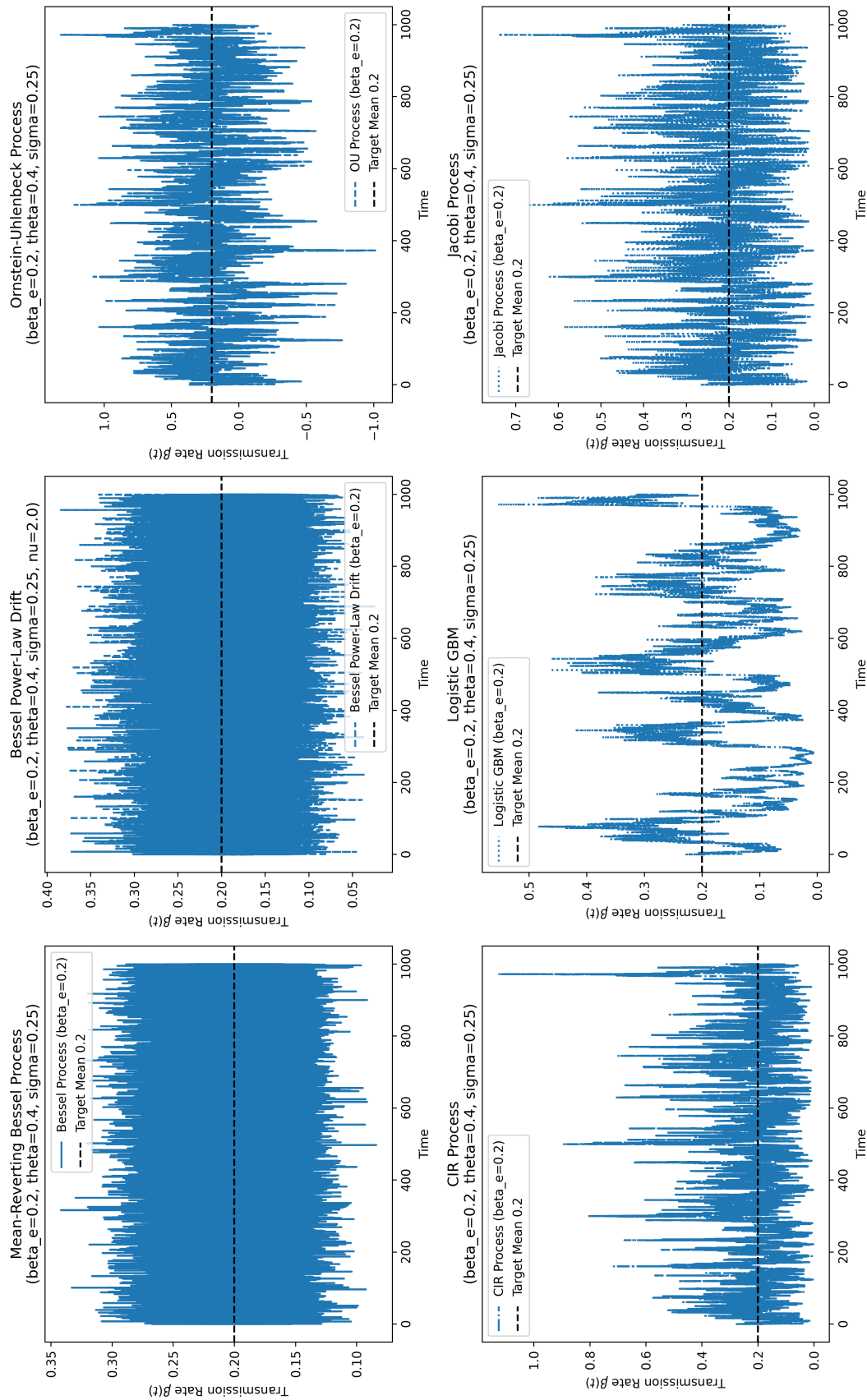


Figure 5.23: Plots of the transmission processes for Set B, Case 1. Top row, from left to right: mean-reverting Bessel process, mean-reverting power law Bessel process, OU process; Bottom row, from left to right: CKLS process, Logistic GBM process, Jacobi process.

Case	Process	Empirical Mean	Empirical Std Dev
Set A			
1	OU	0.25026	0.26783
1	CKLS	0.23149	0.13523
1	Bessel	0.22474	0.03158
1	Bessel (Power-Law Drift)	0.22567	0.04685
1	Logistic GBM	0.19432	0.08778
1	Jacobi	0.23005	0.11183
2	OU	0.19574	0.27219
2	CKLS	0.20675	0.11575
2	Bessel	0.22414	0.03108
2	Bessel (Power-Law Drift)	0.22438	0.04661
2	Logistic GBM	0.14609	0.08370
2	Jacobi	0.20976	0.10317
3	OU	0.21034	0.28471
3	CKLS	0.21678	0.13446
3	Bessel	0.22438	0.03159
3	Bessel (Power-Law Drift)	0.22480	0.04716
3	Logistic GBM	0.15512	0.09940
3	Jacobi	0.21648	0.11414
4	OU	0.23615	0.27488
4	CKLS	0.22721	0.13178
4	Bessel	0.22455	0.03128
4	Bessel (Power-Law Drift)	0.22528	0.04638
4	Logistic GBM	0.17989	0.10256
4	Jacobi	0.22636	0.11193
Set B			
1	OU	0.23026	0.26783
1	CKLS	0.21084	0.12932
1	Bessel	0.20457	0.02957
1	Bessel (Power-Law Drift)	0.19573	0.04405
1	Logistic GBM	0.17141	0.08467
1	Jacobi	0.20955	0.10844
2	OU	0.20141	0.26069
2	CKLS	0.21021	0.11324
2	Bessel	0.22547	0.03560
2	Bessel (Power-Law Drift)	0.22026	0.05322
2	Logistic GBM	0.15584	0.08181
2	Jacobi	0.21238	0.10028
3	OU	0.34500	0.24907
3	CKLS	0.34735	0.14950
3	Bessel	0.35676	0.04998
3	Bessel (Power-Law Drift)	0.35263	0.09129
3	Logistic GBM	0.32390	0.12611
3	Jacobi	0.34730	0.11612
4	OU	0.20415	0.27488
4	CKLS	0.19476	0.12197
4	Bessel	0.19236	0.02826
4	Bessel (Power-Law Drift)	0.17666	0.04196
4	Logistic GBM	0.15510	0.11018
4	Jacobi	0.19406	0.10583

Table 5.4: Empirical means and standard deviations for Set A and Set B across different processes

5.7.7 Feasible extinction region analysis

In this section, we analyze the conditions under which the transmission rate of a disease either persists or leads to extinction using two stochastic processes: the OU process and the CIR process. The key metric for determining whether the disease dynamics lead to extinction or persistence is the stochastic threshold, denoted by $\mathcal{T}_0^{\text{OU}}$ for the OU process and $\mathcal{T}_0^{\text{CIR}}$ for the CIR process. The stochastic threshold $\mathcal{T}_0$ determines the balance between the forces driving the transmission rate upwards (persistence) or towards extinction. When $\mathcal{T}_0 \leq 1$, extinction is guaranteed, while persistence occurs if $\mathcal{T}_0 > 1$.

The OU process is a stochastic model that describes fluctuations around a mean value. In the context of disease modelling, it represents how the transmission rate tends to revert to a mean level but is influenced by random shocks. In order to guarantee extinction, the stochastic threshold must satisfy $\mathcal{T}_0^{\text{OU}} < 1$. This defines the feasible region for extinction, which is identified by plotting $\mathcal{T}_0^{\text{OU}}$ as a function of θ and ρ .

The analysis yields feasible region plots showing the combinations of ρ and θ where $\mathcal{T}_0^{\text{OU}} < 1$, ensuring extinction. As with the OU process, extinction is guaranteed when $\mathcal{T}_0^{\text{CIR}} < 1$. The square root dependence on ρ reflects the dynamics of the CIR model, where fluctuations in the transmission rate are more gradual.

The analysis of feasible extinction regions for both the OU and CIR processes allows us to understand the conditions that lead to the extinction of the transmission rate. By determining the stochastic thresholds $\mathcal{T}_0^{\text{OU}}$ and $\mathcal{T}_0^{\text{CIR}}$, we can identify the combinations of ρ and θ that guarantee extinction, which is essential for designing effective interventions to control disease outbreaks. Visualizations of these regions are provided by plots contained in the Appendix.

5.8 Discussion

This work provides a significant advancement in the field of epidemiological modelling by integrating stochastic mean-reverting processes into the classical SIRS model. The introduction of processes such as the OU and CKLS models allows for a more realistic representation of transmission rate fluctuations, which are often driven by external factors like interventions or seasonal variations. By addressing the limitations of traditional deterministic models, the work presents a flexible framework that better captures the inherent variability in disease dynamics.

A key contribution of the work is the derivation of extinction thresholds under different stochastic transmission processes. These thresholds provide a quantitative measure to predict

whether a disease will persist or die out, offering important insights for public health policies. The analysis of stochastic processes such as the CKLS, OU, and Bessel models allows for a robust comparison of deterministic and stochastic approaches, revealing the advantages of using mean-reverting processes, particularly for maintaining positivity in transmission rates.

The numerical simulations provide strong validation for the theoretical findings. By comparing different parameter sets, the work effectively illustrates the long-term behaviour of the infected population under both deterministic and stochastic models. The inclusion of multiple stochastic processes illustrates the flexibility of the framework in adapting to a varied assortment of epidemiological settings. This numerical validation strengthens the practical relevance of the model under simulated settings and provides clear evidence of the versatility of stochastic models and their suitability for capturing real-world epidemic dynamics.

However, despite the strengths of the work, there are areas for improvement. The complexity of the mathematical framework, while rigorous, may limit the accessibility of the model to non-experts. The reliance on advanced stochastic calculus, such as Itô's lemma and the Fokker–Planck equation, might present a steep learning curve for practitioners in public health or epidemiology who are less familiar with these techniques. Additionally, the model assumptions, such as a constant long-term equilibrium mean transmission rate, may not fully capture the dynamic nature of real epidemics, where transmission rates can shift due to evolving external factors. Such instances may be better suited by incorporating a mechanism for regime changes in the transmission level.

5.9 Conclusion

This work presents a critical advancement in epidemiological modelling by integrating stochastic transmission processes into the classical SIRS framework. The incorporation of mean-reverting processes such as the CKLS and modified Bessel models allows for a more realistic and flexible representation of disease transmission, addressing the limitations of deterministic models. The key contributions include the derivation of stochastic extinction thresholds and the introduction of novel mean-reverting processes like the Bessel and modified Bessel processes.

The theoretical findings are supported by extensive numerical simulations, which highlight the model's capacity to accurately capture the variability in transmission rates and provide insights into disease persistence and eradication. The comprehensive approach contributes to the growing body of literature on stochastic epidemiological models and offers valuable tools for public health policy and intervention strategies.

Nevertheless, there are limitations that should be addressed in future work. Applying the model to real-world epidemic data, such as from past influenza or COVID-19 outbreaks, could significantly enhance its practical relevance. Additionally, the intensive mathematical framework could be adapted and simplified to make the model more accessible to a broader audience. Despite these challenges, the model provides a powerful framework for understanding disease dynamics in the presence of stochastic variability, and it lays a solid foundation for further research at the frontier of the field.

Chapter 6

Multiperiod Forecasting of Epidemic Wave Peak Time and Height

This chapter presents a comprehensive analysis of machine learning models for predicting influenza peaks across multiple seasons. The study evaluates both traditional and deep learning models, including Linear Regression, Random Forest, LSTM, Support Vector Machines (SVM), XGBoost, Adaboost, and Recurrent Neural Networks (RNN), with a focus on improving the accuracy of peak prediction for influenza outbreaks. Using historical data from the Southern Hemisphere and covariate inclusion/exclusion scenarios, we assess model performance across various influenza seasons from 2013 to 2020.

Our analysis reveals that simpler models like Linear Regression and Lasso consistently offer robust and reliable predictions with lower error margins, particularly in early and steady-peak influenza seasons. More complex models, such as LSTM and Random Forest, demonstrate greater predictive power in high-variability seasons but also show higher variability and risk of overfitting. Models incorporating covariates, such as Simple RNN and Temporal CNN, further highlight the importance of context-specific information in improving predictive accuracy.

Despite their potential, deep learning models exhibit several limitations, including sensitivity to data quality, season-specific variability, and a lack of transparency, making them less suitable for public health decision-making without significant calibration. The work concludes by recommending hybrid approaches that integrate mechanistic epidemiological models with machine learning to enhance real-time predictive capacity and account for novel outbreak dynamics. Additionally, we propose future research directions, including the development of models that incorporate external factors, such as vaccination rates and mobility data, to improve influenza peak prediction and contribute to more timely public health interventions.

6.1 Introduction

Influenza remains one of the most significant public health challenges, causing seasonal outbreaks that result in substantial morbidity, mortality, and economic burden. Predicting the peak of influenza cases, both in terms of timing and intensity, is critical for enabling timely public health responses, optimizing healthcare resources, and mitigating the impact of outbreaks. Traditional epidemiological models have provided foundational insights into the spread of infectious diseases, but the growing availability of data and advances in machine learning offer new opportunities to enhance predictive accuracy.

Machine learning, a significant subfield of artificial intelligence, concentrates on developing algorithms and models that empower computers to learn from data and make predictions or decisions without being explicitly programmed. This field involves designing systems capable of automatically improving their performance by analyzing and extracting patterns from vast amounts of data. Such systems adapt through experience, enabling them to handle complex tasks that are difficult to define procedurally.

In the context of influenza peak prediction, machine learning offers powerful tools for forecasting future outbreaks based on historical trends. In this study, we explore the application of various machine learning models for predicting influenza peaks. Our objective is to evaluate and compare the performance of different models, including traditional approaches like Linear Regression and Lasso, more advanced ensemble methods like Random Forest and XGBoost, and deep learning architectures such as Long Short-Term Memory (LSTM) networks and Recurrent Neural Networks (RNN). These models are tested across multiple influenza seasons, from 2013 to 2020, with specific attention to how their performance varies when covariates, such as data from the Southern Hemisphere, are included or excluded.

The motivation behind this analysis stems from the increasing demand for more accurate and adaptive models that can handle the complexities of influenza outbreaks. Traditional models, while useful, often rely heavily on assumptions about the disease's dynamics, which may not capture the full range of variability seen in real-world data. Machine learning models, particularly deep learning methods, have the potential to uncover complex patterns in historical influenza data, potentially leading to better predictions of peak case numbers and timing. However, these models also come with challenges, including overfitting, sensitivity to data quality, and lack of interpretability, which limits their current use in public health contexts.

A critical consideration in predictive analytics, particularly in time series forecasting, is addressing fitting issues, namely overfitting and underfitting [105]–[107]. Overfitting occurs when a model performs exceptionally well on the training data but poorly on the test data. It

indicates that the model has learned the noise and fluctuations specific to the training dataset rather than the underlying general patterns. Overfitting is associated with high variance and low bias, resulting in a model that does not generalize well to new data. Underfitting, in contrast, happens when a model is too simplistic to capture the true patterns in the data. It fails to perform well on both the training and test datasets, reflecting high bias and low variance. An underfit model does not adequately learn from the data, leading to suboptimal predictions.

Understanding the concepts of bias and variance is crucial in building effective machine learning models. In statistical learning theory, bias is defined as the systematic difference between the average predictions of a model (over different training sets drawn from the same distribution) and the true underlying values. A model with low bias produces predictions that, on average, are close to the true target values. In contrast, a model with high bias tends to consistently deviate from the truth, often because it is too simple or makes overly restrictive assumptions, causing it to miss important patterns. Variance measures the variability of the model's predictions across different training datasets. High variance suggests that the model is sensitive to fluctuations in the training data, often capturing noise, which leads to overfitting. Balancing bias and variance is essential to develop models that generalize well to unseen data.

To illustrate these fitting issues, consider the following scenarios:

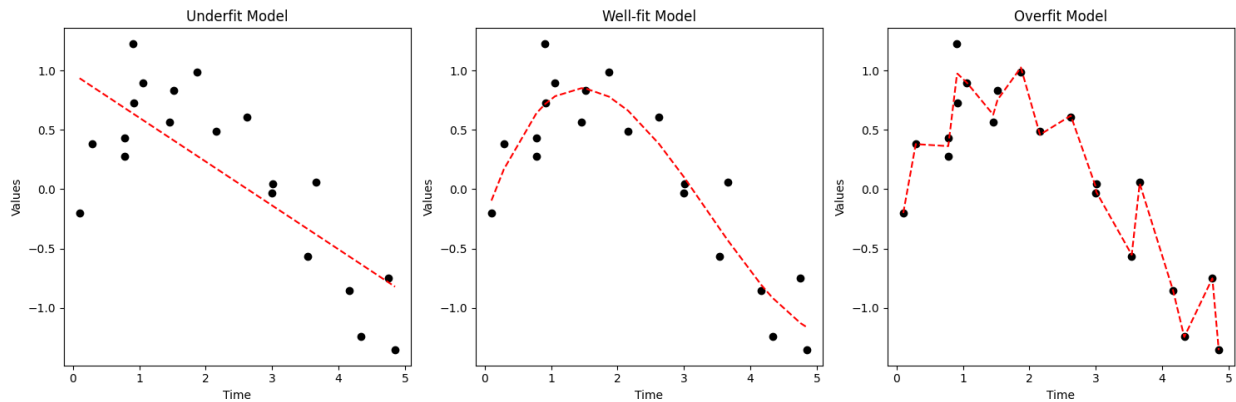


Figure 6.1: Different fitting situations: (a) Underfitting, (b) Appropriate fitting, (c) Overfitting.

An underfit model fails to capture the real relationships within the data, resulting in poor predictions for both historical and future values. A well-fit model effectively captures the general tendencies of the data, providing satisfactory predictions that generalize well to new observations. An overfit model captures every detail of the training data, including noise and outliers, which can lead to inaccurate predictions on new data due to its inability

to generalize.

Several factors contribute to overfitting. Insufficient data can prevent the model from learning the true underlying patterns due to a lack of diverse training data. Overly complex models may learn noise instead of meaningful patterns if they are excessively complex. Additionally, feature overload, which involves including irrelevant or noisy features, can confuse the model, leading to overfitting. Another cause is the lack of regularization, where the absence of proper regularization techniques fails to constrain model complexity effectively. Also, data leakage, which occurs when test data is inadvertently included in the training set, can inflate performance metrics and lead to overfitting.

Underfitting can result from several factors. First, insufficient model complexity occurs when simple models fail to capture the complexity of the data patterns. Additionally, insufficient training, where limited training time or data is available, may prevent the model from learning effectively. Another cause is over-regularization, where excessive regularization overly constrains the model, hindering its ability to learn. Lastly, poor feature selection, using non-informative or irrelevant features, fails to provide the necessary information for accurate predictions.

To mitigate fitting biases in time series prediction, several guidelines are recommended. First, increasing the amount of training data helps the model capture a broader range of patterns and reduces the likelihood of overfitting. Second, selecting relevant features carefully minimizes noise and complexity, enabling the model to learn meaningful relationships. Additionally, employing regularization techniques, such as Lasso or Ridge regression, helps control model complexity by penalizing excessive weights, thus preventing overfitting. Optimizing hyperparameters through methods like cross-validation ensures the model is configured effectively for the given data. Furthermore, applying ensemble methods, which combine predictions from multiple models, enhances robustness and predictive performance. Finally, regularly evaluating the model's performance on validation sets or unseen data helps detect and address fitting issues early in the development process.

Evaluating the model's performance is a critical step in selecting the most appropriate predictive model. It is important to recognize that strong performance on historical data does not guarantee future success. Common evaluation metrics include accuracy, which measures the proportion of correct predictions out of all predictions made and is calculated as:

$$\text{Accuracy} = 100 \times \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}.$$

Another key metric is the Mean Absolute Error (MAE), which quantifies the average magnitude of errors in a set of predictions and is less sensitive to outliers, making it suitable for

datasets with extreme values:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|.$$

The Mean Squared Error (MSE) calculates the average of the squares of the errors, emphasizing larger discrepancies due to squaring differences:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2.$$

Lastly, the Bias Ratio assesses the directional bias of the model’s predictions by comparing the number of overestimates to underestimates:

$$\text{Bias Ratio} = \frac{\text{Number of Overestimates}}{\text{Number of Underestimates}}.$$

An ideal bias ratio is approximately 1.00, indicating a balanced system without systematic overprediction or underprediction.

When evaluating models using these metrics, several practical considerations should be taken into account. Comparative analysis involves comparing the evaluation metrics across multiple regression models to identify the best performer in terms of predictive accuracy and reliability. Establishing thresholds by defining acceptable levels of prediction error helps determine whether a model meets the necessary performance criteria for the application. It is also important to analyze training vs. testing performance, as evaluating the loss functions for both can reveal overfitting or underfitting, which can guide further model refinement. Additionally, sensitivity to outliers should be considered when choosing between MAE and MSE, since MAE is less sensitive to outliers while MSE penalizes larger errors more heavily. Finally, focusing on model improvement by emphasizing metrics like MSE can help reduce significant prediction errors, leading to overall enhancements in model performance.

This work is structured as follows: Section 1 includes a literature review of related work. Section 2 provides an overview of the data and methods used in this study. Section 3 presents the results of model performance across different influenza seasons, highlighting the role of covariate inclusion and exclusion. Section 4 discusses key findings, limitations, and insights on model suitability for various outbreak patterns. Section 5 concludes with recommendations for improving predictive models for influenza peaks and directions for future research.

Through this work, we aim to provide a deeper understanding of the strengths and limitations of machine learning models for influenza prediction and to offer practical recommendations for their application in public health. In particular, we explore how simpler

models like Linear Regression and Lasso maintain reliable performance across seasons, while more complex models like LSTM and RNN can offer improved predictions but require careful calibration and handling of covariates. Our findings highlight the importance of balancing predictive power with model interpretability and real-world applicability.

The primary objective of this research is to improve the accuracy of influenza peak prediction by evaluating and comparing multiple machine learning models. Specific aims include investigating the performance of traditional machine learning models, such as Linear Regression and Lasso, as well as advanced models, including LSTM, RNN, and XGBoost, for influenza peak prediction. Additionally, the study aims to assess the impact of covariate inclusion and exclusion, particularly Southern Hemisphere influenza data, on model performance. Another goal is to identify suitable models for different types of influenza peaks, such as early peaks, severe outbreaks, and late peaks with gradual onset. Finally, the research seeks to provide recommendations for future studies and practical applications in public health decision-making.

By addressing these objectives, this work aims to contribute to the development of more reliable and robust models for real-time influenza surveillance and prediction, ultimately supporting public health efforts to mitigate the impact of seasonal outbreaks.

Review of related work

Kim et al., 2024

Kim et al. (2024) [108] combine Bayesian Markov Switching Models with the Susceptible-Infectious-Recovered (SIR) model to detect season onset through dynamic state changes and predict outbreak peaks using regime-switching probabilities. Their approach aims to predict the peak timing of influenza-like illness (ILI) outbreaks by identifying regime shifts rather than directly forecasting peaks. The study emphasizes the importance of accurate peak timing for effective public health response.

The model utilizes advanced statistical techniques to detect shifts in epidemic states. However, peak timing predictions show significant variability across seasons, with errors ranging from -10 to $+26$ weeks. Error statistics for the Markov Switching model across different prediction windows are as follows: 3 weeks after onset, the mean error is 1.93 weeks with a standard deviation of 10.67 weeks; 4 weeks after onset, the mean error is 3.40 weeks with a standard deviation of 12.15 weeks; 5 weeks after onset, the mean error is 2.60 weeks with a standard deviation of 9.41 weeks; and 6 weeks after onset, the mean error is 3.93 weeks with a standard deviation of 9.59 weeks.

The model does not account for the severity of outbreaks, focusing solely on timing. Large prediction errors are observed, frequently exceeding two months (8 weeks). For the 2014 season, the error was in excess of +20 weeks (5 months), regardless of the number of weeks from the time of onset to prediction.

The model exhibits a lack of reliability, with high error variability across seasons and no consistent improvement. Performance degrades significantly in atypical or high-variability seasons.

The model does not estimate the intensity or severity of the ILI outbreak. Accurate severity prediction is crucial for planning and resource allocation, as hospitals and health services need to know not just when but also how severe outbreaks will be. Failure to predict severity limits the model’s utility in emergency preparedness.

The methodological shortcomings of the current model primarily arise from its overemphasis on detecting regime shifts rather than directly predicting peak times. While regime shifts are important, they do not necessarily correlate with the timing of influenza-like illness case peaks. Furthermore, the model’s high computational complexity provides limited practical advantage over simpler approaches such as linear regression, which have demonstrated lower errors in certain scenarios. Another significant issue is the unrealistic assumption that epidemic transitions can be captured by discrete regime switches, whereas in reality, influenza dynamics are more continuous and complex.

In terms of challenges for public health adoption, the model suffers from high prediction errors, which diminish its credibility and usability for decision-makers who rely on precise and reliable forecasts. Additionally, the model lacks actionable insights regarding outbreak severity, which limits its practical utility in decision-making processes. Another barrier to adoption is the model’s technical complexity, making it difficult for non-experts to use effectively. A simpler, more user-friendly model would enhance adoption and ease of use.

To improve the model, several recommendations can be made. First, the model should be simplified to make it more intuitive and easier to interpret. Second, it should be extended to incorporate predictions about the severity of outbreaks, not just their timing. Finally, validating the model with real-world, diverse, and realistic datasets would help ensure its broad applicability and reliability across different scenarios.

While the model by Kim et al. (2024) is innovative, its practical utility is limited due to high prediction errors and lack of severity estimation. Future improvements should focus on increasing accuracy, simplicity, and relevance to real-world public health needs.

Morbey et al., 2023

Morbey et al. (2023) [109] utilize a range of machine learning models to forecast seasonal epidemic peaks, focusing on Respiratory Syncytial Virus (RSV) activity, including atypical ‘black swan’ seasons. The aim is to provide early warning of peak activity for public health interventions, highlighting the importance of forecasting epidemic peaks in real-time to mitigate public health impacts.

They tested multiple regression models, including linear regression, generalized linear models, random forest, and support vector machines (SVM). A novel ‘peak error’ measure was developed to evaluate models on peak timing and intensity accuracy. It was found that models optimized for daily forecasts were not necessarily the best for peak forecasts. The models were validated using both typical and atypical seasons.

Models that performed well in typical seasons did not generalize well to atypical seasons. High variability in peak error rates was noted, particularly in seasons with unexpected patterns. Additionally, the models struggled with ‘black swan’ events where seasonality diverged significantly from historical data. Overfitting to historical patterns reduced adaptability to novel epidemic scenarios.

Including seasonality variables improved forecasts for typical seasons but worsened them for atypical seasons. Models were unable to adequately adjust to unexpected seasonal shifts without significant error. Moreover, forecasts are heavily dependent on historical seasonality patterns, limiting their robustness due to a lack of mechanisms to adapt dynamically to real-time changes in disease spread patterns.

Multiple model variants and combinations increase complexity but do not necessarily improve practical utility. Complex models such as random forests or SVMs are difficult to interpret, reducing transparency for public health officials. Additionally, these models require constant recalibration and validation, limiting their usability in real-time public health scenarios.

The models face a high risk of overfitting to typical seasons, which reduces their reliability for atypical events. Predictions provide limited guidance on specific public health actions or interventions. Furthermore, a lack of clear communication on forecast uncertainty, especially during atypical seasons, limits trust in model outputs.

Models with fixed seasonality assumptions fail to capture the dynamics of atypical outbreaks. There is a need for more flexible models that can handle both typical and atypical patterns without compromising accuracy. Excessive reliance on seasonality can lead to poor performance during unexpected shifts, as seen in the 2020/21 ‘black swan’ season.

To improve the models, it is recommended to incorporate real-time data adjustments,

allowing models to adapt dynamically based on real-time surveillance data. Simplifying model complexity by reducing the number of model variants would improve interpretability and ease of use. Additionally, better methods to communicate the uncertainty of forecasts to public health officials should be developed. Finally, ensuring that models are generalizable across different seasonal patterns and outbreak scenarios is critical.

The machine learning approach by Morbey et al. (2023) shows promise but has significant limitations in real-world application. Future research should focus on enhancing adaptability, reducing complexity, and improving the robustness of models against atypical seasonal events.

Zhang et al., 2019

Zhang et al. (2019) [110] utilize the Seasonal Autoregressive Integrated Moving Average (SARIMA) model, integrating cross-hemisphere influenza surveillance data and local internet query data. The aim is to predict seasonal influenza outbreaks and provide early warning for public health interventions, emphasizing the importance of accurate and timely predictions for effective public health response.

The SARIMA model effectively captures seasonal patterns and trends, integrating real-time data (internet search queries) to improve prediction timeliness. Performance metrics show high Pearson correlation coefficients and low Root Mean Squared Error (RMSE) values.

The model depends heavily on historical data patterns and is unable to adapt to unexpected changes or novel strains. It struggles to predict peak timing and intensity accurately in seasons with irregular patterns, with limited capability to handle non-stationary data or sudden epidemic changes.

The model does not incorporate key contextual factors affecting influenza transmission, such as demographic changes, vaccination rates, and public health interventions. Simplistic assumptions limit the model's applicability across diverse geographic regions and populations.

The model faces several methodological issues. First, it has limited real-time adaptability, relying on fixed seasonal patterns and not adjusting dynamically to real-time changes. Additionally, model performance is highly sensitive to parameter selection and tuning, where small parameter changes can lead to vastly different predictions, affecting reliability. Lastly, the complexity in model tuning and interpretation poses challenges for non-expert users, with a lack of transparency in understanding the impact of various inputs on predictions.

The model faces challenges in public health adoption due to its limited predictive utility. Predictions are often too generalized and lack specificity for localized outbreaks, making it

ineffective in predicting sudden epidemic spikes or downturns. The model’s outputs do not provide clear guidance on specific public health actions, including severity, resource needs, or optimal intervention timing. Moreover, the model is inadequate for emerging threats, as it is ineffective in adapting to new influenza strains or unforeseen epidemiological trends, and is not robust enough to handle rapidly evolving public health emergencies.

There is a risk of overfitting to historical data when the model is excessively tailored to past patterns, leading to poor generalization to future outbreaks or new data. Regular recalibration is required to maintain model accuracy, which poses significant resource challenges.

To address these issues, the model should incorporate additional data sources such as demographic trends, vaccination rates, and mobility data. Developing more adaptive models that dynamically adjust to real-time changes and new data inputs is essential. Additionally, simplifying model parameters and enhancing user understanding of prediction drivers would improve model transparency and interpretability. Finally, the model’s real-world validation should be strengthened by using diverse datasets from different regions and outbreak scenarios.

While the SARIMA model by Zhang et al. (2019) offers some utility, it has limitations that hinder its effectiveness in real-world public health settings. Future work should focus on developing more robust, adaptive models with greater practical utility for predicting influenza dynamics.

6.2 Methodology

6.2.1 Data collection and preprocessing

The dataset used in this study consists of influenza case counts collected from multiple regions, spanning the 2013-2020 influenza seasons [111]. Historical influenza data is sourced from public health databases, including both confirmed case counts and surveillance data. Covariate data, such as the influenza activity in the Southern Hemisphere, is also included to assess its impact on predictive performance.

The raw influenza data was preprocessed to handle missing values, outliers, and noise. For model inputs, case counts were normalized to reduce the impact of seasonality and allow for better model training. Covariates, including Southern Hemisphere influenza activity, were incorporated as time-lagged features to account for potential lead-lag relationships between the regions.

6.2.2 Modelling approaches

This study evaluates multiple machine learning models [105], [107], ranging from traditional regression techniques to more complex deep learning architectures. The models assessed are grouped into two categories: 1) traditional machine learning models and 2) deep learning models. Each model's performance is evaluated based on its ability to predict the peak of influenza case counts in terms of timing and intensity.

Linear Regression is a simple model that assumes a linear relationship between the predictors and the response variable, which in this case is influenza peak case counts. It serves as a baseline for comparison due to its simplicity and interpretability. Lasso Regression, or Least Absolute Shrinkage and Selection Operator, adds a regularization term to the standard linear regression model to prevent overfitting by penalizing large coefficients. This model is particularly useful for variable selection and simplifying models that include many covariates. The Stochastic Gradient Descent (SGD) Regressor is an iterative method for optimizing the loss function in linear regression models, making it particularly useful for handling large datasets and offering faster convergence compared to traditional gradient descent methods.

Random Forest is an ensemble learning method that constructs a multitude of decision trees during training. It offers robustness to overfitting and performs well with both covariate inclusion and exclusion. However, while Random Forest was initially tested, it was not among the top performers across all seasons and was omitted from the final performance rankings. XGBoost (Extreme Gradient Boosting) is another ensemble method that builds additive models sequentially, focusing on correcting mistakes made by prior models. XGBoost performed well in specific seasons and was included in the final analysis. Adaboost (Adaptive Boosting) is an ensemble model that adjusts the weights of misclassified observations at each iteration, improving the prediction accuracy over time.

Recurrent Neural Networks (RNNs) are designed for sequential data and are well-suited for time series forecasting. They maintain an internal state to capture temporal dependencies in the data. The Simple RNN model was effective at predicting influenza peaks, particularly when covariates from the Southern Hemisphere were included. Long Short-Term Memory (LSTM) networks are an extension of RNNs that can better capture long-term dependencies by using a gating mechanism to regulate the flow of information. The Simple LSTM model was tested with and without covariates and performed well in certain influenza seasons, though it exhibited variability across different data configurations.

A more complex LSTM model, the Deep LSTM, was employed to capture deeper, more intricate patterns in the influenza data. While this model showed promising results in some seasons, it demonstrated higher levels of variability and overfitting compared to simpler

models. Temporal Convolutional Neural Networks (CNNs) were also tested for their ability to capture local patterns in time series data through convolutional layers. Though not included in the final performance rankings, this model demonstrated moderate success in predicting the peaks during certain seasons.

6.2.3 Methodology

This section outlines the methodology used to predict future values of influenza-like illness (ILI) case counts in the northern hemisphere, denoted as y_t^{north} . The primary objective is to forecast not only the future trajectory of ILI incidences but also to accurately determine the timing and magnitude of the seasonal peaks. To achieve this, we begin with a simple linear regression framework that employs lagged values of the target variable as features. Initially, no additional covariates (such as southern hemisphere data or autocorrelation measures) are included. After establishing a baseline using this simpler model, we progressively expand the feature set by integrating more informative covariates, including southern hemisphere ILI data, measures of autocorrelation, and seasonal harmonics.

The target variable y_t^{north} represents the number of ILI cases at time t in the northern hemisphere. Our initial feature set, X_t , consists of the $L = 52$ most recent lagged values of y_t^{north} . Scaling techniques, such as MinMax scaling, are applied to ensure that features share comparable value ranges, thereby aiding in model stability and interpretability.

The initial linear regression model can be expressed as:

$$\hat{y}_t^{\text{north}} = \beta_0 + \sum_{i=1}^L \beta_i y_{t-i}^{\text{north}} + \epsilon_t,$$

where ϵ_t is the error term. In matrix form, this model is succinctly represented as $\hat{y}_t^{\text{north}} = \beta_0 + \mathbf{X}_t \boldsymbol{\beta} + \epsilon_t$, with $\mathbf{X}_t = [1 \ y_{t-1}^{\text{north}} \ \dots \ y_{t-52}^{\text{north}}]$ and $\boldsymbol{\beta} = [\beta_0 \ \beta_1 \ \dots \ \beta_{52}]^T$.

Parameter estimation is performed by minimizing the sum of squared errors (SSE) between observed and predicted values. The ordinary least squares (OLS) solution, often derived from the normal equation $\boldsymbol{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$, ensures that the chosen coefficients provide the best fit to the training data in a least-squares sense.

Prediction is then straightforward, as future values $\hat{y}_{t+1}^{\text{north}}$ are generated using the previously estimated coefficients $\boldsymbol{\beta}$ and the most recent lagged observations. Model performance is evaluated using various metrics, including the Root Mean Squared Error (RMSE), correlation between predicted and actual values, accuracy (based on the correct prediction of upward or downward trends), and model bias, defined as the ratio of positive to negative forecast errors.

After establishing this baseline, we enhance the model by incorporating additional covariates to improve predictive accuracy. For instance, southern hemisphere ILI data (y_t^{south}) can be included to capture inter-hemispheric relationships that might influence northern hemisphere ILI dynamics. Autocorrelation measures, introduced as additional features, help the model account for memory and feedback effects in the time series. Seasonal harmonics, constructed using sinusoidal functions, allow the model to represent cyclical patterns commonly observed in ILI data.

By refining the feature set, we aim to capture more complex relationships and improve the model’s predictive performance, especially in forecasting peak timing and intensity. While the initial approach relies solely on past northern hemisphere ILI counts, the subsequent steps systematically incorporate richer information to approach a more realistic and accurate representation of the underlying disease dynamics.

6.2.4 Model training and evaluation

Each model was trained on the influenza case counts from previous seasons, with hyperparameters tuned using grid search and rolling-window validation techniques. For deep learning models, the training process involved backpropagation and gradient descent algorithms, with early stopping employed to prevent overfitting using PyTorch and TensorFlow [112], [113].

Two covariate configurations were tested: one with Southern Hemisphere data included as a covariate, and the other excluding this data. The inclusion of Southern Hemisphere data was based on the hypothesis that the Southern Hemisphere’s influenza outbreak leads the Northern Hemisphere’s outbreak by several weeks, thus improving peak prediction accuracy. The second configuration, with Southern Hemisphere covariates excluded, allowed models to be trained solely on Northern Hemisphere data, reflecting their robustness under limited information.

To evaluate model performance, several metrics were used. Prediction Error (%) was the primary metric for assessing each model’s ability to predict the peak of influenza case counts, calculated as the percentage difference between predicted and actual peak values. Confidence Interval (CI) Error (%) was used to evaluate the accuracy of the predicted confidence intervals for peak case counts. Root Mean Squared Error (RMSE) measured the overall prediction accuracy, penalizing larger errors, while Mean Absolute Error (MAE) focused on the magnitude of errors without the squaring effect of RMSE. Finally, correlation coefficients were used to assess how well the predicted values correlated with actual case counts over time.

Several techniques were employed to address overfitting. Regularization terms were ap-

plied to models like Lasso and SGD Regressor to penalize large coefficients and reduce model complexity, improving generalization to unseen data. In deep learning models, such as LSTM and RNN, early stopping was employed during training to prevent overfitting on the training data. Rolling-window validation techniques were used to ensure model performance was evaluated over different time periods, providing a robust test of the model’s generalizability.

Models were evaluated across multiple influenza seasons (2013-2020) to capture the variability in their predictive performance. The peaks of each season varied in timing, height, and growth dynamics. For example, the 2013-14 season exhibited rapid growth and early peaks, while the 2014-15 season had severe, high curvature peaks. Each model’s performance was ranked by prediction error, with the results presented both for cases where covariates were included and excluded.

By analyzing the performance across different seasonal patterns, we identify which models are better suited to handle specific characteristics, such as rapid growth, high curvature, or late peaks with gradual onset. We comment on this in more detail later on.

6.3 Applications to understanding the relationship between influenza case counts in the Southern and Northern hemispheres

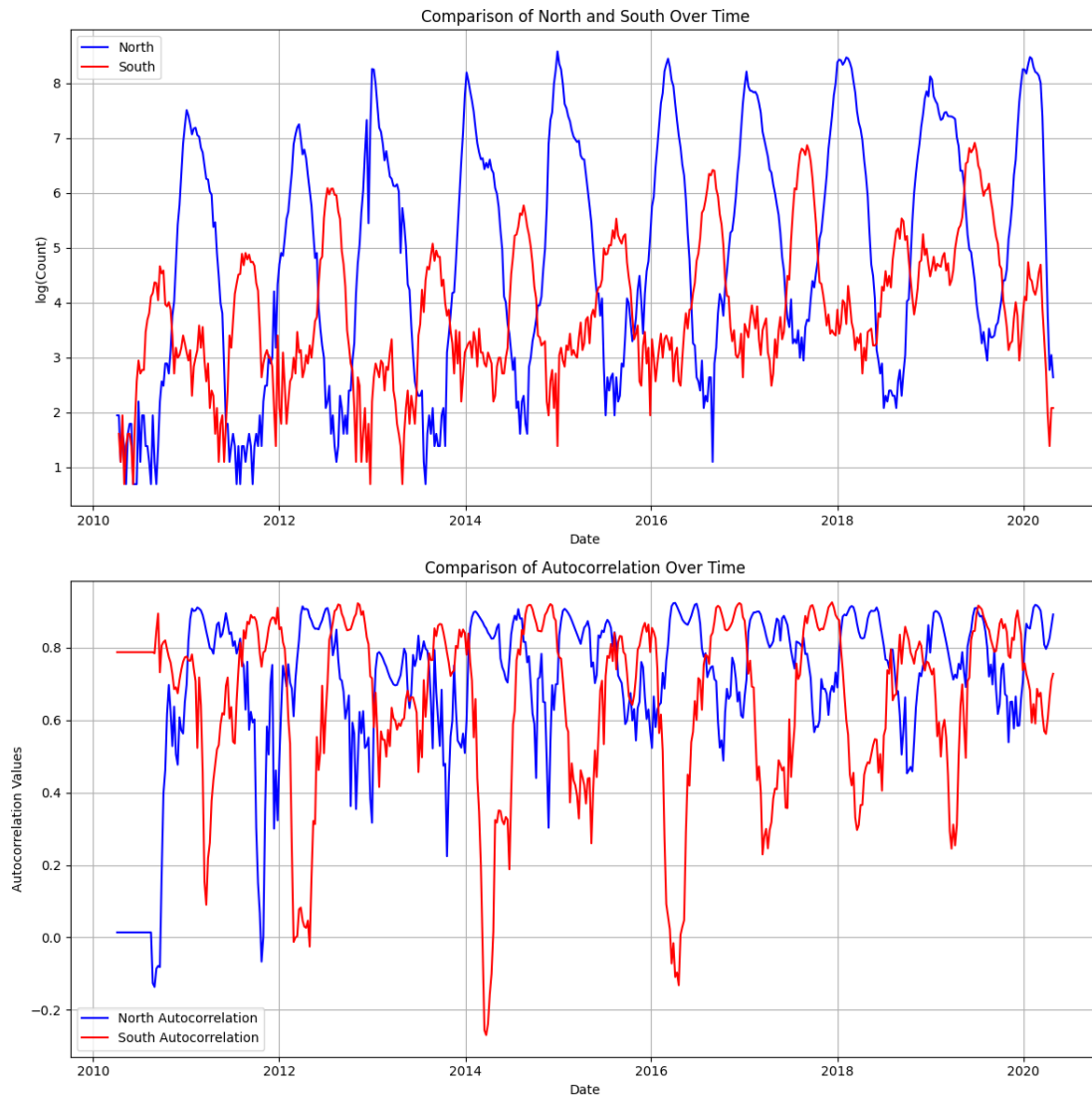


Figure 6.2: WHO FluNet Hemisphere Data Sets [111]: The top panel shows the natural logarithm of the actual counts of influenza viruses detected by hemisphere (data from the Global Influenza Programme). The bottom panel displays the dynamic rolling autocorrelation values of the data with a window size of 20.

For each season, we provide three separate plots, both with and without the Southern Hemisphere covariate, in the model. To illustrate the results, we show the results from the Linear

Regression model, since it is consistently one of the best performing models. First, we show the season prediction, including the peak prediction. Then, we focus on the last 10% of the training data along with the test set to show a clearer view of the predictions and confidence intervals. Lastly, we isolate the test set and show both the actual peak and predicted peak for that season, along with confidence intervals for the predictions. We show the plots for the 2019-2020 season. The plots for the other seasons analyzed are provided in the Appendix.

Prediction for 2019-2020 Season by Linear Regression

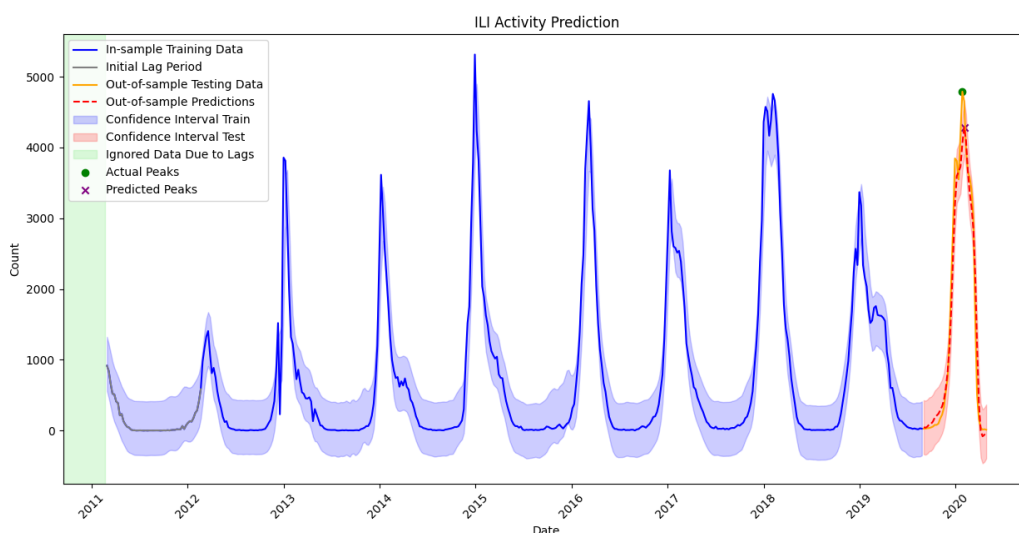


Figure 6.3: Prediction for 2019-2020 Season by Linear Regression without Southern Hemisphere covariate.

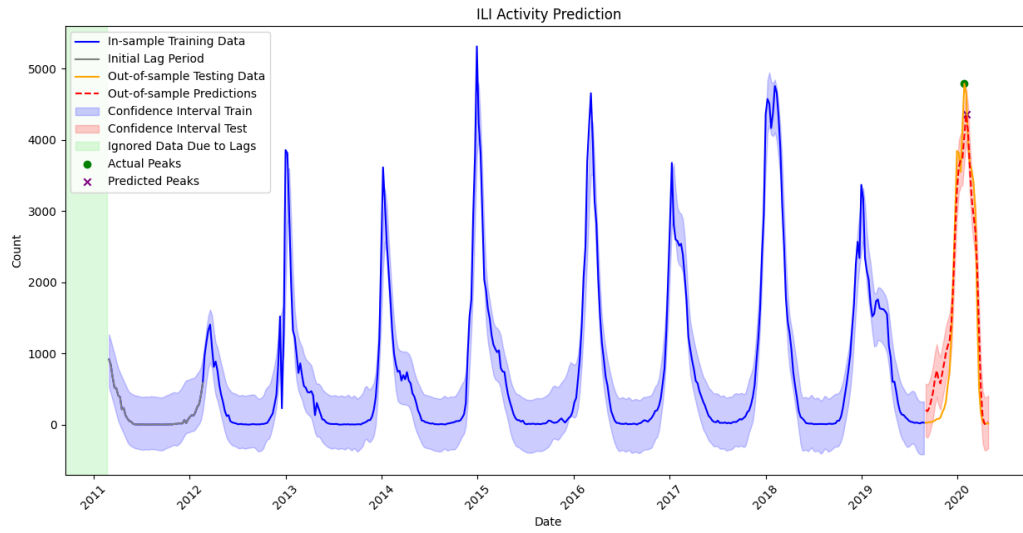


Figure 6.4: Prediction for 2019-2020 Season by Linear Regression including Southern Hemisphere covariate.

2019-2020 Season (Last 10% Training Focus)

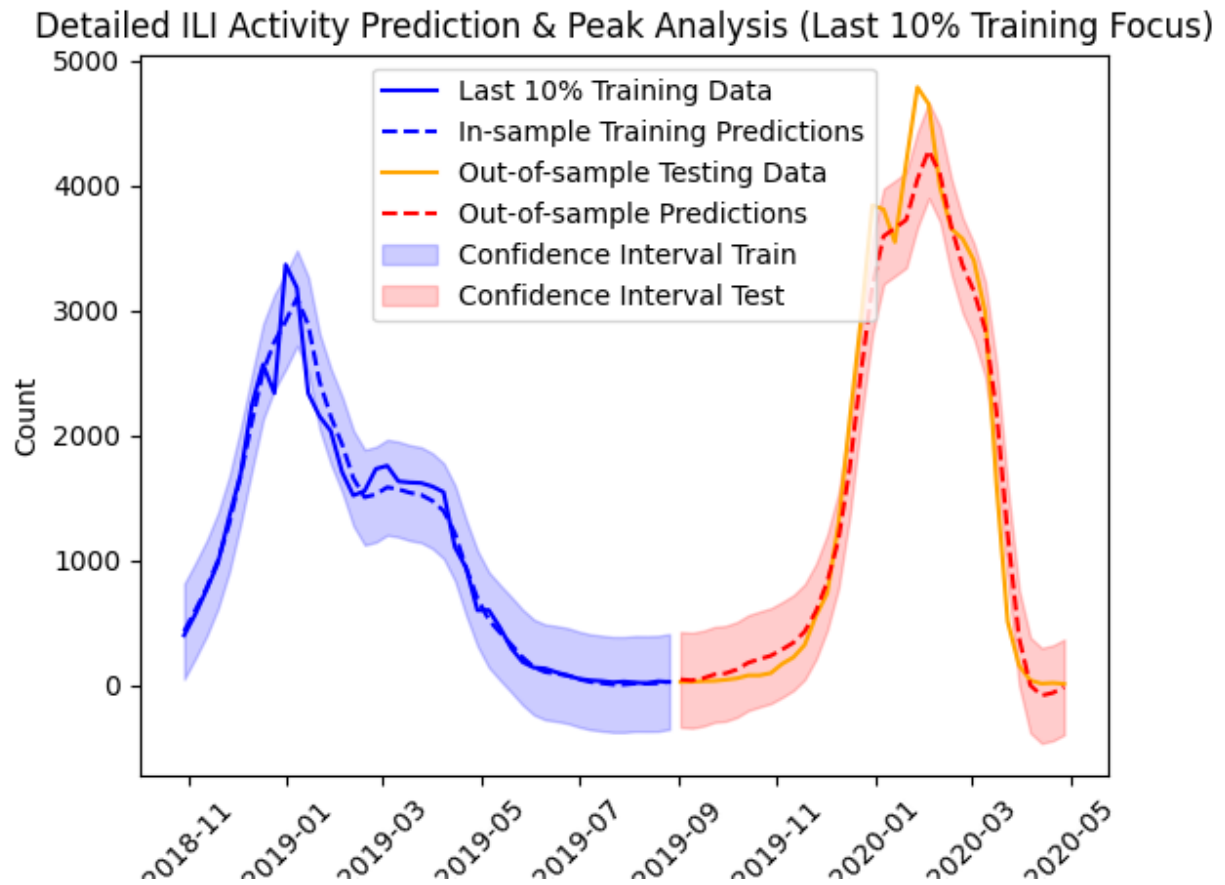


Figure 6.5: 2019-2020 Season Prediction Focused on Last 10% of Training Data without Southern Hemisphere covariate.

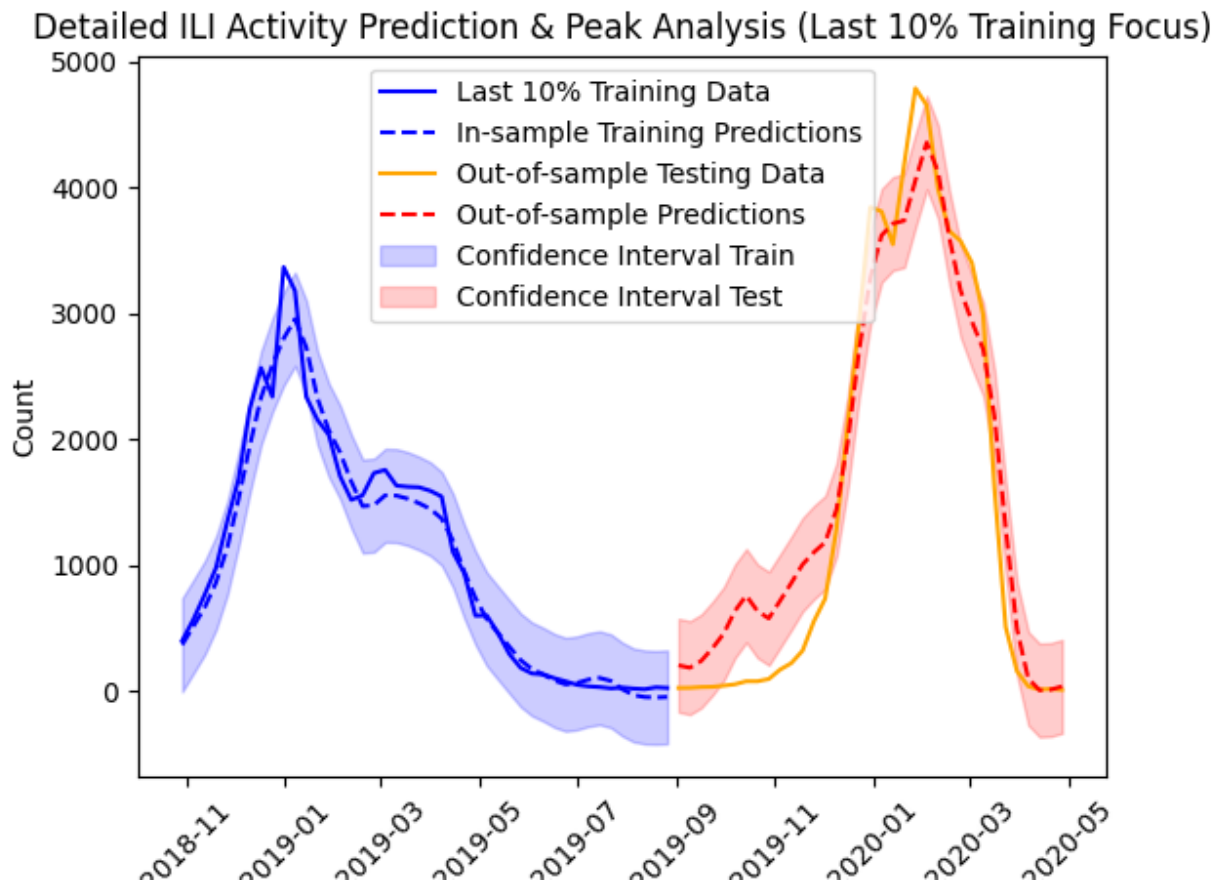


Figure 6.6: 2019-2020 Season Prediction Focused on Last 10% of Training Data including Southern Hemisphere covariate.

2019-2020 Season (Test Set Focus)

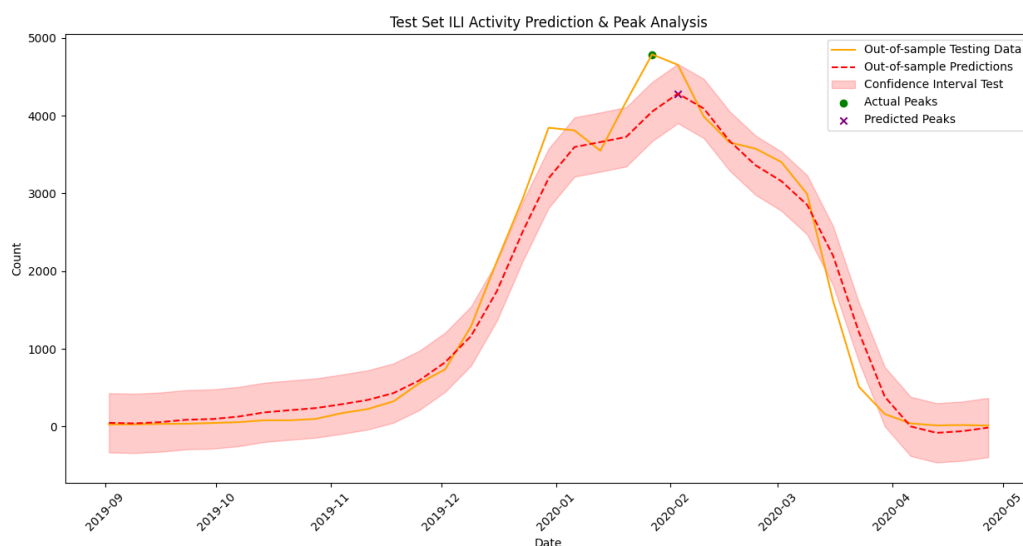


Figure 6.7: 2019-2020 Season Prediction on Test Set without Southern Hemisphere covariate.

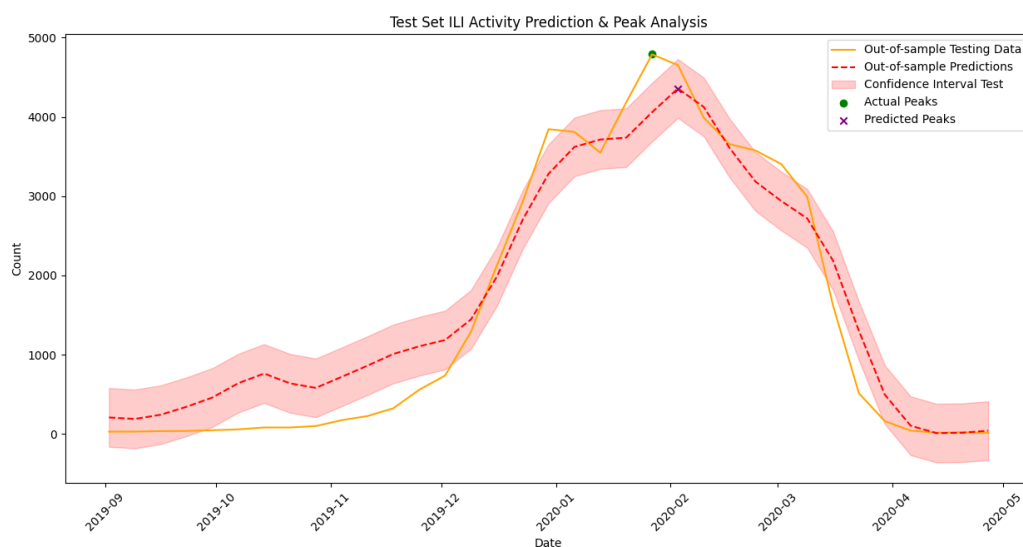


Figure 6.8: 2019-2020 Season Prediction on Test Set including Southern Hemisphere covariate.

Summary of Results

The *Model Performance Results by Season* present the performance metrics of various models for each season, focusing on key metrics such as Peak Date, Peak Height, Accuracy, and RMSE. The aim is to understand how well each model predicts peak values and their associated errors across different seasons.

The *Performance Ranked by Prediction Error by Season* organizes models based on their prediction errors for each season. This ranking helps to identify which models are the most accurate in terms of minimizing prediction errors over different seasons.

The *Mean Percentage Error Across 2016-2020 Seasons* provides the average percentage error of models across multiple seasons (2016-2020). The objective is to evaluate the overall performance consistency of the models over a longer time period.

Model Performance Results for 2013-2014 Season

Model Type	Peak Date	Peak Height	Accuracy Train (%)	Accuracy Test (%)	RMSE Train	RMSE Test	Correlation Train	Correlation Test	CI Lower at Peak	CI Upper at Peak
Linear regression	2014-01-13	3040.78	68.46	69.23	186.56	443.73	0.961	0.881	2282.68	3013.97
Lasso	2014-01-20	2530.51	70.77	69.23	267.71	356.64	0.918	0.939	1468.81	2518.25
SGD regressor	2014-01-20	2614.86	68.46	69.23	292.63	500.87	0.891	0.837	1456.32	2598.58
K-neighbours	2014-01-20	1401.76	69.23	66.67	370.46	774.59	0.862	0.7	-163.19	1273.3
Decision tree	2014-01-13	1322.75	77.69	76.92	136.99	570.71	0.978	0.888	1053.29	1590.31
Random forest	2014-01-20	1798.24	75.38	66.67	188.08	455.78	0.961	0.935	1257.51	1994.43
Adaboost	2014-01-20	1507.08	63.08	71.79	177.13	585.56	0.974	0.874	901.65	1534.45
XGBoost	2014-01-13	1293.28	73.08	76.92	141.54	592.76	0.978	0.894	982.72	1537.47
Simple RNN	2014-01-13	2914.75	62.31	66.67	229.6	306.33	0.944	0.942	2029.32	2922.52
Temporal CNN	2014-01-20	2430.32	69.23	64.1	198.26	668.29	0.953	0.84	1284.83	2054.39
Deep LSTM	2014-01-13	1473.59	59.23	69.23	570.29	735.66	0.558	0.637	348.45	2570.06
Simple LSTM	2014-01-20	2668.21	71.54	76.92	309.15	381.0	0.891	0.916	1329.39	2496.9
Actual Peak	2014-01-06	3614.0	—	—	—	—	—	—	—	—

Table 6.1: Model Performance Results for 2013-2014 Season including Southern Hemisphere covariate.

Model Type	Peak Date	Peak Height	Accuracy Train (%)	Accuracy Test (%)	RMSE Train	RMSE Test	Correlation Train	Correlation Test	CI Lower at Peak	CI Upper at Peak
Linear regression	2014-01-13	2045.21	70.77	58.97	149.44	565.23	0.974	0.834	1499.14	2084.96
Lasso	2014-01-20	1766.42	70.0	66.67	253.04	541.18	0.921	0.905	864.44	1856.36
SGD regressor	2014-01-20	1243.19	68.46	64.1	269.75	810.25	0.91	0.738	302.34	1357.63
K-neighbours	2014-02-03	877.02	70.0	71.79	317.77	935.28	0.885	0.52	-421.89	819.95
Decision tree	2014-01-27	877.7	77.69	56.41	136.99	780.74	0.978	0.746	373.35	910.37
Random forest	2014-01-20	1153.77	78.46	76.92	184.32	665.88	0.96	0.818	645.02	1367.57
Adaboost	2014-02-24	938.06	73.08	58.97	171.87	773.19	0.974	0.706	402.15	1030.52
XGBoost	2014-01-27	1055.03	72.31	71.79	141.97	732.7	0.977	0.791	502.26	1058.69
Simple RNN	2014-01-20	2829.51	66.15	61.54	456.1	482.72	0.881	0.888	1402.28	2592.92
Temporal CNN	2014-02-17	929.88	69.23	56.41	239.77	959.94	0.936	0.269	-128.02	754.28
Deep LSTM	2014-02-10	641.99	70.77	69.23	345.47	908.89	0.843	0.705	-258.8	1087.62
Simple LSTM	2014-01-20	2734.4	71.54	71.79	302.14	384.17	0.889	0.912	1323.07	2485.67
Actual Peak	2014-01-06	3614.0	—	—	—	—	—	—	—	—

Table 6.2: Model Performance Results for 2013-2014 Season excluding Southern Hemisphere covariate.

Model Performance Results for 2014-2015 Season

Model Type	Peak Date	Peak Height	Accuracy Train (%)	Accuracy Test (%)	RMSE Train	RMSE Test	Correlation Train	Correlation Test	CI Lower at Peak	CI Upper at Peak
Linear regression	2015-01-05	4321.11	67.21	87.18	203.7	359.73	0.962	0.968	3545.14	4343.65
Lasso	2015-01-05	3965.7	73.77	92.31	267.44	450.79	0.935	0.95	2917.95	3966.3
SGD regressor	2015-01-12	4032.65	68.85	82.05	292.73	581.93	0.916	0.913	2766.24	3910.79
K-neighbours	2015-01-05	2064.34	70.49	84.62	375.32	1037.47	0.887	0.89	1027.69	2475.82
Decision tree	2014-12-22	2405.66	77.6	84.62	127.21	964.13	0.985	0.89	1596.8	2095.48
Random forest	2014-12-22	2534.99	77.6	92.31	171.27	696.62	0.974	0.955	2084.31	2755.11
Adaboost	2014-12-29	2141.34	69.95	92.31	186.44	780.68	0.981	0.949	1823.03	2459.65
XGBoost	2014-12-15	2838.87	72.68	84.62	130.89	870.35	0.984	0.871	1684.84	2197.88
Simple RNN	2015-01-05	4312.39	66.12	89.74	196.34	347.07	0.968	0.976	3594.37	4342.1
Temporal CNN	2015-01-12	3642.45	70.49	71.79	152.31	660.65	0.978	0.904	2681.37	3268.92
Deep LSTM	2014-12-29	3930.52	71.58	94.87	207.98	484.95	0.956	0.981	3524.23	4336.81
Simple LSTM	2015-01-05	3903.67	74.32	92.31	216.71	341.93	0.954	0.99	3376.34	4222.04
Actual Peak	2014-12-29	5313.0	—	—	—	—	—	—	—	—

Table 6.3: Model Performance Results for 2014-2015 Season including Southern Hemisphere covariate.

Model Type	Peak Date	Peak Height	Accuracy Train (%)	Accuracy Test (%)	RMSE Train	RMSE Test	Correlation Train	Correlation Test	CI Lower at Peak	CI Upper at Peak
Linear regression	2015-01-05	4593.5	71.04	79.49	155.53	290.84	0.977	0.977	3982.78	4592.46
Lasso	2015-01-05	3827.11	73.77	84.62	254.46	492.28	0.939	0.94	2781.32	3778.79
SGD regressor	2015-01-12	4040.68	67.21	79.49	266.09	609.94	0.927	0.898	2771.25	3812.07
K-neighbours	2015-01-19	1897.65	69.95	82.05	367.76	1048.5	0.893	0.873	880.12	2298.86
Decision tree	2014-12-22	2902.41	77.6	76.92	127.21	928.8	0.985	0.83	1788.59	2287.27
Random forest	2014-12-22	2542.34	79.23	92.31	174.39	677.27	0.974	0.957	2119.07	2802.08
Adaboost	2014-12-22	2482.05	73.22	92.31	185.07	692.48	0.984	0.963	2143.21	2762.93
XGBoost	2014-12-22	2040.83	74.32	89.74	131.56	937.02	0.984	0.917	1486.19	2001.85
Simple RNN	2015-01-05	4555.24	63.93	82.05	170.86	331.17	0.977	0.97	3829.85	4452.12
Temporal CNN	2015-01-12	4256.45	73.77	69.23	138.23	637.16	0.982	0.902	3160.94	3700.46
Deep LSTM	2015-01-19	2215.66	69.4	82.05	426.71	955.8	0.8	0.857	794.16	2458.14
Simple LSTM	2015-01-05	3658.21	72.13	92.31	222.38	389.92	0.955	0.989	3083.59	3944.7
Actual Peak	2014-12-29	5313.0	—	—	—	—	—	—	—	—

Table 6.4: Model Performance Results for 2014-2015 Season excluding Southern Hemisphere covariate.

Model Performance Results for 2015-2016 Season

Model Type	Peak Date	Peak Height	Accuracy Train (%)	Accuracy Test (%)	RMSE Train	RMSE Test	Correlation Train	Correlation Test	CI Lower at Peak	CI Upper at Peak
Linear regression	2016-03-07	3774.01	70.64	74.36	218.74	400.97	0.971	0.97	3345.29	4202.74
Lasso	2016-03-14	3641.6	75.74	74.36	275.89	451.04	0.954	0.96	2992.97	4074.46
SGD regressor	2016-03-14	3490.98	75.74	69.23	328.78	755.31	0.927	0.851	2668.67	3956.31
K-neighbours	2016-03-14	2177.37	72.34	66.67	406.5	1025.71	0.923	0.679	1386.71	2945.74
Decision tree	2016-02-15	2263.39	78.72	74.36	129.96	786.89	0.99	0.891	1545.47	2054.91
Random forest	2016-02-29	2141.97	80.0	79.49	174.76	708.15	0.982	0.946	1777.66	2462.24
Adaboost	2016-03-07	2074.63	68.51	74.36	220.43	834.97	0.987	0.905	1724.97	2424.29
XGBoost	2016-03-07	1849.18	77.45	89.74	137.75	837.29	0.989	0.931	1579.26	2119.1
Simple RNN	2016-03-07	3777.6	68.09	71.79	219.15	380.89	0.971	0.973	3348.19	4207.01
Temporal CNN	2015-12-14	2844.9	71.91	64.1	156.62	1251.22	0.985	0.484	2232.72	2846.38
Deep LSTM	2016-03-21	1721.62	73.62	79.49	233.03	911.51	0.968	0.878	1190.88	2095.27
Simple LSTM	2016-03-07	3318.96	74.89	84.62	222.4	419.78	0.97	0.988	2883.48	3754.43
Actual Peak	2016-03-07	4655.0	—	—	—	—	—	—	—	—

Table 6.5: Model Performance Results for 2015-2016 Season including Southern Hemisphere covariate.

Model Type	Peak Date	Peak Height	Accuracy Train (%)	Accuracy Test (%)	RMSE Train	RMSE Test	Correlation Train	Correlation Test	CI Lower at Peak	CI Upper at Peak
Linear regression	2016-03-14	3920.94	70.21	69.23	185.92	459.58	0.979	0.957	3280.01	4008.81
Lasso	2016-03-14	3348.77	75.74	71.79	267.42	534.48	0.956	0.949	2727.73	3776.03
SGD regressor	2016-03-14	2521.58	66.81	61.54	332.66	916.18	0.943	0.762	1804.85	3107.84
K-neighbours	2015-12-21	1671.57	73.19	69.23	393.13	1273.9	0.91	0.435	436.07	1960.44
Decision tree	2016-02-08	1967.3	78.72	69.23	129.96	926.47	0.99	0.811	1275.48	1784.92
Random forest	2016-02-22	2101.18	80.85	79.49	184.32	710.94	0.98	0.924	1723.96	2446.23
Adaboost	2016-03-07	2195.44	75.74	74.36	206.02	783.84	0.987	0.914	1858.9	2531.98
XGBoost	2016-02-15	1704.15	77.45	74.36	137.38	895.86	0.989	0.886	1384.42	1922.83
Simple RNN	2016-03-14	3557.19	70.21	69.23	229.36	655.4	0.976	0.899	2785.2	3562.53
Temporal CNN	2015-12-21	2871.32	72.34	69.23	149.26	1286.97	0.987	0.438	1970.01	2546.15
Deep LSTM	2016-01-04	2719.23	64.68	61.54	649.78	1541.12	0.701	0.053	-219.56	2257.91
Simple LSTM	2016-03-07	2995.06	72.77	79.49	225.73	512.62	0.971	0.983	2552.63	3437.49
Actual Peak	2016-03-07	4655.0	—	—	—	—	—	—	—	—

Table 6.6: Model Performance Results for 2015-2016 Season excluding Southern Hemisphere covariate.

Model Performance Results for 2016-2017 Season

Model Type	Peak Date	Peak Height	Accuracy Train (%)	Accuracy Test (%)	RMSE Train	RMSE Test	Correlation Train	Correlation Test	CI Lower at Peak	CI Upper at Peak
Linear regression	2017-01-16	3049.07	70.73	92.31	211.66	217.54	0.978	0.979	2418.15	3247.85
Lasso	2017-01-16	2940.34	76.31	89.74	274.68	284.02	0.962	0.963	2066.54	3143.28
SGD regressor	2017-01-23	2705.07	71.78	84.62	318.95	374.02	0.948	0.942	1642.46	2892.2
K-neighbours	2017-02-13	2576.9	73.17	74.36	409.59	862.61	0.915	0.696	-334.0	1264.99
Decision tree	2017-01-23	4016.54	80.84	79.49	123.22	579.45	0.992	0.891	2615.91	3098.93
Random forest	2017-01-16	3243.44	81.53	92.31	175.73	293.06	0.985	0.973	2672.95	3361.42
Adaboost	2017-01-16	3366.59	71.08	87.18	239.71	398.79	0.986	0.966	2805.33	3560.1
XGBoost	2017-01-23	3459.23	79.44	87.18	130.28	293.53	0.992	0.969	2771.72	3282.31
Simple RNN	2017-01-16	3041.29	70.73	89.74	212.87	226.45	0.978	0.977	2405.88	3237.88
Temporal CNN	2017-01-16	2928.5	73.17	64.1	169.25	353.51	0.987	0.941	2266.13	2929.36
Deep LSTM	2017-01-23	3586.63	75.26	87.18	244.34	396.41	0.969	0.928	2277.7	3210.78
Simple LSTM	2017-01-16	3268.58	75.96	89.74	216.82	173.5	0.976	0.987	2680.06	3512.67
Actual Peak	2017-01-09	3677.0	—	—	—	—	—	—	—	—

Table 6.7: Model Performance Results for 2016-2017 Season including Southern Hemisphere covariate.

Model Type	Peak Date	Peak Height	Accuracy Train (%)	Accuracy Test (%)	RMSE Train	RMSE Test	Correlation Train	Correlation Test	CI Lower at Peak	CI Upper at Peak
Linear regression	2017-01-16	3441.69	71.43	76.92	195.28	220.02	0.981	0.983	2847.59	3613.09
Lasso	2017-01-16	3248.22	77.0	89.74	271.35	213.08	0.963	0.98	2416.13	3479.82
SGD regressor	2017-01-16	3123.41	70.03	76.92	328.42	282.28	0.952	0.963	2058.98	3345.95
K-neighbours	2017-01-30	3201.08	73.52	76.92	405.16	428.92	0.92	0.924	1445.0	3027.42
Decision tree	2017-01-23	3782.5	80.84	82.05	123.22	525.14	0.992	0.901	2509.22	2992.24
Random forest	2017-01-16	3294.98	81.53	89.74	182.7	277.52	0.984	0.979	2737.61	3453.58
Adaboost	2017-01-16	3613.61	73.17	87.18	238.29	395.4	0.987	0.972	2948.44	3700.43
XGBoost	2017-01-16	3676.34	79.44	87.18	129.59	311.13	0.992	0.98	3213.73	3721.64
Simple RNN	2017-01-16	3360.64	71.78	76.92	194.89	169.53	0.981	0.987	2813.18	3575.22
Temporal CNN	2017-01-16	3965.0	72.47	79.49	182.84	913.29	0.988	0.898	3272.6	3860.87
Deep LSTM	2017-01-23	3331.21	75.26	87.18	186.64	545.57	0.982	0.879	1740.53	2469.34
Simple LSTM	2017-01-16	3426.28	74.91	87.18	214.37	172.31	0.978	0.987	2871.61	3711.93
Actual Peak	2017-01-09	3677.0	—	—	—	—	—	—	—	—

Table 6.8: Model Performance Results for 2016-2017 Season excluding Southern Hemisphere covariate.

Model Performance Results for 2017-2018 Season

Model Type	Peak Date	Peak Height	Accuracy Train (%)	Accuracy Test (%)	RMSE Train	RMSE Test	Correlation Train	Correlation Test	CI Lower at Peak	CI Upper at Peak
Linear regression	2018-02-12	4198.55	72.86	87.18	204.18	248.26	0.98	0.995	3721.78	4522.18
Lasso	2018-02-12	4191.84	77.29	87.18	214.42	258.12	0.978	0.995	3683.06	4523.59
SGD regressor	2018-01-22	4052.14	72.27	84.62	308.65	418.65	0.954	0.989	3197.9	4407.49
K-neighbours	2018-01-15	3400.7	73.75	87.18	423.56	984.1	0.908	0.964	1839.89	3493.32
Decision tree	2018-01-08	4353.17	81.71	87.18	118.8	328.74	0.993	0.984	3597.97	4063.65
Random forest	2018-01-08	3652.35	81.71	89.74	166.47	530.54	0.987	0.981	2890.76	3543.08
Adaboost	2018-01-08	3772.38	71.09	84.62	263.85	501.99	0.984	0.984	2963.76	3752.86
XGBoost	2018-01-08	3821.0	80.24	87.18	125.46	501.99	0.993	0.978	3003.3	3494.99
Simple RNN	2018-01-15	4403.57	70.5	87.18	203.46	208.93	0.979	0.996	3802.97	4593.25
Temporal CNN	2018-01-15	4042.11	74.04	82.05	176.08	488.77	0.985	0.985	2959.79	3642.35
Deep LSTM	2018-01-15	3699.25	75.52	87.18	195.52	515.71	0.98	0.988	3033.14	3798.79
Simple LSTM	2018-01-15	4002.41	78.17	84.62	200.52	426.61	0.979	0.993	3101.21	3886.62
Actual Peak	2018-02-05	4757.0	–	–	–	–	–	–	–	–

Table 6.9: Model Performance Results for 2017-2018 Season including Southern Hemisphere covariate.

Model Type	Peak Date	Peak Height	Accuracy Train (%)	Accuracy Test (%)	RMSE Train	RMSE Test	Correlation Train	Correlation Test	CI Lower at Peak	CI Upper at Peak
Linear regression	2018-01-15	4783.89	71.68	82.05	195.04	217.76	0.981	0.992	4059.19	4823.73
Lasso	2018-01-15	4551.98	78.17	87.18	262.77	218.34	0.966	0.993	3778.4	4808.46
SGD regressor	2018-01-15	4703.23	70.21	76.92	312.74	301.65	0.952	0.987	3550.09	4775.89
K-neighbours	2018-01-08	3331.14	74.34	89.74	380.95	950.73	0.928	0.935	1714.64	3200.63
Decision tree	2018-01-08	3286.33	81.71	87.18	118.8	646.67	0.993	0.976	2761.66	3227.34
Random forest	2018-01-08	3614.85	81.71	89.74	170.01	553.83	0.986	0.979	2848.95	3515.14
Adaboost	2018-01-15	3702.44	75.52	87.18	262.23	477.45	0.985	0.985	3185.08	3978.7
XGBoost	2018-01-15	3594.5	79.94	87.18	124.64	523.94	0.993	0.984	3038.48	3527.0
Simple RNN	2018-01-15	4290.65	69.91	84.62	212.58	285.5	0.978	0.994	3478.05	4309.98
Temporal CNN	2018-01-08	4948.66	73.16	82.05	157.32	670.09	0.987	0.927	3029.9	3632.75
Deep LSTM	2018-01-01	3288.88	75.81	89.74	169.94	822.49	0.985	0.981	2487.32	3149.15
Simple LSTM	2018-01-15	3982.99	76.11	76.92	212.33	420.38	0.978	0.993	3137.2	3965.34
Actual Peak	2018-02-05	4757.0	–	–	–	–	–	–	–	–

Table 6.10: Model Performance Results for 2017-2018 Season excluding Southern Hemisphere covariate.

Model Performance Results for 2018-2019 Season

Model Type	Peak Date	Peak Height	Accuracy Train (%)	Accuracy Test (%)	RMSE Train	RMSE Test	Correlation Train	Correlation Test	CI Lower at Peak	CI Upper at Peak
Linear regression	2019-01-07	3150.35	72.89	87.18	196.95	170.31	0.985	0.984	2572.89	3344.93
Lasso	2019-01-07	3188.0	76.73	87.18	207.51	178.6	0.984	0.983	2567.37	3380.81
SGD regressor	2019-01-07	3256.54	72.63	79.49	320.49	268.69	0.96	0.968	2457.57	3713.82
K-neighbours	2019-01-21	3218.89	73.91	76.92	416.7	581.21	0.931	0.843	1861.1	3490.62
Decision tree	2018-12-17	3050.09	81.84	74.36	115.31	578.36	0.995	0.85	2148.07	2600.07
Random forest	2018-12-24	3063.46	82.1	82.05	161.55	232.71	0.99	0.972	2744.44	3377.56
Adaboost	2019-01-07	2878.17	73.4	82.05	270.54	227.15	0.988	0.976	2400.2	3213.49
XGBoost	2018-12-31	2767.98	81.07	89.74	120.81	211.86	0.995	0.973	2531.25	3004.72
Simple RNN	2019-01-07	3002.22	73.91	87.18	208.52	171.9	0.985	0.984	2405.68	3211.04
Temporal CNN	2019-01-07	3337.56	74.94	76.92	193.11	251.55	0.988	0.968	2708.0	3442.82
Deep LSTM	2018-12-17	2716.81	77.24	82.05	254.32	285.22	0.98	0.962	2167.76	3134.49
Simple LSTM	2019-01-07	3005.01	76.47	84.62	196.67	161.33	0.985	0.985	2393.9	3164.84
Actual Peak	2018-12-31	3369.0	–	–	–	–	–	–	–	–

Table 6.11: Model Performance Results for 2018-2019 Season including Southern Hemisphere covariate.

Model Type	Peak Date	Peak Height	Accuracy Train (%)	Accuracy Test (%)	RMSE Train	RMSE Test	Correlation Train	Correlation Test	CI Lower at Peak	CI Upper at Peak
Linear regression	2019-01-07	2888.8	72.12	79.49	190.67	206.73	0.986	0.98	2328.14	3075.55
Lasso	2019-01-07	2887.4	78.77	82.05	253.65	245.06	0.975	0.965	2141.45	3135.75
SGD regressor	2019-01-14	2711.0	73.66	76.92	304.19	310.18	0.963	0.955	1901.82	3094.03
K-neighbours	2019-01-28	2733.41	75.19	71.79	381.78	840.11	0.949	0.69	597.57	2082.94
Decision tree	2019-03-04	3051.84	81.84	79.49	115.31	537.57	0.995	0.843	1884.73	2336.73
Random forest	2019-01-07	3160.44	81.33	84.62	163.14	251.97	0.99	0.975	2774.06	3413.5
Adaboost	2019-01-07	3385.3	74.94	82.05	276.22	261.09	0.987	0.972	2760.62	3614.16
XGBoost	2019-01-07	2797.97	80.05	87.18	120.67	169.97	0.995	0.983	2446.24	2919.21
Simple RNN	2019-01-07	2902.95	73.4	87.18	204.12	204.17	0.984	0.976	2322.78	3105.34
Temporal CNN	2019-01-14	2461.77	75.45	64.1	198.88	368.33	0.988	0.928	1775.34	2495.64
Deep LSTM	2019-01-07	2706.71	77.24	82.05	206.57	243.03	0.983	0.971	2187.32	2988.21
Simple LSTM	2019-01-07	2889.45	74.68	82.05	212.51	184.95	0.984	0.982	2288.44	3104.49
Actual Peak	2018-12-31	3369.0	–	–	–	–	–	–	–	–

Table 6.12: Model Performance Results for 2018-2019 Season excluding Southern Hemisphere covariate.

Model Performance Results for 2019-2020 Season

Model Type	Peak Date	Peak Height	Accuracy Train (%)	Accuracy Test (%)	RMSE Train	RMSE Test	Correlation Train	Correlation Test	CI Lower at Peak	CI Upper at Peak
Linear regression	2020-02-03	4284.11	75.23	82.86	194.5	285.27	0.985	0.989	3670.37	4432.8
Lasso	2020-02-03	4277.05	77.7	85.71	204.71	295.42	0.984	0.989	3632.05	4434.51
SGD regressor	2020-02-03	3890.41	74.32	82.86	305.6	455.84	0.964	0.975	3140.48	4338.1
K-neighbours	2020-01-13	3799.68	75.23	82.86	393.69	535.52	0.939	0.969	2886.27	4424.9
Decision tree	2020-02-24	4542.32	82.43	71.43	113.24	502.87	0.995	0.967	3944.79	4388.68
Random forest	2020-02-03	4302.86	83.11	82.86	157.87	340.55	0.99	0.981	3894.33	4513.06
Adaboost	2020-02-03	4279.89	75.0	82.86	272.86	483.56	0.986	0.971	3666.47	4504.92
XGBoost	2020-02-10	4364.55	80.63	77.14	118.96	418.48	0.995	0.971	3990.91	4457.17
Simple RNN	2020-02-03	4240.76	75.45	82.86	197.86	290.45	0.985	0.989	3644.3	4417.81
Temporal CNN	2020-02-03	4035.04	73.2	74.29	196.9	529.98	0.985	0.955	3360.63	4125.94
Deep LSTM	2020-02-03	3945.84	77.03	85.71	249.86	356.19	0.982	0.987	3175.17	4105.11
Simple LSTM	2020-02-03	4118.54	77.7	82.86	199.16	318.07	0.984	0.987	3530.5	4310.99
Actual Peak	2020-01-27	4789.0	—	—	—	—	—	—	—	—

Table 6.13: Model Performance Results for 2019-2020 Season including Southern Hemisphere covariate.

Model Type	Peak Date	Peak Height	Accuracy Train (%)	Accuracy Test (%)	RMSE Train	RMSE Test	Correlation Train	Correlation Test	CI Lower at Peak	CI Upper at Peak
Linear regression	2020-02-03	4358.09	73.2	80.0	189.04	420.9	0.986	0.984	3687.78	4428.82
Lasso	2020-02-03	4161.96	79.05	82.86	204.03	422.42	0.983	0.984	3554.38	4354.17
SGD regressor	2020-02-03	3810.14	75.68	82.86	297.01	732.08	0.963	0.945	3126.76	4290.93
K-neighbours	2019-11-25	3409.58	75.0	65.71	379.74	1891.34	0.949	0.119	498.5	1969.35
Decision tree	2020-01-06	2523.85	82.43	65.71	113.24	855.0	0.995	0.951	2291.05	2734.95
Random forest	2020-02-03	4039.34	82.21	82.86	158.8	348.72	0.99	0.981	3711.89	4334.34
Adaboost	2020-01-13	3955.06	75.9	80.0	252.01	520.3	0.986	0.974	3377.15	4176.54
XGBoost	2020-01-27	3321.5	80.63	85.71	118.33	578.34 bias refer	0.995	0.975	3089.6	3553.39
Simple RNN	2020-02-03	3839.86	72.07	80.0	201.35	487.93	0.99	0.978	3291.0	4055.19
Temporal CNN	2020-02-10	3797.16	76.35	74.29	158.05	706.79	0.99	0.931	3156.46	3774.71
Deep LSTM	2020-02-03	3763.71	76.13	85.71	204.13	435.72	0.986	0.987	3250.21	4037.19
Simple LSTM	2020-02-03	3976.49	75.23	82.86	196.33	400.43	0.985	0.98	3461.42	4229.27
Actual Peak	2020-01-27	4789.0	—	—	—	—	—	—	—	—

Table 6.14: Model Performance Results for 2019-2020 Season excluding Southern Hemisphere covariate.

Performance Ranked by Prediction Error – 2013-14

Model Type	Error (%)	CI Lower Error (%)	CI Upper Error (%)
Simple RNN	21.71	61.2	28.25
Simple LSTM	24.34	63.39	31.22
Linear regression	43.41	58.52	42.31
Lasso	51.12	76.08	48.63
SGD regressor	65.6	91.63	62.43

Table 6.15: Performance Ranked by Prediction Error – 2013-14 including Southern Hemisphere covariate.

Model Type	Error (%)	CI Lower Error (%)	CI Upper Error (%)
Linear regression	15.86	36.84	16.6
Simple RNN	19.35	43.85	19.13
Simple LSTM	26.17	63.22	30.91
SGD regressor	27.65	59.7	28.1
Lasso	29.98	59.36	30.32

Table 6.16: Performance Ranked by Prediction Error – 2013-14 excluding Southern Hemisphere covariate.

Performance Ranked by Prediction Error – 2014-15

Model Type	Error (%)	CI Lower Error (%)	CI Upper Error (%)
Linear regression	13.54	25.04	13.56
Simple RNN	14.26	27.92	16.2
Temporal CNN	19.89	40.51	30.35
SGD regressor	23.95	47.84	28.25
Lasso	27.97	47.65	28.88

Table 6.17: Performance Ranked by Prediction Error – 2014-15 including Southern Hemisphere covariate.

Model Type	Error (%)	CI Lower Error (%)	CI Upper Error (%)
Linear regression	18.67	33.27	18.24
Simple RNN	18.83	32.35	18.27
SGD regressor	24.1	47.93	26.39
Lasso	25.36	45.08	25.35
Deep LSTM	26.02	33.67	18.37

Table 6.18: Performance Ranked by Prediction Error – 2014-15 excluding Southern Hemisphere covariate.

Performance Ranked by Prediction Error – 2015-16

Model Type	Error (%)	CI Lower Error (%)	CI Upper Error (%)
Linear regression	15.77	29.54	13.88
Simple RNN	23.58	40.17	23.47
Lasso	28.06	41.4	18.88
Simple LSTM	35.66	45.16	26.15
Temporal CNN	38.32	57.68	45.3

Table 6.19: Performance Ranked by Prediction Error – 2015-16 including Southern Hemisphere covariate.

Model Type	Error (%)	CI Lower Error (%)	CI Upper Error (%)
Simple RNN	18.85	28.07	9.62
Linear regression	18.93	28.14	9.72
Lasso	21.77	35.7	12.47
SGD regressor	25.01	42.67	15.01
Simple LSTM	28.7	38.06	19.35

Table 6.20: Performance Ranked by Prediction Error – 2015-16 excluding Southern Hemisphere covariate.

Performance Ranked by Prediction Error – 2016-17

Model Type	Error (%)	CI Lower Error (%)	CI Upper Error (%)
XGBoost	0.02	12.6	1.21
Adaboost	1.72	19.81	0.64
Decision tree	2.87	31.76	18.62
Linear regression	6.4	22.56	1.74
Simple LSTM	6.82	21.9	0.95

Table 6.21: Performance Ranked by Prediction Error – 2016-17 including Southern Hemisphere covariate.

Model Type	Error (%)	CI Lower Error (%)	CI Upper Error (%)
Deep LSTM	2.46	38.06	12.68
XGBoost	5.92	24.62	10.73
Adaboost	8.44	23.71	3.18
Decision tree	9.23	28.86	15.72
Simple LSTM	11.11	27.11	4.47

Table 6.22: Performance Ranked by Prediction Error – 2016-17 excluding Southern Hemisphere covariate.

Performance Ranked by Prediction Error – 2017-18

Model Type	Error (%)	CI Lower Error (%)	CI Upper Error (%)
Linear regression	0.57	14.67	1.4
SGD regressor	1.13	25.37	0.4
Temporal CNN	4.03	36.31	23.63
Lasso	4.31	20.57	1.08
Simple RNN	9.8	26.89	9.4

Table 6.23: Performance Ranked by Prediction Error – 2017-18 including Southern Hemisphere covariate.

Model Type	Error (%)	CI Lower Error (%)	CI Upper Error (%)
Simple RNN	7.43	20.06	3.44
Decision tree	8.49	24.36	14.58
Lasso	10.08	22.58	4.91
Linear regression	10.43	21.76	4.94
SGD regressor	14.82	32.77	7.35

Table 6.24: Performance Ranked by Prediction Error – 2017-18 excluding Southern Hemisphere covariate.

Performance Ranked by Prediction Error – 2018-19

Model Type	Error (%)	CI Lower Error (%)	CI Upper Error (%)
Adaboost	0.48	18.06	7.28
Random forest	6.19	17.66	1.32
Decision tree	9.41	44.06	30.64
Simple RNN	13.83	31.05	7.83
Simple LSTM	14.23	32.07	7.85

Table 6.25: Performance Ranked by Prediction Error – 2018-19 including Southern Hemisphere covariate.

Model Type	Error (%)	CI Lower Error (%)	CI Upper Error (%)
Temporal CNN	0.93	19.62	2.19
SGD regressor	3.34	27.05	10.24
K-neighbours	4.46	44.76	3.61
Lasso	5.37	23.79	0.35
Linear regression	6.49	23.63	0.71

Table 6.26: Performance Ranked by Prediction Error – 2018-19, Southern Hemisphere covariate excluded.

Performance Ranked by Prediction Error – 2019-20

Model Type	Error (%)	CI Lower Error (%)	CI Upper Error (%)
Linear regression	9.0	22.99	7.52
Lasso	13.09	25.78	9.08
Random forest	15.65	22.49	9.49
Simple LSTM	16.97	27.72	11.69
Adaboost	17.41	29.48	12.79

Table 6.27: Performance Ranked by Prediction Error – 2019-20 including Southern Hemisphere covariate.

Model Type	Error (%)	CI Lower Error (%)	CI Upper Error (%)
XGBoost	8.86	16.67	6.93
Random forest	10.15	18.68	5.76
Linear regression	10.54	23.36	7.44
Adaboost	10.63	23.44	5.93
Lasso	10.69	24.16	7.4

Table 6.28: Performance Ranked by Prediction Error – 2019-20 excluding Southern Hemisphere covariate.

From the analysis, several key patterns emerged across the different seasons. Linear regression consistently ranked among the top performers, especially in the earlier seasons, showcasing its reliability as a simple yet effective model. Lasso regression performed similarly to linear regression, with only slightly higher error margins, indicating its utility in feature selection and model simplification. The SGD regressor exhibited more variable performance, often showing higher prediction errors compared to other models, which suggests its limitations in handling the data consistently. Simple RNN and LSTM models demonstrated strong performance, particularly when covariates were included, highlighting their strength in modelling time-dependent data with additional features.

In the 2013-2014 season, Simple RNN and Simple LSTM outperformed other models when covariates were included, while linear regression showed the lowest prediction error when covariates were excluded, indicating robustness even without covariates. The 2014-2015 season saw both Linear regression and Simple RNN maintain their positions as top performers with and without covariates, showcasing their consistency across data configurations. Similarly, in the 2015-2016 season, Linear regression and Simple RNN continued to

demonstrate strong performance, maintaining low prediction errors in both configurations.

The 2016-2017 season introduced a shift, with XGBoost and Adaboost emerging as the best-performing models when covariates were included, while Deep LSTM and XGBoost were the top models when covariates were excluded, showing improvement in Deep LSTM's ability to capture complex patterns. In the 2017-2018 season, Linear regression and SGD regressor performed best with covariates included, while Simple RNN and Decision tree models excelled when covariates were excluded, indicating adaptability to different data setups.

For the 2018-2019 season, Adaboost and Random forest led with covariates included, while Temporal CNN and SGD regressor maintained low errors without covariates. Finally, in the 2019-2020 season, Linear regression and Lasso performed well with covariates, whereas XGBoost and Random forest showed robustness and emerged as top performers without covariates.

In terms of overall model robustness, Random forest and XGBoost consistently performed well across seasons, both with and without covariates, demonstrating their flexibility and adaptability to varying data setups. Simple RNN and LSTM models showed strong performance, particularly when covariates were included, highlighting their ability to handle complex, time-dependent data effectively. Linear regression remained reliable across multiple seasons, maintaining low prediction errors, further solidifying its status as a robust and consistent model for influenza peak forecasting.

The mean percentage error for various models across the 2016-2020 seasons provides valuable insights into model performance. When covariates from the Southern Hemisphere were included, Linear regression demonstrated the lowest mean error at 7.55%, indicating a strong predictive performance. Adaboost and Lasso exhibited moderate mean errors of 10.45% and 10.84%, respectively, but their wide confidence intervals reflect some uncertainty. Simple RNN and Simple LSTM had higher mean errors of 13.02% and 13.57%, suggesting more variability in their predictions.

In contrast, when covariates were excluded, Linear regression's mean error increased to 11.26%, though it remained among the top performers. Lasso and Simple RNN showed similar mean errors of 11.61% and 11.76%, indicating consistency across both configurations. However, Adaboost and Simple LSTM displayed higher mean errors of 12.58% and 12.94%, suggesting that these models rely more heavily on covariate information for accurate predictions.

When covariates were included, Linear regression showed the lowest mean error at 7.55%, affirming its robustness in prediction tasks. Adaboost and Lasso also performed moderately well, with mean errors of 10.45% and 10.84%, though their wide confidence intervals indicate

greater variability. Simple RNN and Simple LSTM, on the other hand, demonstrated higher mean errors of 13.02% and 13.57%, indicating increased variability in their predictions.

Without covariates, Linear regression's mean error increased to 11.26%, although it remained one of the best-performing models. Lasso and Simple RNN had similar mean errors of 11.61% and 11.76%, showcasing consistent performance in both scenarios. However, Adaboost and Simple LSTM experienced higher mean errors without covariates, reflecting their sensitivity to the exclusion of additional information.

The comparison between covariate settings shows that Linear regression's performance degrades slightly when covariates are excluded, but it remains a robust model. Adaboost exhibited a noticeable increase in mean error without covariates, which suggests a greater reliance on supplementary information for accurate predictions. Lasso maintained stable performance across both covariate settings, demonstrating its adaptability. Simple RNN and Simple LSTM, however, displayed higher errors in both configurations, with greater sensitivity to covariate exclusion, indicating that these models may benefit more from the presence of additional data.

The characterization of the peak curve was analyzed using three key metrics: Mean Growth Rate which is the average rate at which cases increased during the period leading up to the peak, mean slope, or the average slope of the case increase, representing the overall trend of rising cases, and mean curvature which measures the acceleration of the rise towards the peak. Higher values indicate a rapid increase in the rate of new cases.

The table below provides a detailed characterization of growth metrics for the 12 weeks leading up to the peak for each season.

Season	Peak Date	Peak Height	Mean Growth Rate	Mean Slope	Mean Curvature
2013-14	Jan 6, 2014	3614	1.07	325.85	94.21
2014-15	Dec 29, 2014	5313	0.58	464.69	128.79
2015-16	Mar 9, 2015	4655	0.40	369.19	34.37
2016-17	Jan 11, 2016	3677	0.40	310.62	64.63
2017-18	Feb 6, 2017	4757	0.25	356.77	22.23
2018-19	Jan 1, 2018	3369	0.44	296.62	100.02
2019-20	Jan 28, 2019	4789	0.34	380.42	43.10

Table 6.29: Growth metrics for the 12 weeks prior to the peak for each season.

The 2013-14 season featured a rapid onset with the fastest mean growth rate of 1.07, a high mean slope of 325.85, and a moderate mean curvature of 94.21, indicating a sharp rise towards the peak. The 2014-15 season was more severe, with the highest peak height and an intense rise characterized by a mean slope of 464.69 and a curvature of 128.79. The

2015-16 season showed a late peak with the lowest mean growth rate (0.40) and a moderate rise towards the peak, indicated by a mean slope of 369.19 and a curvature of 34.37. The 2016-17 season had a steadier progression, with moderate growth and a more gradual slope compared to other seasons.

The 2017-18 season was an anomaly, showing extraordinarily high peak height but a slow growth rate (0.25) and a low curvature (22.23), indicating a gradual increase in cases. The 2018-19 season had an early peak with significant outbreak levels, a high curvature of 100.02, and a steady mean slope. The 2019-20 season saw a balanced mid-season peak, with moderate growth and curvature, and the highest peak value, reflecting a controlled outbreak but a significant height of cases.

Based on the observed dynamics of the peaks and model performance, specific models are recommended for each season type. For rapid growth early peaks, as seen in the 2013-14 season, Linear regression and Simple RNN were the most suitable models. Linear regression demonstrated the lowest error rates, particularly when covariates were excluded, while Simple RNN performed well with covariates included. In the 2014-15 season, which exhibited severe and high curvature peaks, Simple RNN proved to be highly adaptable, while Linear regression provided stable predictions. For seasons with late peaks and a gradual onset, like 2015-16, both Linear regression and Simple RNN were effective, handling the gradual buildup and delayed peak well.

In the 2016-17 season, which experienced steady growth and an early peak, XGBoost and Adaboost emerged as the top performers, displaying the lowest errors in this scenario. Anomalous peaks with slow growth, such as those in the 2017-18 season, were best handled by Linear regression, SGD regressor, and Simple RNN. Linear regression and SGD regressor worked well with covariates, while Simple RNN demonstrated robustness even without covariates. In the 2018-19 season, which saw an early peak and a significant outbreak, Adaboost, Random Forest, and Temporal CNN were the top models, effectively managing the early peak and significant case counts. Finally, for seasons with balanced progression and moderate peaks, like 2019-20, Linear regression and XGBoost were the most suitable models, reflecting steady growth and balanced progression throughout the season.

6.4 Discussion

The results of this study provide a comprehensive analysis of the performance of various machine learning models in predicting influenza peaks across multiple seasons from 2013 to 2020. This analysis highlights the strengths and weaknesses of each model under different conditions, particularly the inclusion and exclusion of Southern Hemisphere covariates. The

discussion focuses on key insights and broader implications derived from these results while also considering limitations and areas for future research.

From the analysis, several general insights emerged regarding the suitability of different models. Linear regression consistently ranked among the top performers across multiple seasons, particularly in early seasons characterized by rapid growth and early peaks. It exhibited the lowest mean errors, which shows its strong predictive performance and reliability. Ensemble methods, such as XGBoost, Adaboost, and Random Forest, excelled in capturing complex patterns, particularly when covariates were included, making them ideal for seasons with intricate dynamics. Recurrent Neural Networks (Simple RNN and LSTM) also showed strong performance, especially with the inclusion of covariates. This indicates their capability to handle complex, time-dependent data, with Simple RNN consistently ranking as one of the top performers. The inclusion of Southern Hemisphere covariates generally improved performance across all models, especially for more complex models like Simple RNN and LSTM, suggesting that leveraging global influenza data enhances model accuracy.

While the study provides valuable insights, several limitations must be acknowledged. Models such as Random Forest and LSTM exhibited signs of overfitting, particularly when trained on seasons with volatile patterns, which reduced their generalizability. Furthermore, there was inconsistency in model performance across different seasons, reflecting a lack of robustness in handling the inherent variability of influenza outbreaks. Additionally, many models heavily relied on historical influenza patterns, making them less effective at adapting to novel strains or atypical outbreak patterns. Lastly, data quality and completeness posed significant challenges, as missing data or noise could significantly degrade model performance.

To enhance model performance and applicability, several recommendations are proposed. First, incorporating mechanistic models alongside data-driven approaches could improve both robustness and interpretability. Second, enhancing real-time adaptability by developing models that can dynamically update with real-time data would allow for a better response to emerging influenza patterns. Third, improving the communication of uncertainty by utilizing probabilistic forecasting methods could provide a range of possible outcomes, supporting risk-based decision-making. Finally, integrating diverse data sources, such as vaccination rates and public health interventions, would likely improve prediction accuracy by capturing factors influencing the spread of influenza.

6.5 Conclusion

This study evaluated a range of machine learning models for predicting influenza peak characteristics across several seasons from 2013 to 2020. The findings demonstrate the importance

of selecting appropriate models based on the specific characteristics of each influenza season.

Linear Regression consistently emerged as a reliable and robust model, particularly in seasons with rapid growth and early peaks. Ensemble methods like XGBoost and Adaboost excelled in seasons with steady growth and complex dynamics. Recurrent Neural Networks (Simple RNN, LSTM) demonstrated strong performance with covariate inclusion, showing potential in handling time-series data for influenza forecasting.

However, limitations such as overfitting, inconsistent performance across seasons, and sensitivity to data quality highlight the need for further research. Future work should focus on enhancing real-time adaptability, integrating mechanistic insights, and incorporating diverse data sources to improve model robustness and utility for public health decision-making.

Overall, machine learning models offer significant potential for improving influenza prediction. By addressing current limitations through methodological advancements and the integration of high-quality, diverse data, these models can become valuable tools for timely and effective public health responses during influenza outbreaks.

Chapter 7

Conclusion and Future Directions

This thesis has presented a multi-faceted approach to analyzing and forecasting epidemic dynamics, drawing on rigorous probabilistic modelling, state-space frameworks, Bayesian inference, stochastic simulation methods, mechanistic compartmental models, and advanced stochastic mean-reverting processes, along with modern machine learning approaches. Each chapter contributed a key methodological component, advancing from foundational concepts in time series and Bayesian learning, through state-space modelling and stochastic simulation, to mechanistic epidemic modelling and machine learning-based peak forecasting. By unifying these diverse strands, the thesis provides a richer, more robust toolkit for understanding and anticipating epidemic behaviour in real-world settings.

7.1 Novelty of the work

The novelty of this thesis emerges from the integrative approach that spans several conceptual and methodological domains, culminating in a comprehensive framework suited to contemporary public health challenges.

In Chapter 1, the groundwork was laid with background material and terminology related to time series analysis and Bayesian learning. A common foundation that transcends traditional compartmental approaches is established, which embraces uncertainty quantification and flexible model structures. By emphasizing Bayesian principles, the thesis set the stage for a cohesive probabilistic paradigm that informs subsequent chapters.

Chapter 2 introduced the state-space modelling framework, initially focusing on linear and normal assumptions, as well as dynamic regression techniques. The novelty in this chapter is twofold. First, by adopting state-space representations, the thesis enabled the separation of signal and noise, facilitating better inference of latent states like underlying epidemic levels or time-varying parameters. Second, the integration of online, real-time

multivariate covariance estimation methods provided tools to adaptively track evolving uncertainties. This adaptive capability is crucial in public health, where the underlying system may shift due to new interventions or emerging pathogen variants.

Advancing to Chapter 3, the thesis goes beyond linear and Gaussian restrictions by employing stochastic simulation methods, including Markov chain Monte Carlo approaches, to handle general and potentially non-linear state-space models. The novelty here was the application of advanced simulation-based inference for non-linear regression case studies and the introduction of online structural break detection using sequential Monte Carlo methods. This broke new ground by allowing the model to detect regime changes in real-time, which is akin to identifying sudden changes in transmission patterns or the emergence of a new strain during an epidemic, without relying solely on static assumptions.

Chapter 4 transitioned the thesis from purely statistical frameworks to mechanistic modelling. By formulating a compartmental model linked to a mobility-enhanced, probabilistic, data-driven framework, this chapter introduced a crucial epidemiological dimension. The novelty stemmed from combining mechanistic insights (e.g., from extended SIR-type models) with data-driven inference, thus allowing for the calibration of model parameters against observed data, as well as the incorporation of mobility patterns to reflect realistic disease spread. Applying these methods to real data and comparing them to established results demonstrated the practical utility and versatility of the proposed methodologies.

Chapter 5 added onto these foundations by extending a classical SIRS model with stochastic, mean-reverting processes to capture realistic epidemic variability. Traditional deterministic models often assume constant or overly simplistic time-varying transmission rates, whereas this thesis introduced the CKLS, and other mean-reverting frameworks. The novelty here lies in bridging epidemiological theory with stochastic processes that preserve positivity and exhibit realistic reversion to long-term means, ultimately yielding more flexible and accurate models. Deriving extinction thresholds under these stochastic dynamics offered new theoretical insights into when diseases fade out or persist, enhancing our conceptual understanding of epidemic control under uncertainty.

Finally, Chapter 6 introduced a machine and deep learning framework for analyzing epidemic peak characteristics, integrating global and hemispheric data sources. The novelty here was the synergistic use of advanced predictive models which range from linear regression and Lasso to ensemble methods and neural architectures (RNNs, LSTMs), together with cross-hemisphere data. This approach provided improvements in forecasting accuracy, demonstrating that global contextual information, combined with powerful machine learning techniques, can improve early warning capabilities for epidemic peaks.

Taken together, these chapters form a cohesive narrative that moves from foundational

time series and Bayesian concepts in Chapters 1-3, through mechanistic epidemiological modelling in Chapter 4, toward innovative stochastic and machine learning based epidemic predictions in Chapters 5-6. The overarching novelty is the integration of diverse methodologies into a unified, adaptive, and context-aware toolkit for epidemic analysis and forecasting, providing both theoretical insights and practical guidance for real-world decision-making.

7.2 Limitations and future studies

Despite these advancements, the thesis encountered several limitations that suggest avenues for further development.

Model complexity and overfitting emerged as ongoing challenges, particularly for sophisticated machine learning architectures that performed exceptionally well in some scenarios but less reliably in others. While the thesis demonstrated various methods to tune hyperparameters and employ regularization, future work could integrate Bayesian inference more deeply or use model compression [114] and transfer learning [115] to ensure that complexity does not undermine generalization.

Adapting to novel pathogens, newly emerging strains, or atypical epidemiological regimes remains difficult. The frameworks presented, though flexible, were calibrated primarily on COVID-19 or influenza data. To extend these methods to other diseases or entirely new outbreak conditions, future research should develop pathogen-specific parameters, incorporate evolutionary models that track changes in transmissibility or immunity, and stress-test the frameworks under extreme black swan conditions to ensure robust performance.

Data quality and completeness also pose persistent hurdles. Missing values, delays in reporting, and evolving data streams complicate inference. Future directions include developing more sophisticated anomaly detection techniques, robust data imputation approaches, and strategies to integrate multiple heterogeneous data sources simultaneously. Incorporating mobility data, vaccination coverage, environmental variables, and genomic sequencing results could potentially improve situational awareness and predictive power.

7.3 Incorporating behavioural changes and public health interventions

Perhaps the most significant opportunity for improvement lies in explicitly modelling human behaviour, policy interventions, and public perception. While the mechanistic models and stochastic processes in Chapters 4 and 5 capture epidemiological fundamentals, real-world

outbreaks are strongly influenced by how people respond to perceived risks and how health authorities implement interventions.

Future research can integrate behavioural adaptation directly into transmission models. Agent-based or network models could encode how contact patterns change in response to health warnings, media coverage, or policy mandates. By coupling these behavioural adaptations with mean-reverting processes that reflect changing baseline conditions, models can evolve more dynamically and realistically.

Public health interventions, such as school closures, travel restrictions, and vaccination campaigns, can be included as explicit covariates or as control variables. Evaluating counterfactual scenarios using machine learning optimization techniques could identify which interventions would have had the greatest impact on reducing outbreak peaks. Such scenario planning would move from passive forecasting toward actionable guidance, helping health officials tailor interventions more effectively.

Hybrid models that blend mechanistic insight with data-driven predictions will likely prove most powerful. Validating these hybrid methods against historical events, such as changes in vaccination policy or the emergence of antiviral drugs, can ensure that models remain grounded in empirical reality. Over time, these models can continuously update their parameters as new data arrive, adapting to changing conditions in ways that purely static or deterministic models cannot.

7.4 Conclusion

This thesis advanced epidemic forecasting by pushing beyond conventional models, harnessing both theoretical rigour and empirical flexibility. By uniting time series analysis, Bayesian state-space modelling, stochastic simulation, mechanistic compartmental methods, and machine learning, it has laid a robust methodological foundation for next-generation epidemic forecasting tools. The findings demonstrate that model choice, data context, and global covariates all matter, and that real-time adaptability is essential for effective epidemic prediction and management.

Opportunities for future improvement are abundant. Addressing overfitting, enhancing real-time updates, incorporating richer data sources, and explicitly modelling behavioural changes and interventions can yield more robust, socially relevant, and actionable forecasts. Such improvements will guide public health authorities and communities to anticipate and mitigate outbreaks, safeguarding lives and societal well-being.

Continuous refinement of these methods and broadening of their scope can provide the impetus for future research which produce epidemiological models that are not only techni-

cally sound but practically relevant. By pursuing these collective efforts, we advance toward a future in which timely, evidence-based interventions preempt widespread epidemics and effectively mitigate their human and economic burdens.

Appendix A

A.1 Additional figures for Chapter 5

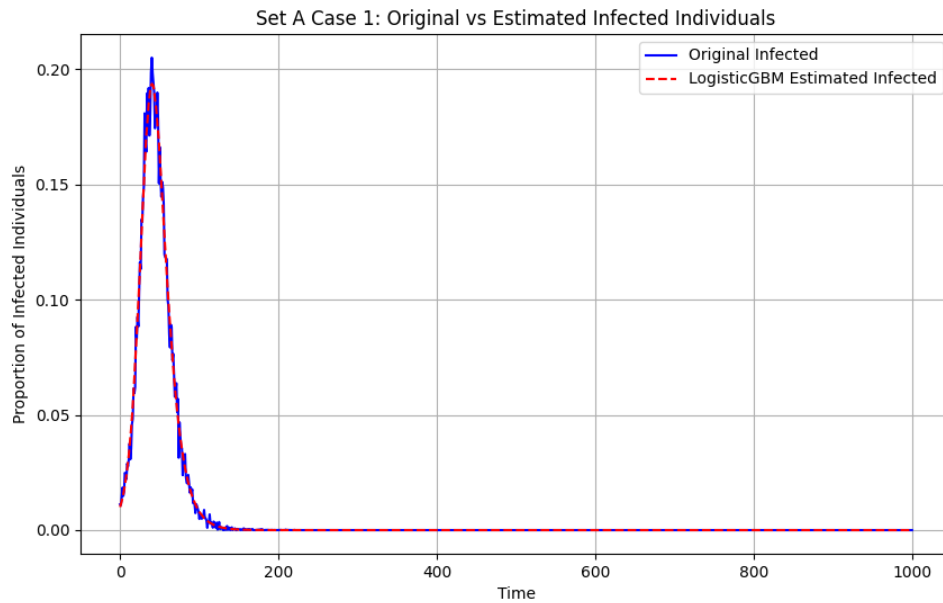


Figure A.1: A plot of the original synthetic infected data along with the estimated SIRS model with LogisticGBM transmission process. The data set and case number are A and 1, respectively.

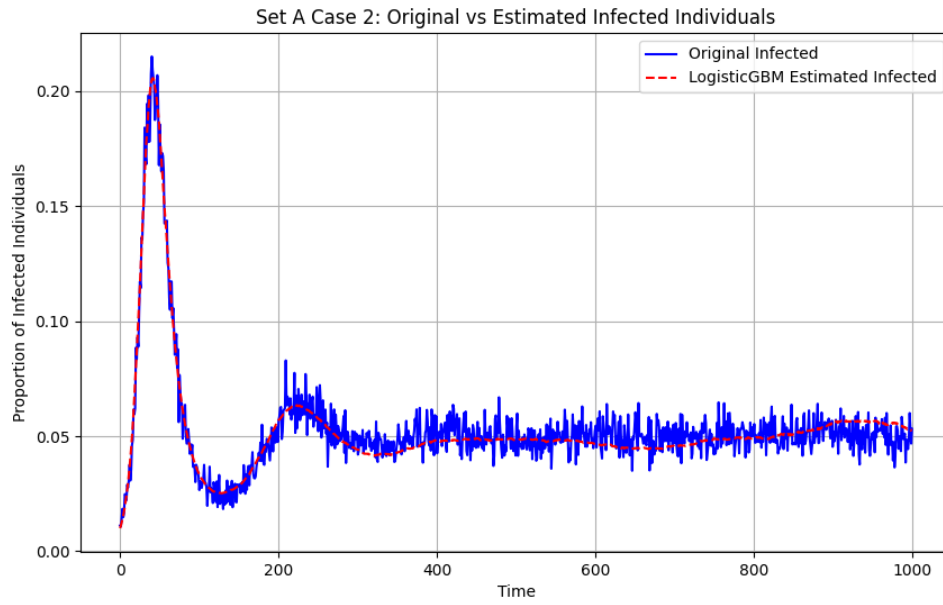


Figure A.2: A plot of the original synthetic infected data along with the estimated SIRS model with LogisticGBM transmission process. The data set and case number are A and 2, respectively.

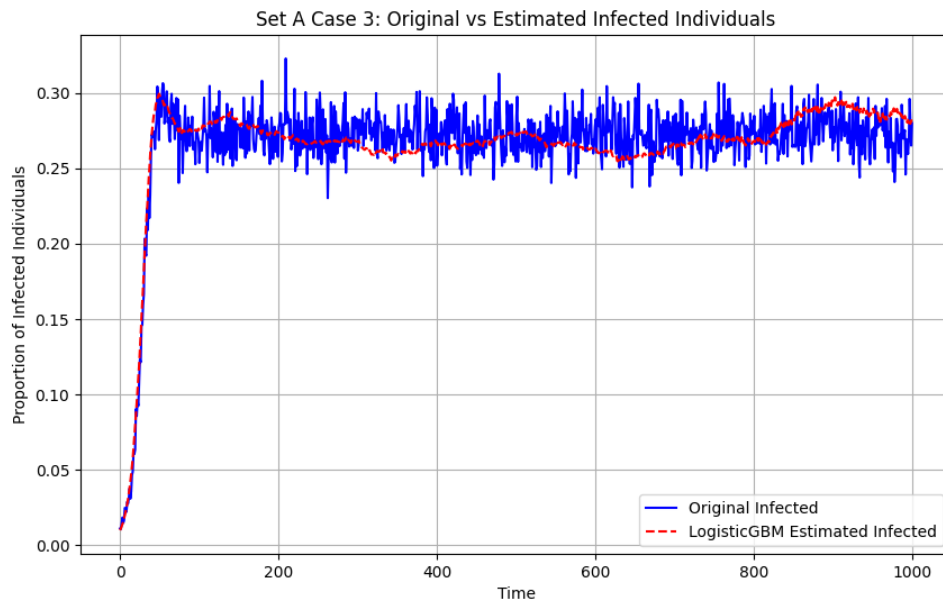


Figure A.3: A plot of the original synthetic infected data along with the estimated SIRS model with LogisticGBM transmission process. The data set and case number are A and 3, respectively.

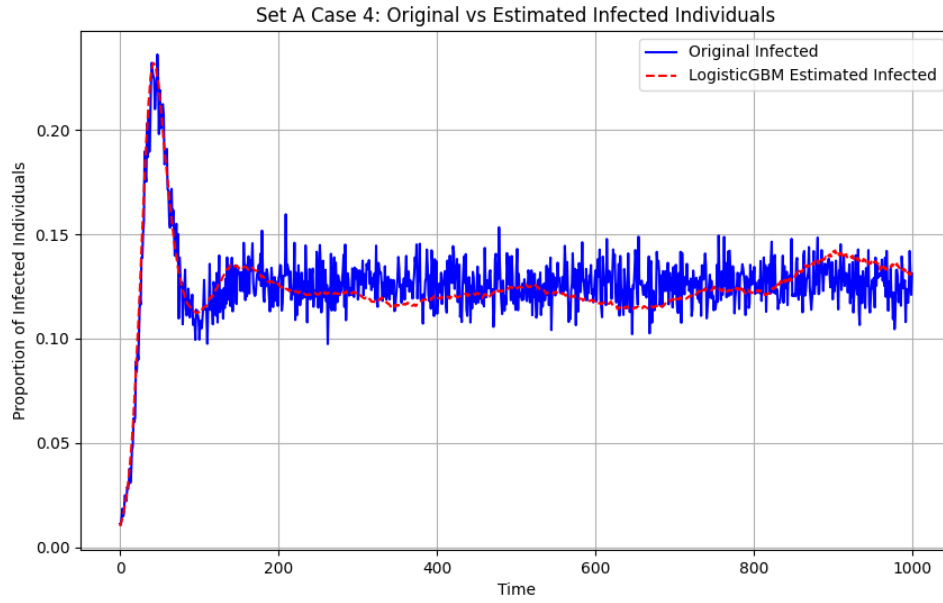


Figure A.4: A plot of the original synthetic infected data along with the estimated SIRS model with LogisticGBM transmission process. The data set and case number are A and 4, respectively.

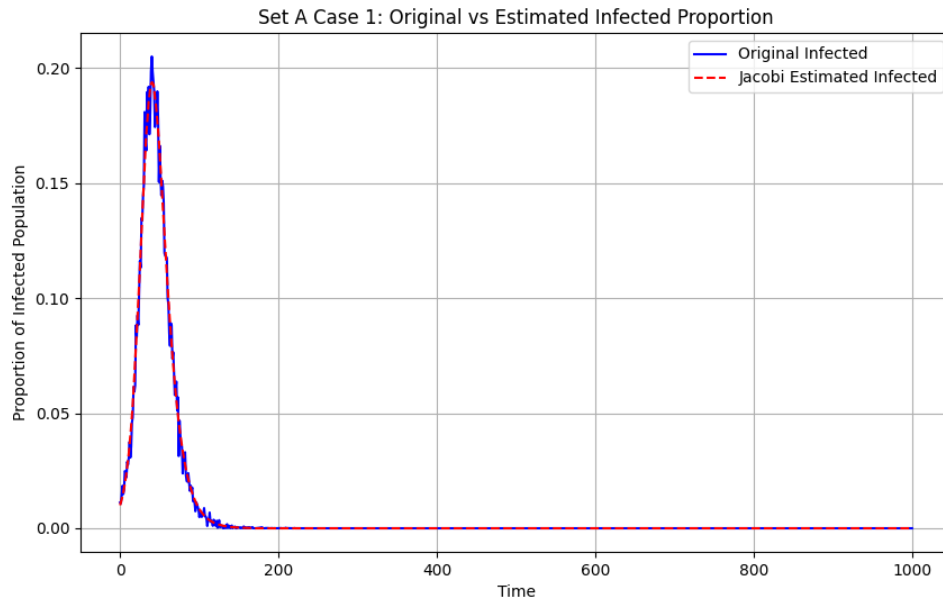


Figure A.5: A plot of the original synthetic infected data along with the estimated SIRS model with Jacobi transmission process. The data set and case number are A and 1, respectively.

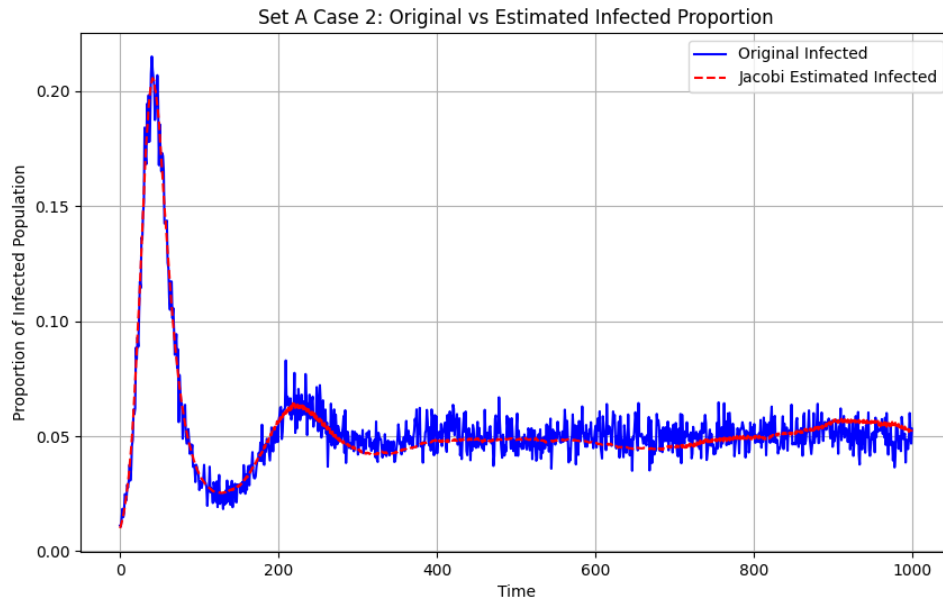


Figure A.6: A plot of the original synthetic infected data along with the estimated SIRS model with Jacobi transmission process. The data set and case number are A and 2, respectively.

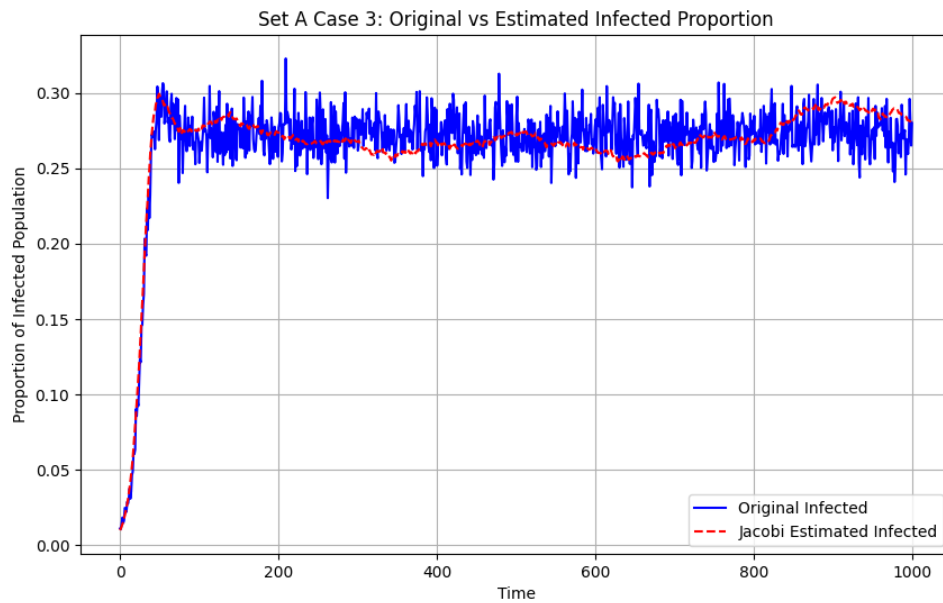


Figure A.7: A plot of the original synthetic infected data along with the estimated SIRS model with Jacobi transmission process. The data set and case number are A and 3, respectively.

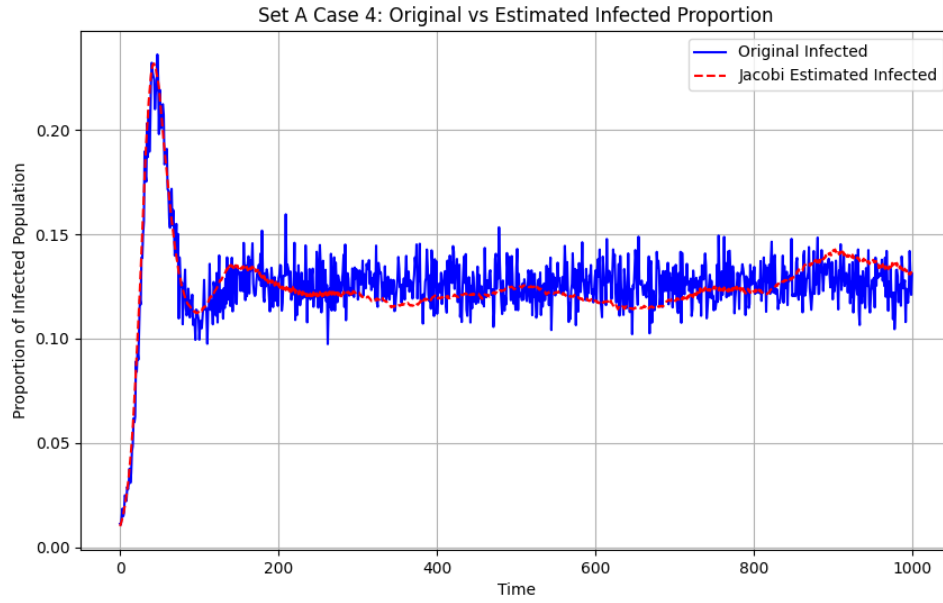


Figure A.8: A plot of the original synthetic infected data along with the estimated SIRS model with Jacobi transmission process. The data set and case number are A and 4, respectively.

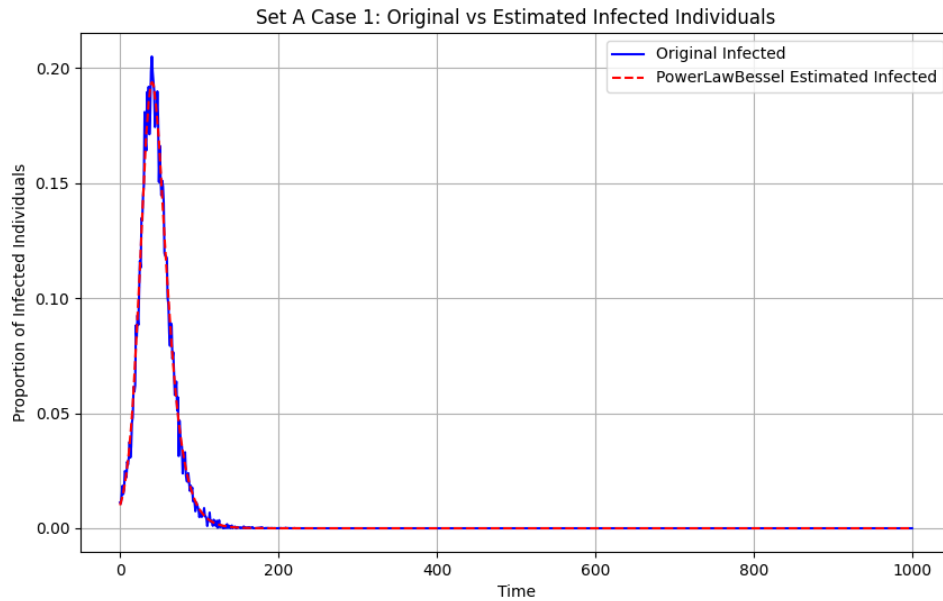


Figure A.9: A plot of the original synthetic infected data along with the estimated SIRS model with PowerLawBessel transmission process. The data set and case number are A and 1, respectively.

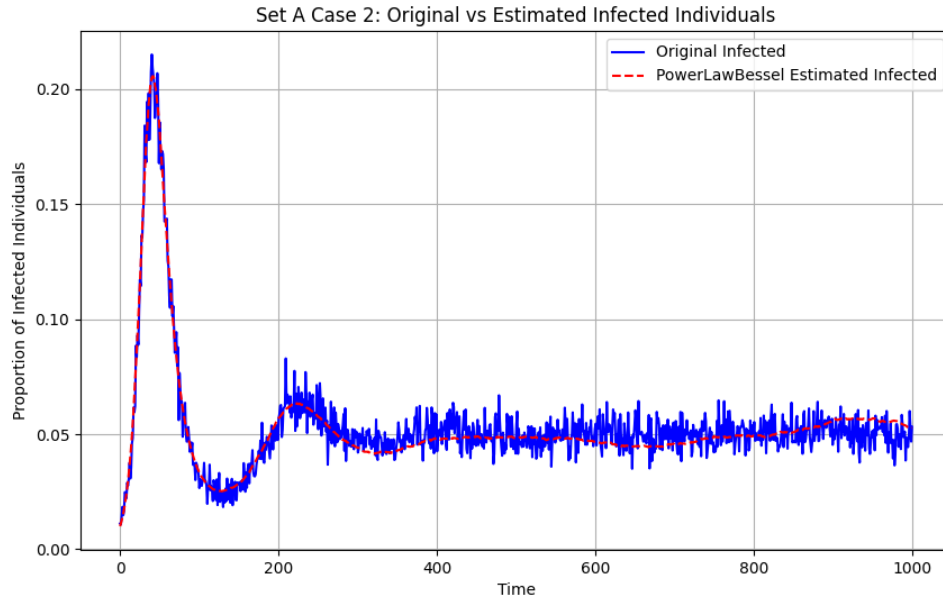


Figure A.10: A plot of the original synthetic infected data along with the estimated SIRS model with PowerLawBessel transmission process. The data set and case number are A and 2, respectively.

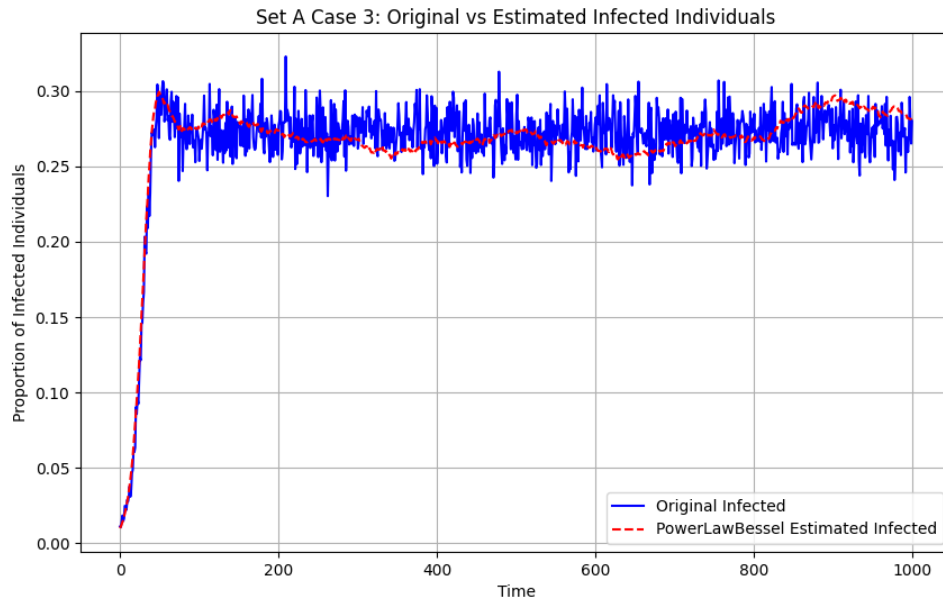


Figure A.11: A plot of the original synthetic infected data along with the estimated SIRS model with PowerLawBessel transmission process. The data set and case number are A and 3, respectively.

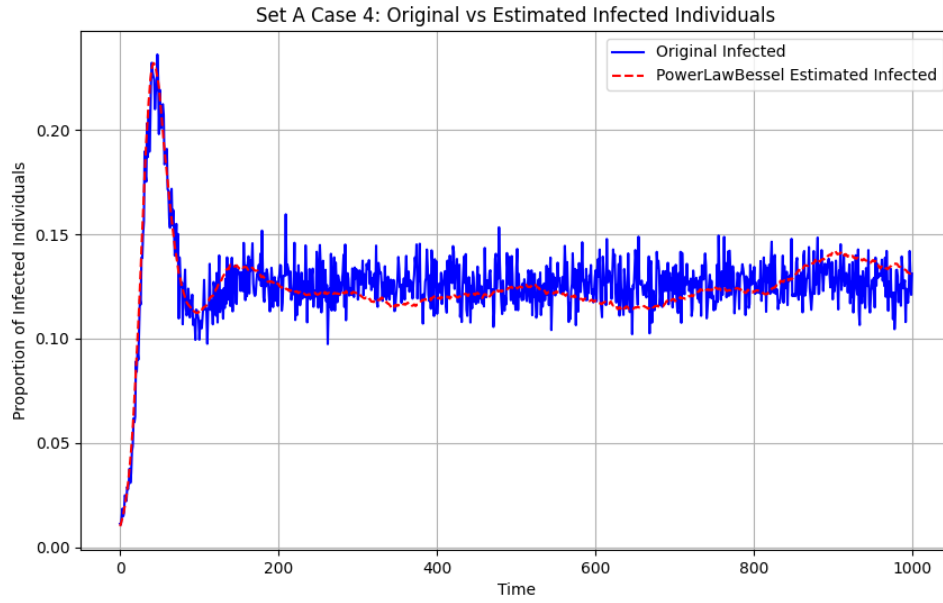


Figure A.12: A plot of the original synthetic infected data along with the estimated SIRS model with PowerLawBessel transmission process. The data set and case number are A and 4, respectively.

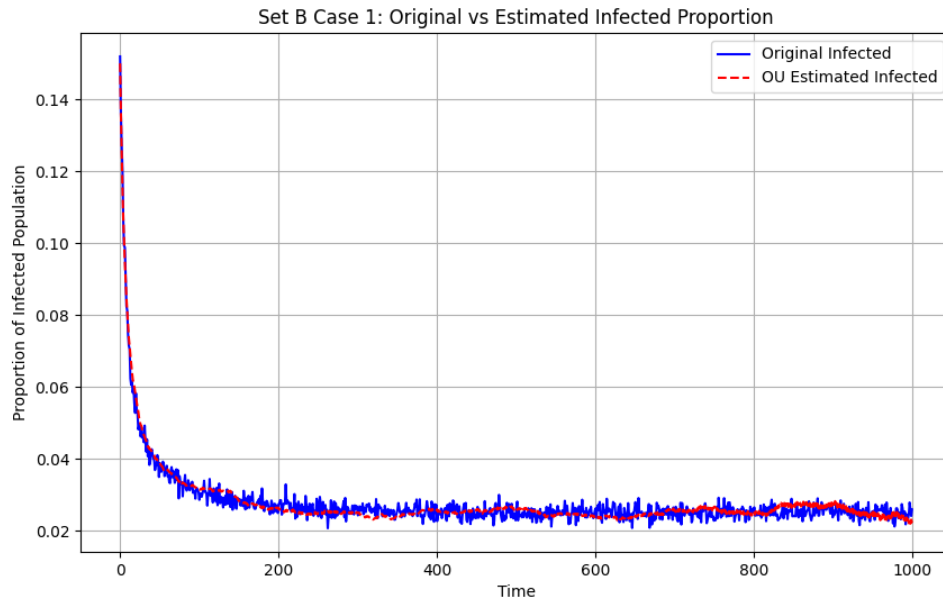


Figure A.13: A plot of the original synthetic infected data along with the estimated SIRS model with OU transmission process. The data set and case number are B and 1, respectively.

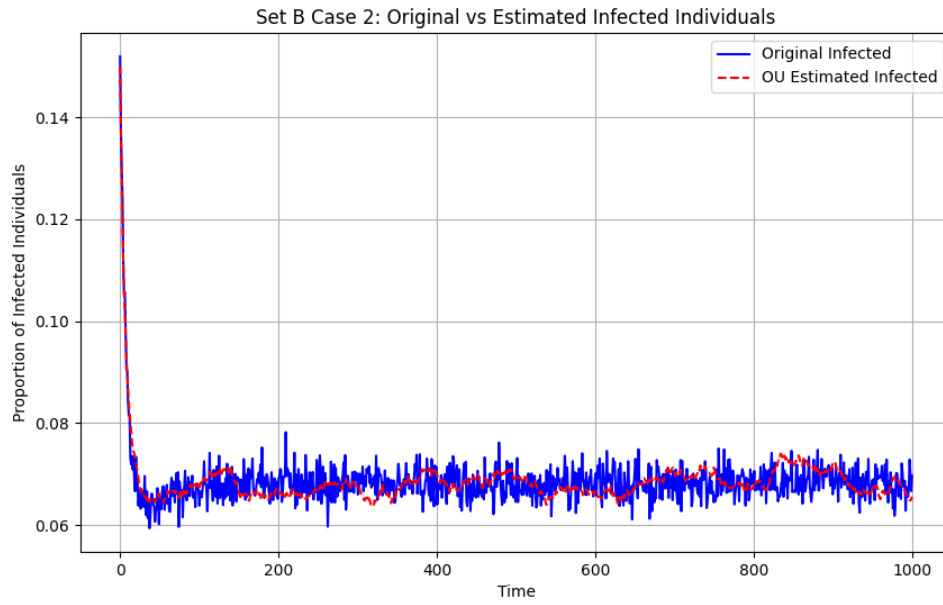


Figure A.14: A plot of the original synthetic infected data along with the estimated SIRS model with OU transmission process. The data set and case number are B and 2, respectively.

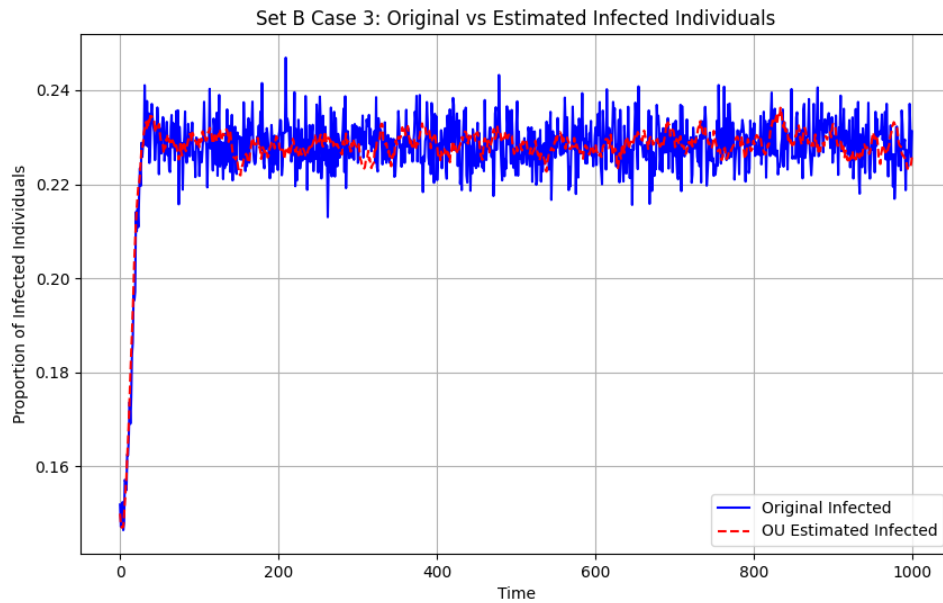


Figure A.15: A plot of the original synthetic infected data along with the estimated SIRS model with OU transmission process. The data set and case number are B and 3, respectively.

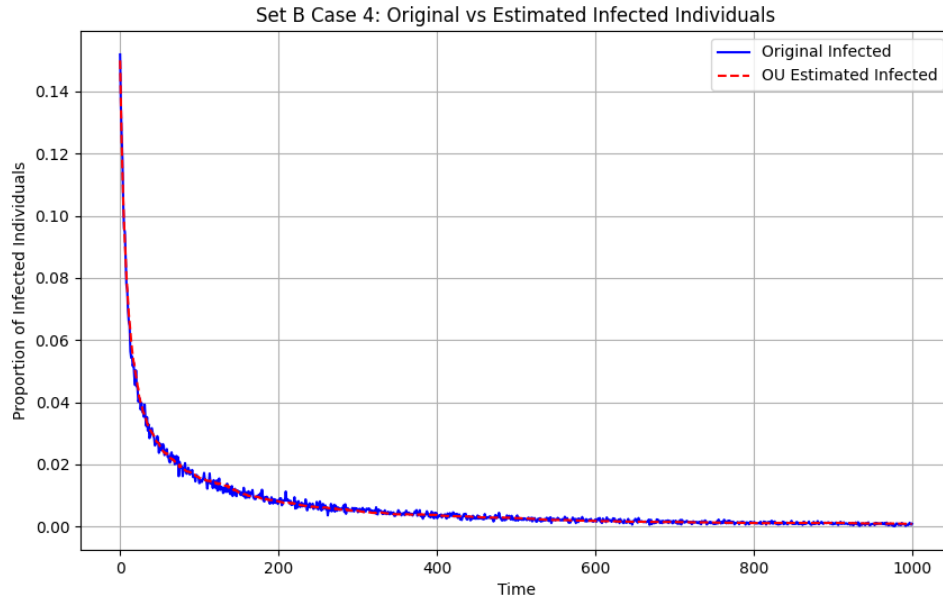


Figure A.16: A plot of the original synthetic infected data along with the estimated SIRS model with OU transmission process. The data set and case number are B and 4, respectively.

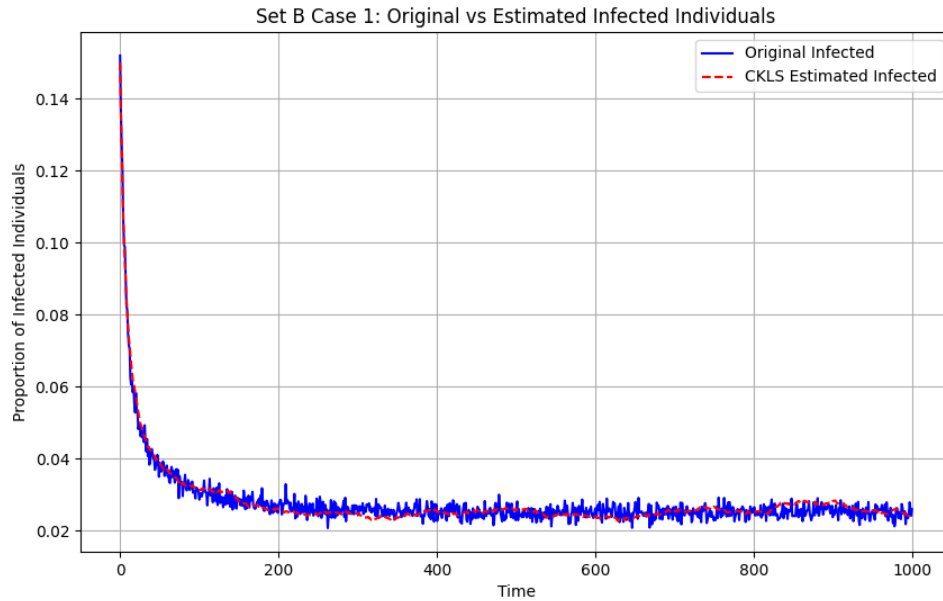


Figure A.17: A plot of the original synthetic infected data along with the estimated SIRS model with CKLS transmission process. The data set and case number are B and 1, respectively.

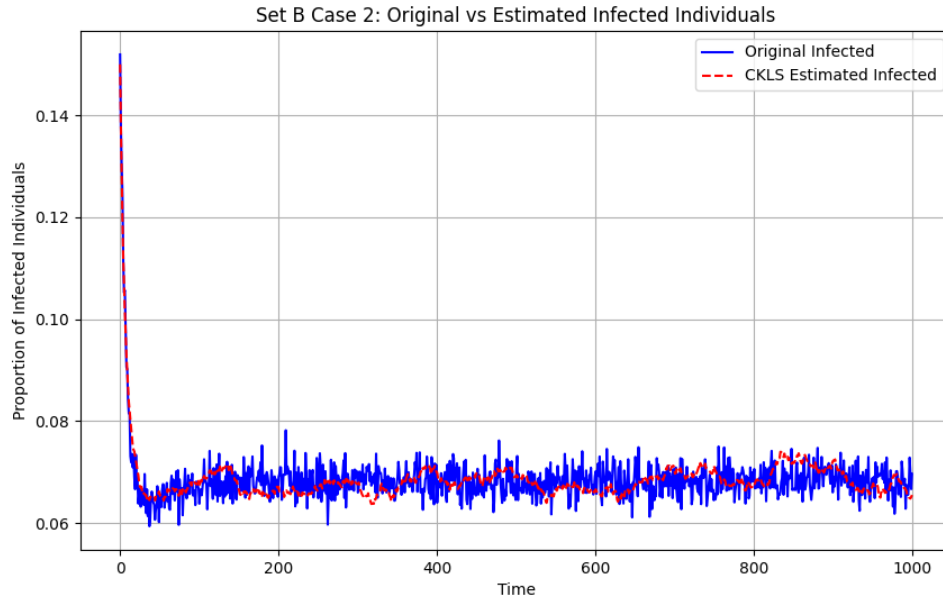


Figure A.18: A plot of the original synthetic infected data along with the estimated SIRS model with CKLS transmission process. The data set and case number are B and 2, respectively.

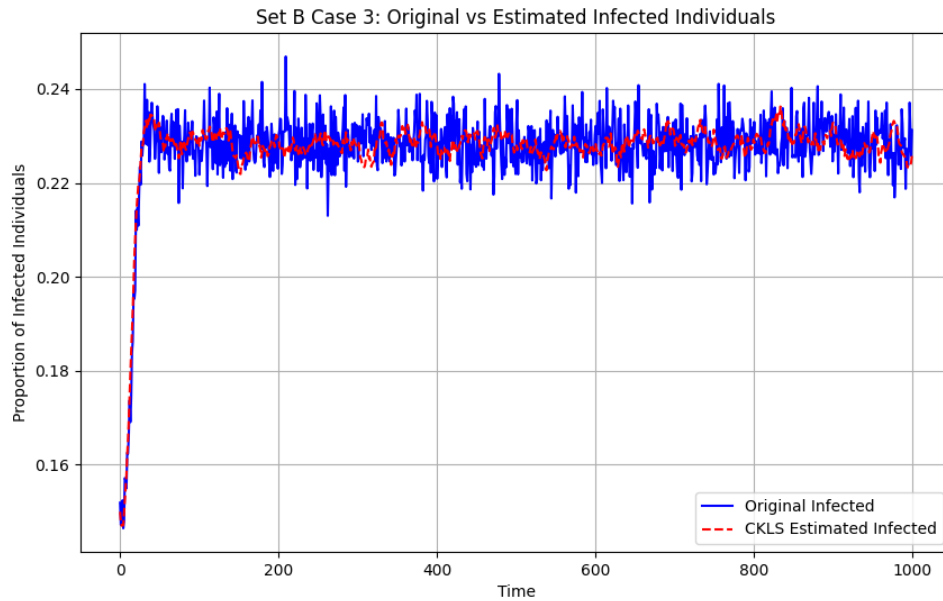


Figure A.19: A plot of the original synthetic infected data along with the estimated SIRS model with CKLS transmission process. The data set and case number are B and 3, respectively.

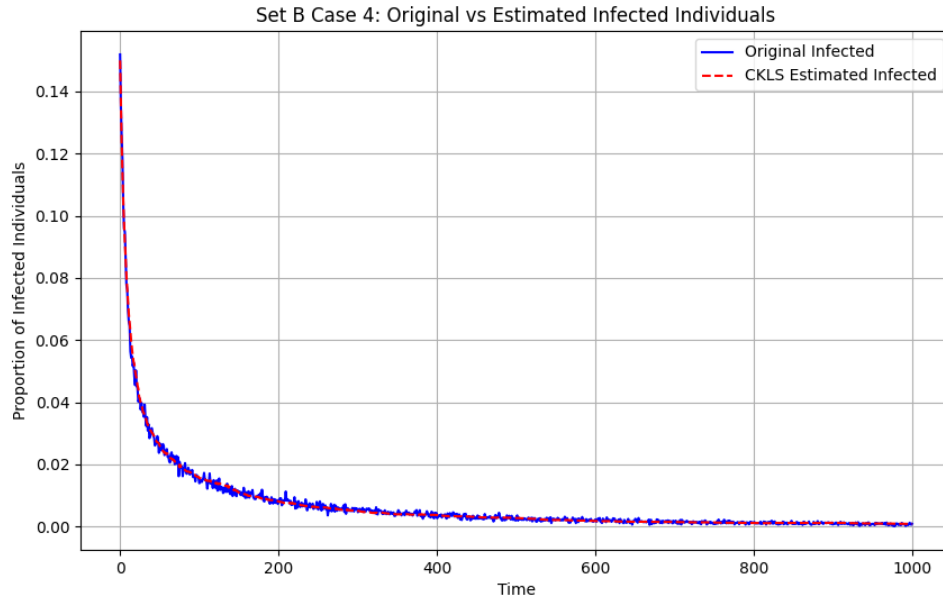


Figure A.20: A plot of the original synthetic infected data along with the estimated SIRS model with CKLS transmission process. The data set and case number are B and 4, respectively.

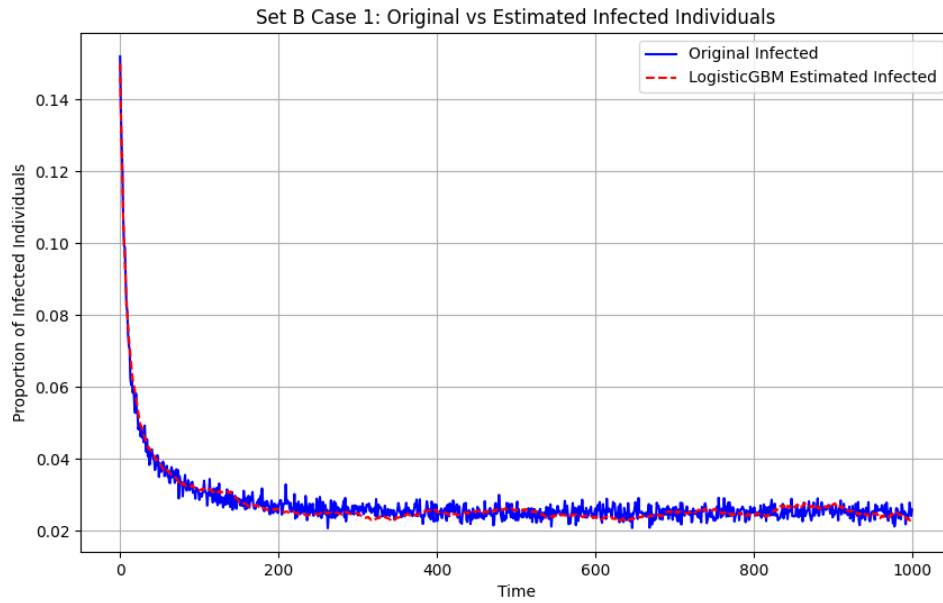


Figure A.21: A plot of the original synthetic infected data along with the estimated SIRS model with LogisticGBM transmission process. The data set and case number are B and 1, respectively.

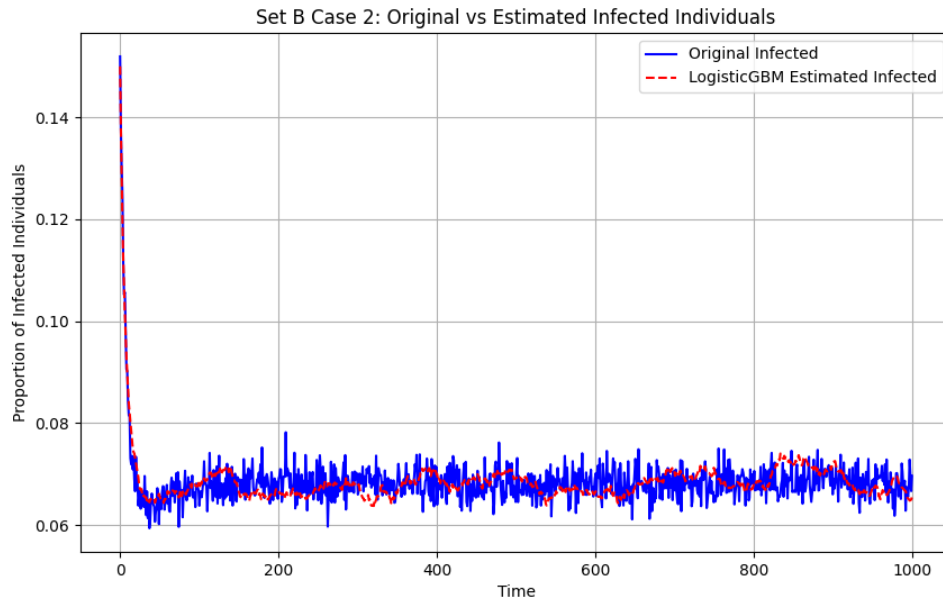


Figure A.22: A plot of the original synthetic infected data along with the estimated SIRS model with LogisticGBM transmission process. The data set and case number are B and 2, respectively.

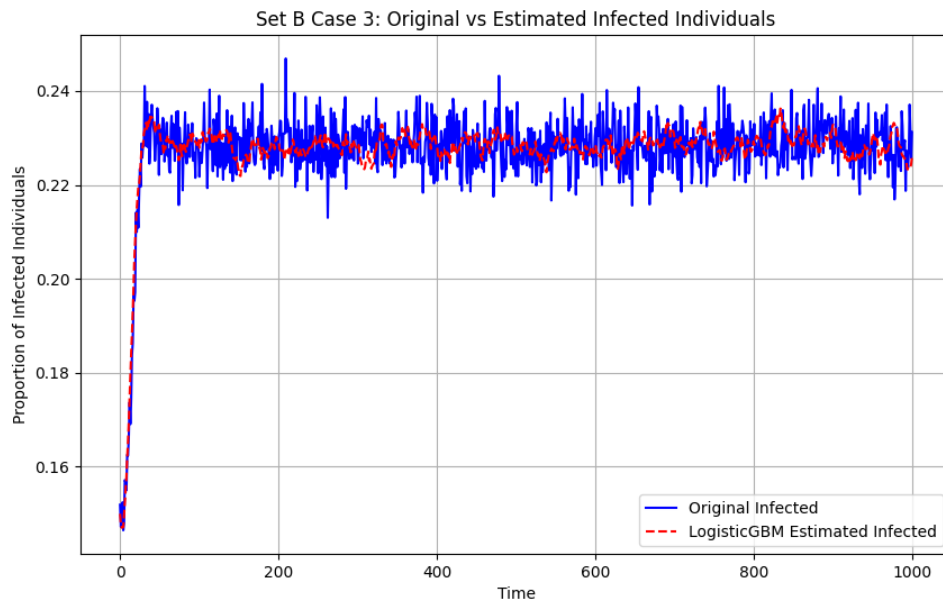


Figure A.23: A plot of the original synthetic infected data along with the estimated SIRS model with LogisticGBM transmission process. The data set and case number are B and 3, respectively.

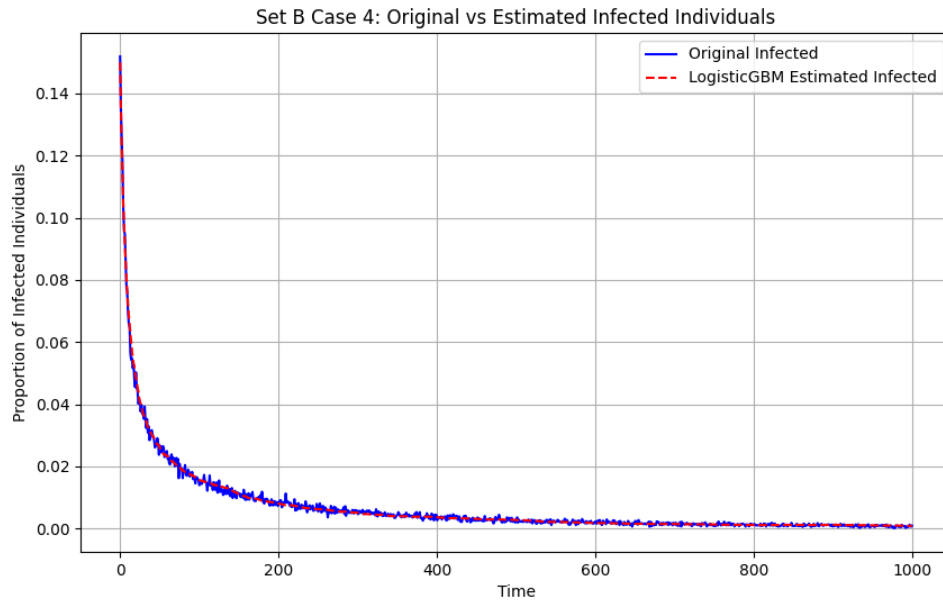


Figure A.24: A plot of the original synthetic infected data along with the estimated SIRS model with LogisticGBM transmission process. The data set and case number are B and 4, respectively.

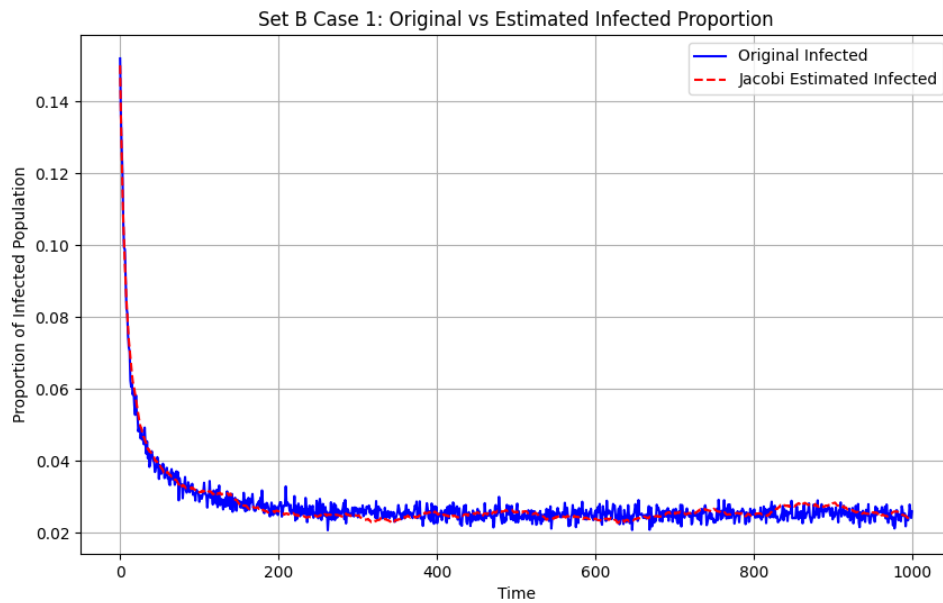


Figure A.25: A plot of the original synthetic infected data along with the estimated SIRS model with Jacobi transmission process. The data set and case number are B and 1, respectively.

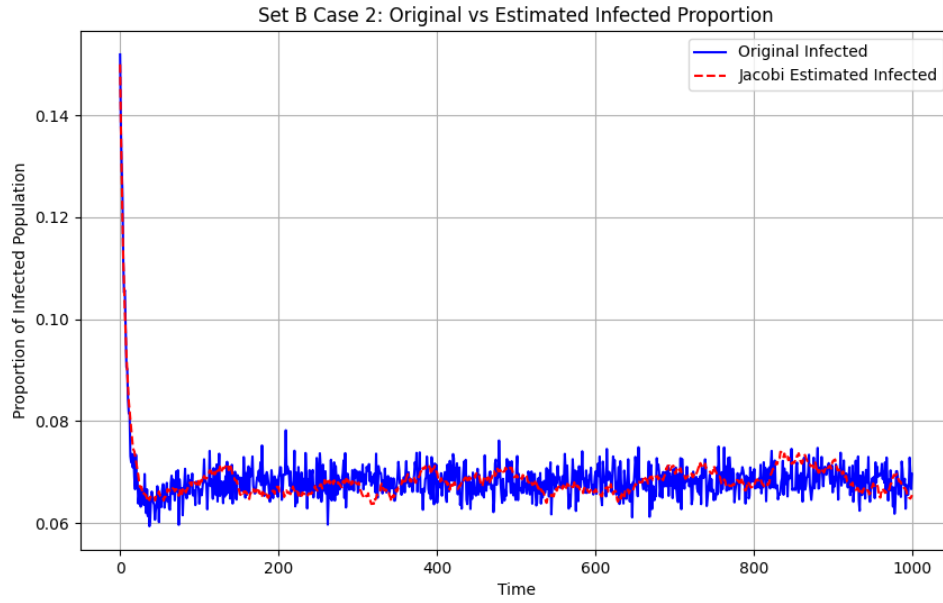


Figure A.26: A plot of the original synthetic infected data along with the estimated SIRS model with Jacobi transmission process. The data set and case number are B and 2, respectively.

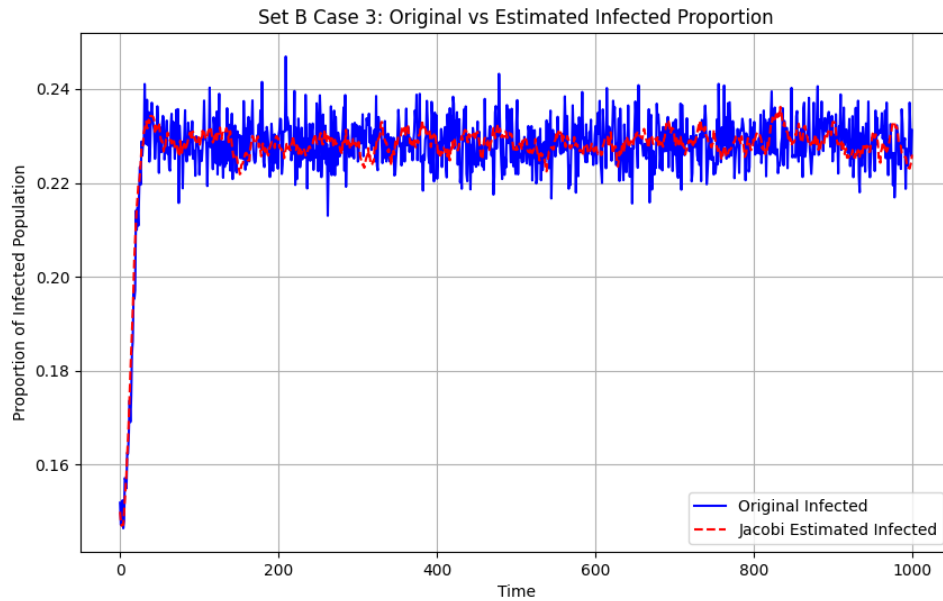


Figure A.27: A plot of the original synthetic infected data along with the estimated SIRS model with Jacobi transmission process. The data set and case number are B and 3, respectively.

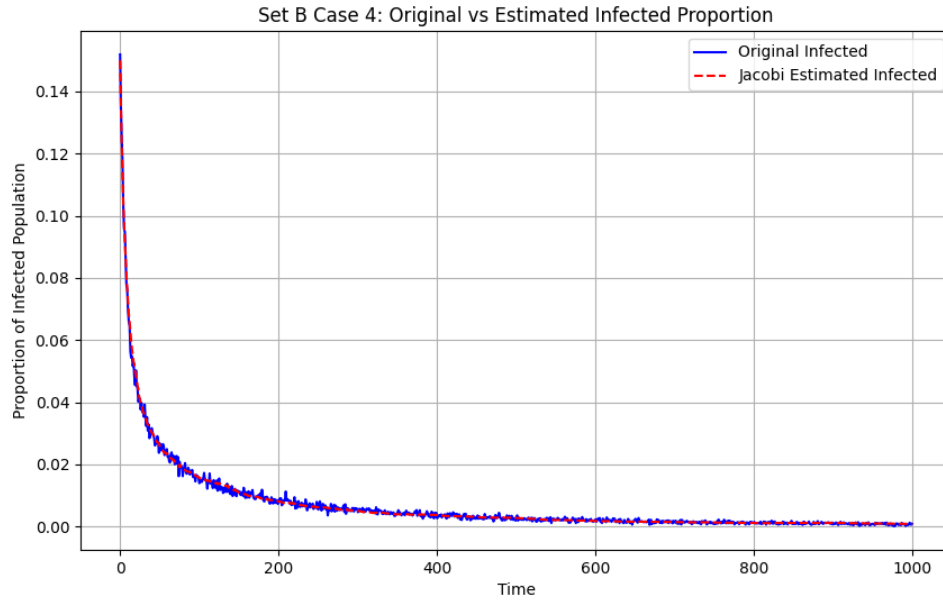


Figure A.28: A plot of the original synthetic infected data along with the estimated SIRS model with Jacobi transmission process. The data set and case number are B and 4, respectively.

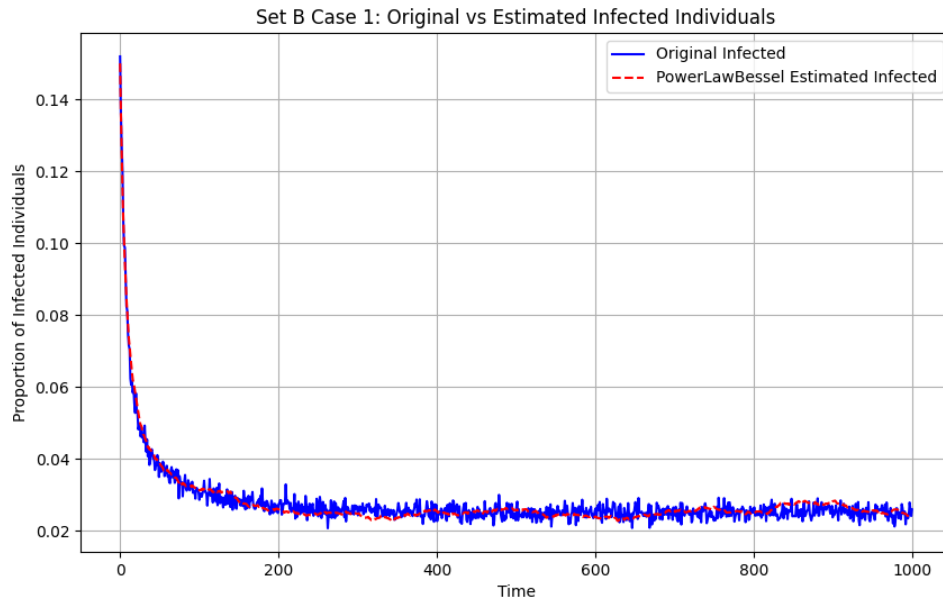


Figure A.29: A plot of the original synthetic infected data along with the estimated SIRS model with PowerLawBessel transmission process. The data set and case number are B and 1, respectively.

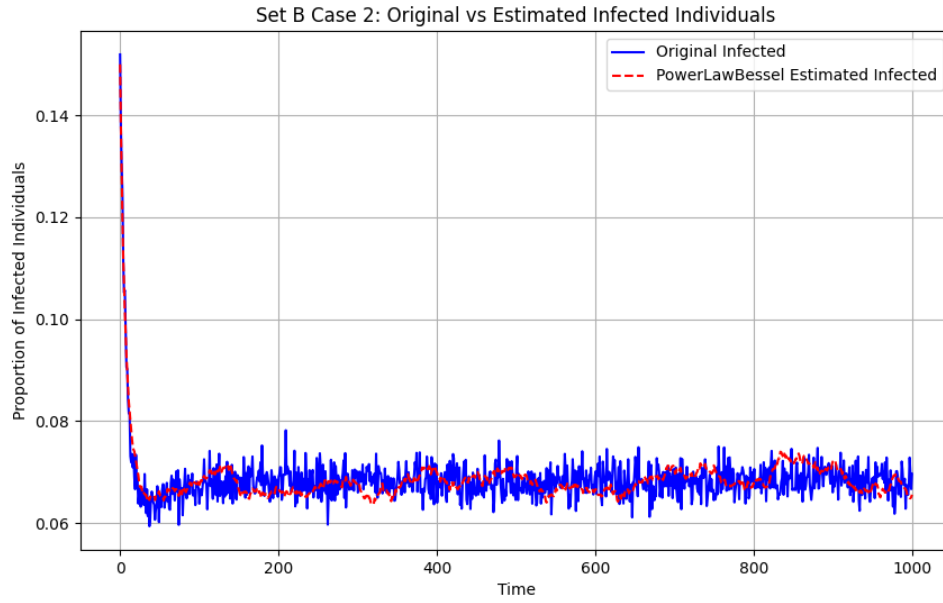


Figure A.30: A plot of the original synthetic infected data along with the estimated SIRS model with PowerLawBessel transmission process. The data set and case number are B and 2, respectively.

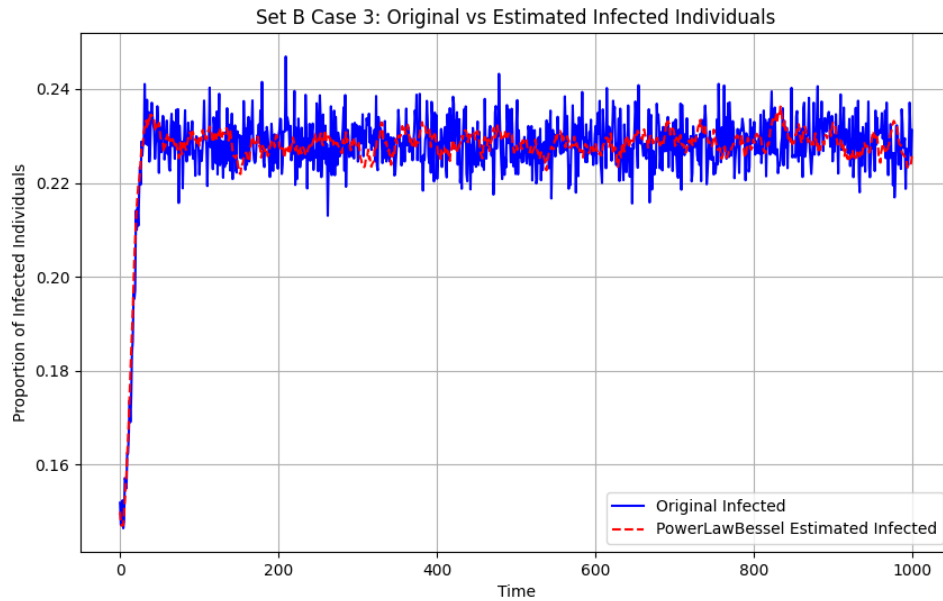


Figure A.31: A plot of the original synthetic infected data along with the estimated SIRS model with PowerLawBessel transmission process. The data set and case number are B and 3, respectively.

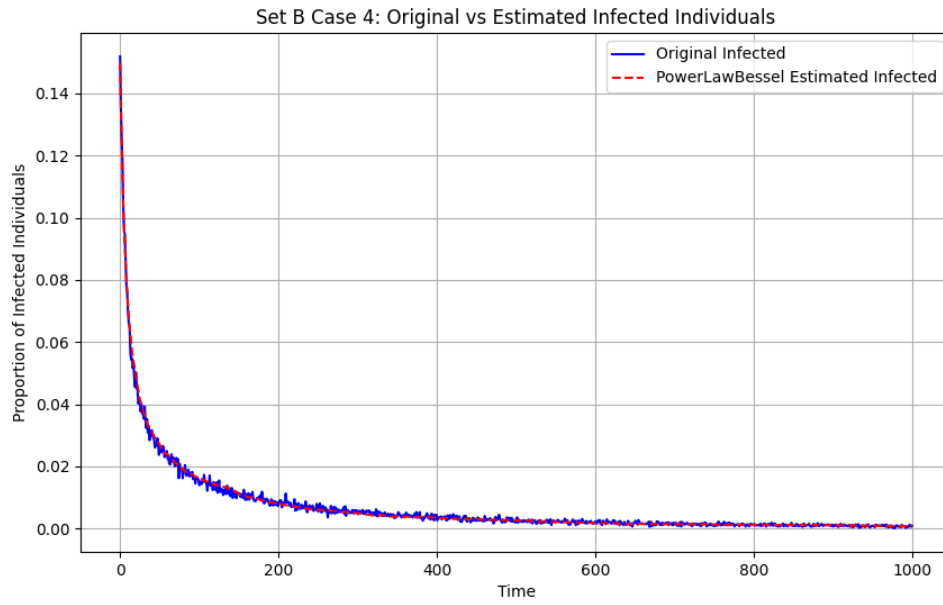


Figure A.32: A plot of the original synthetic infected data along with the estimated SIRS model with PowerLawBessel transmission process. The data set and case number are B and 4, respectively.

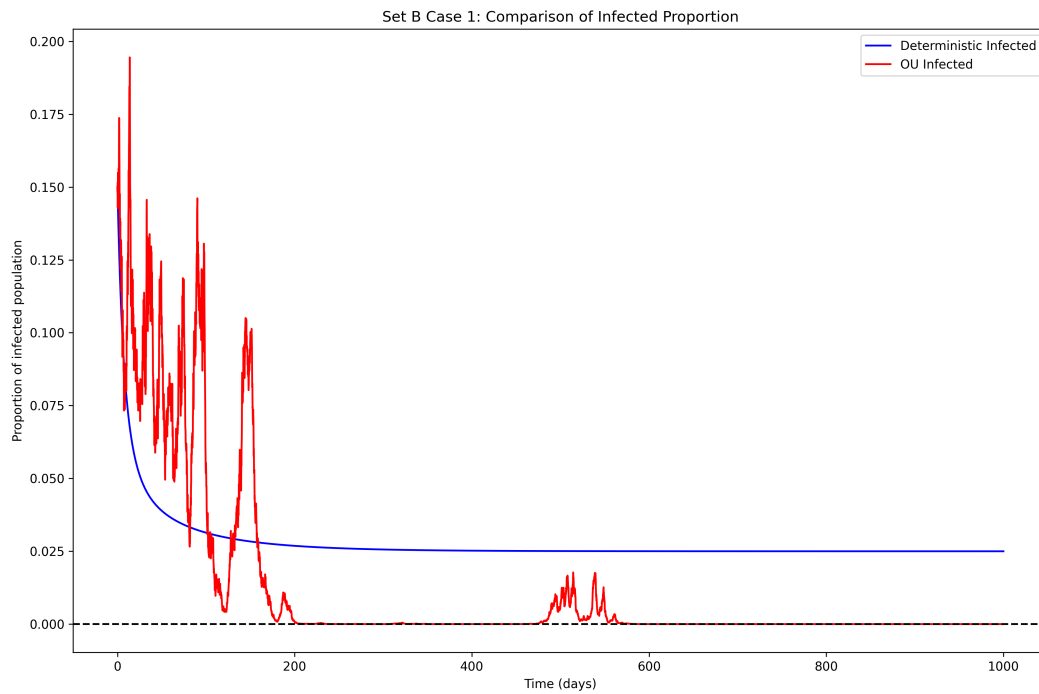


Figure A.33: A plot of the deterministic solution along with the simulated SIRS model with the OU transmission process. The data set is B and the case number is 1.

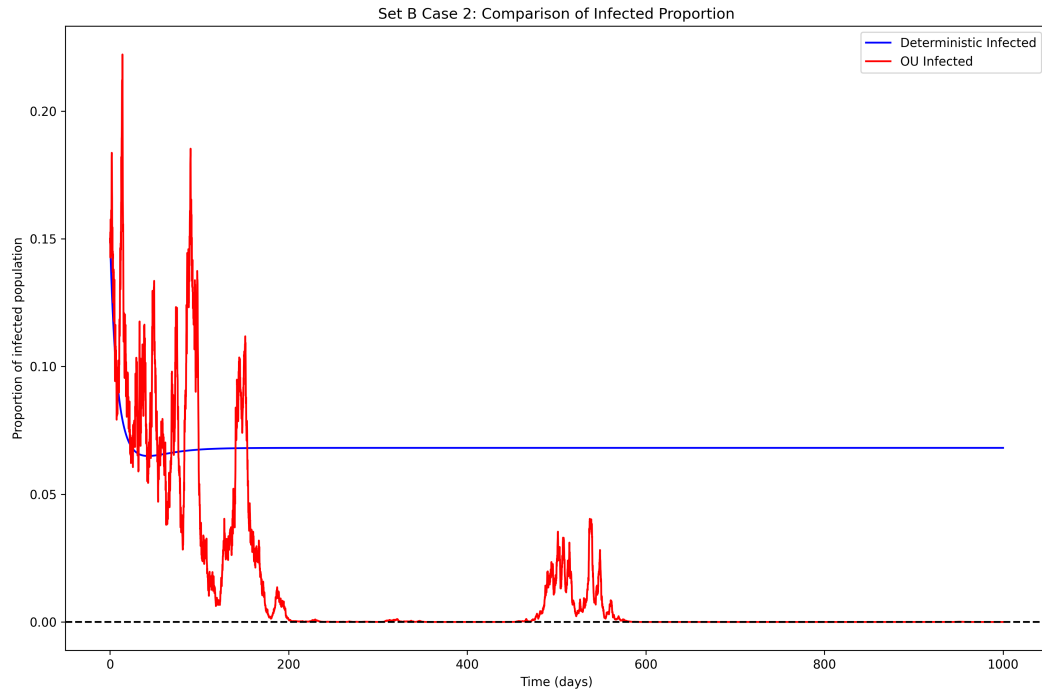


Figure A.34: A plot of the deterministic solution along with the simulated SIRS model with the OU transmission process. The data set is B and the case number is 2.

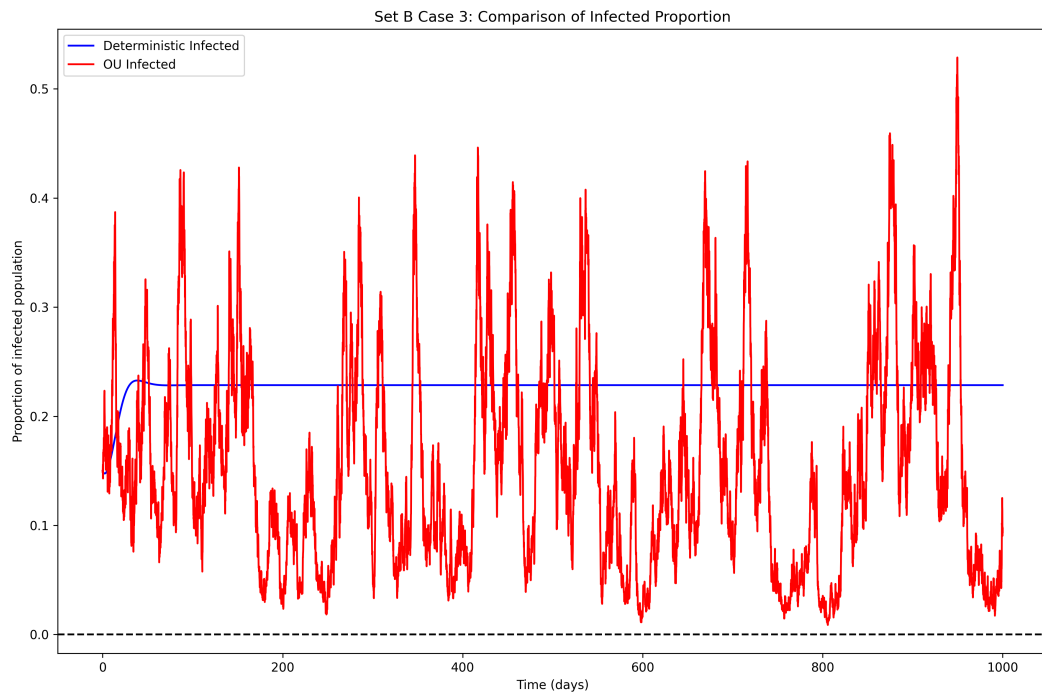


Figure A.35: A plot of the deterministic solution along with the simulated SIRS model with the OU transmission process. The data set is B and the case number is 3.

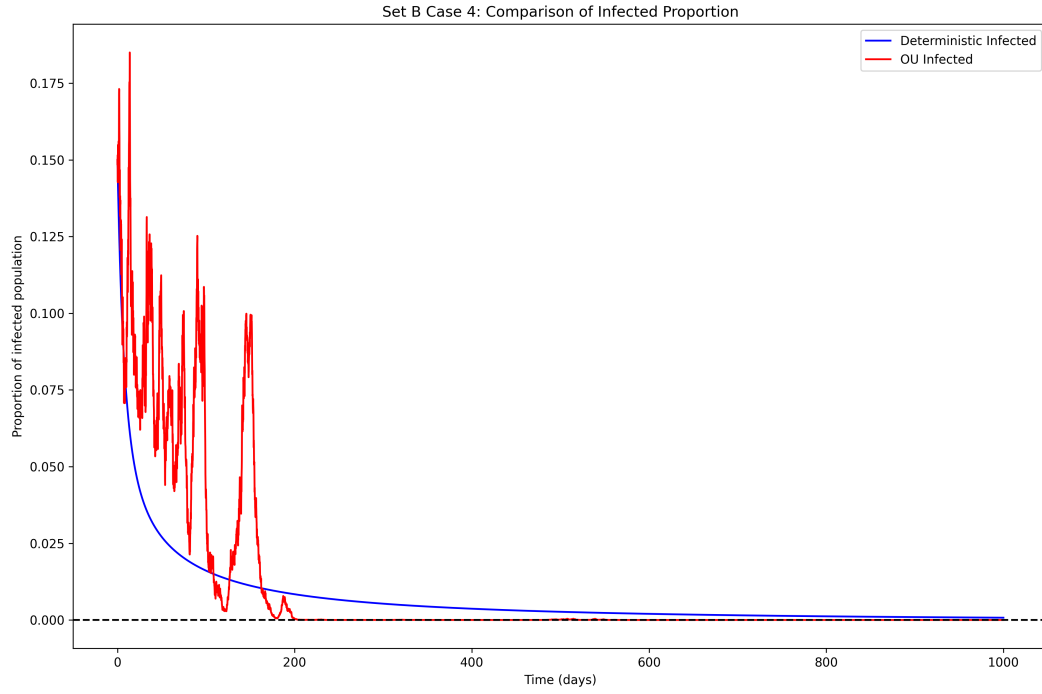


Figure A.36: A plot of the deterministic solution along with the simulated SIRS model with the OU transmission process. The data set is B and the case number is 4.

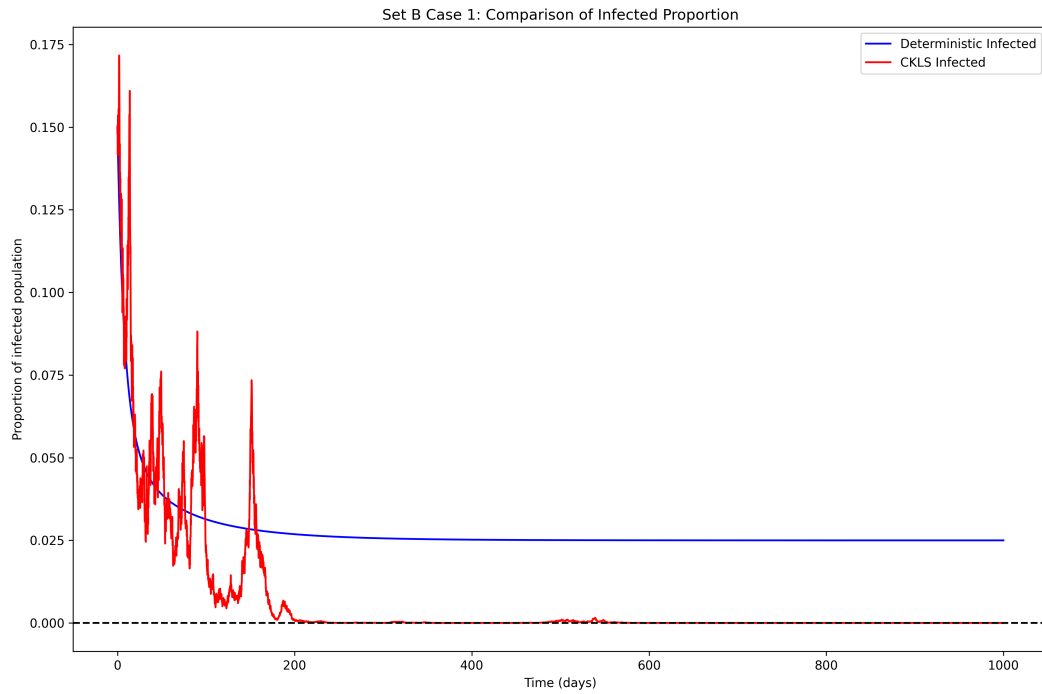


Figure A.37: A plot of the deterministic solution along with the simulated SIRS model with the CKLS transmission process. The data set is B and the case number is 1.

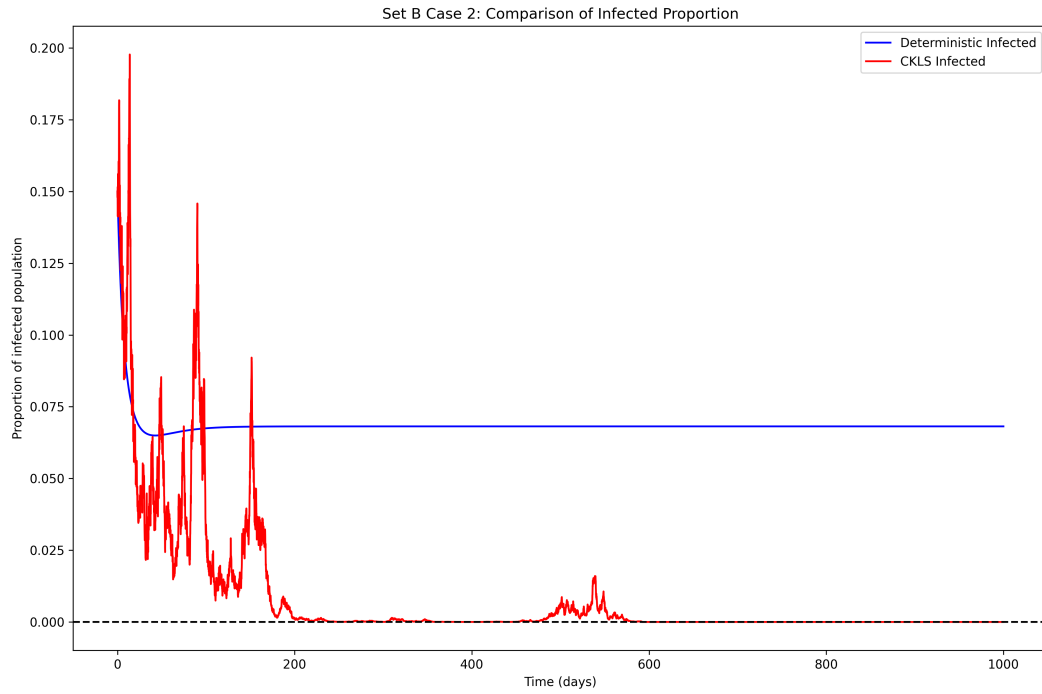


Figure A.38: A plot of the deterministic solution along with the simulated SIRS model with the CKLS transmission process. The data set is B and the case number is 2.

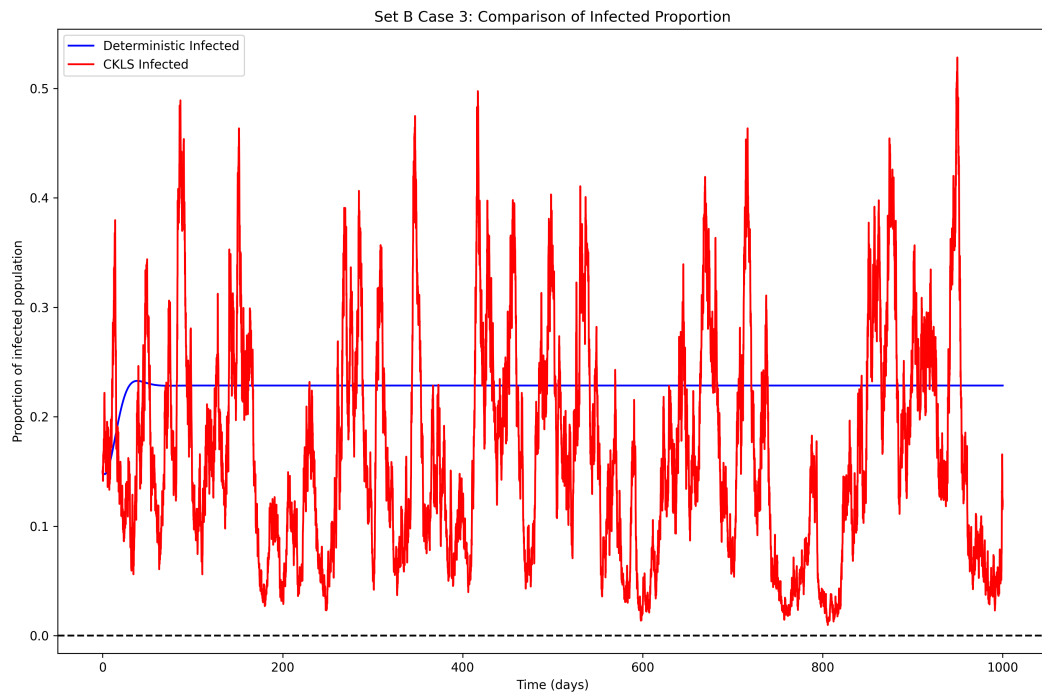


Figure A.39: A plot of the deterministic solution along with the simulated SIRS model with the CKLS transmission process. The data set is B and the case number is 3.

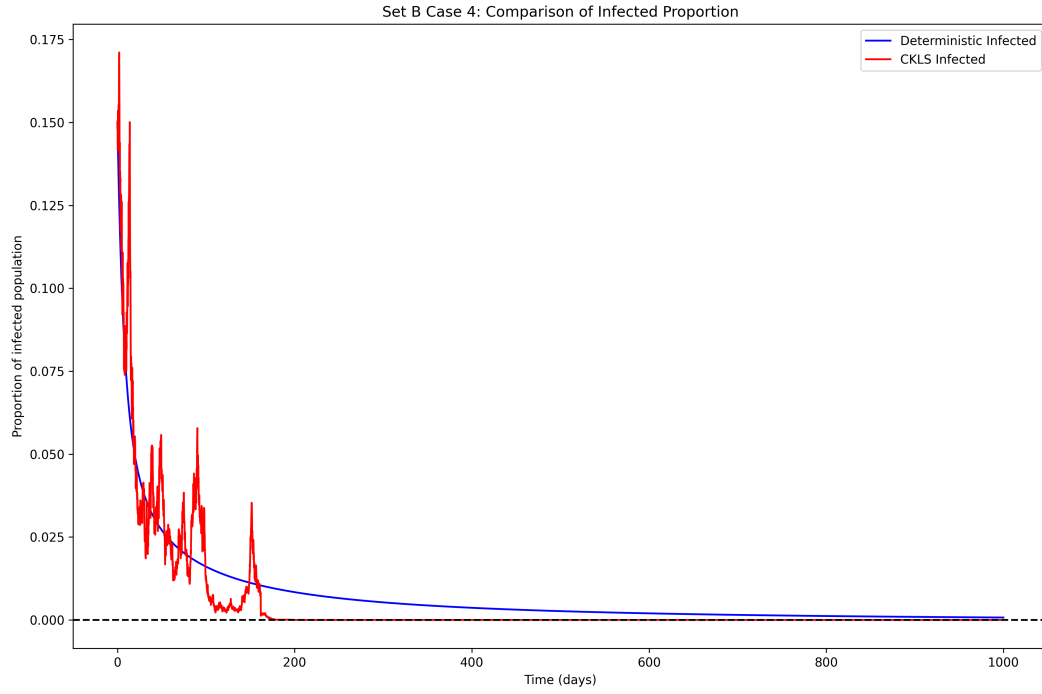


Figure A.40: A plot of the deterministic solution along with the simulated SIRS model with the CKLS transmission process. The data set is B and the case number is 4.

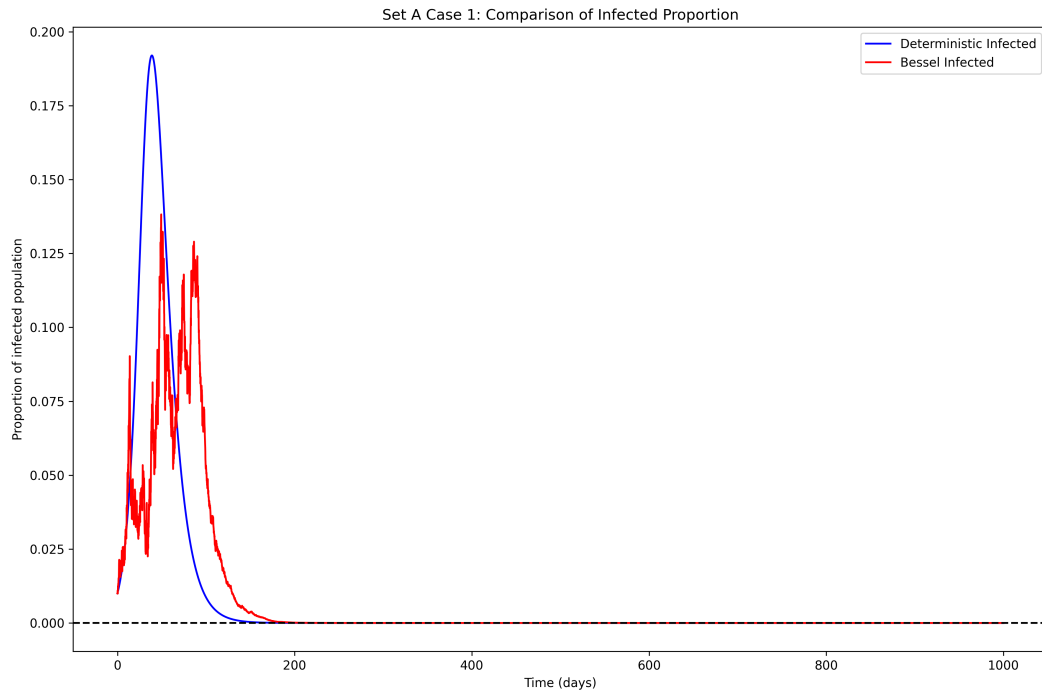


Figure A.41: A plot of the deterministic solution along with the simulated SIRS model with the Bessel transmission process. The data set is A and the case number is 1.

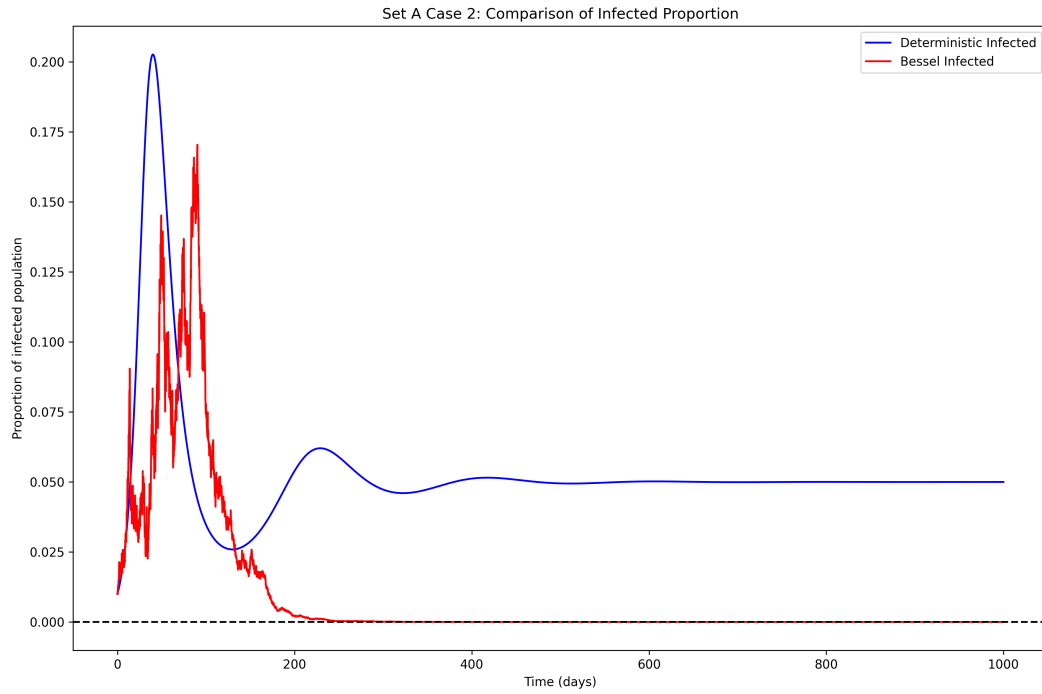


Figure A.42: A plot of the deterministic solution along with the simulated SIRS model with the Bessel transmission process. The data set is A and the case number is 2.

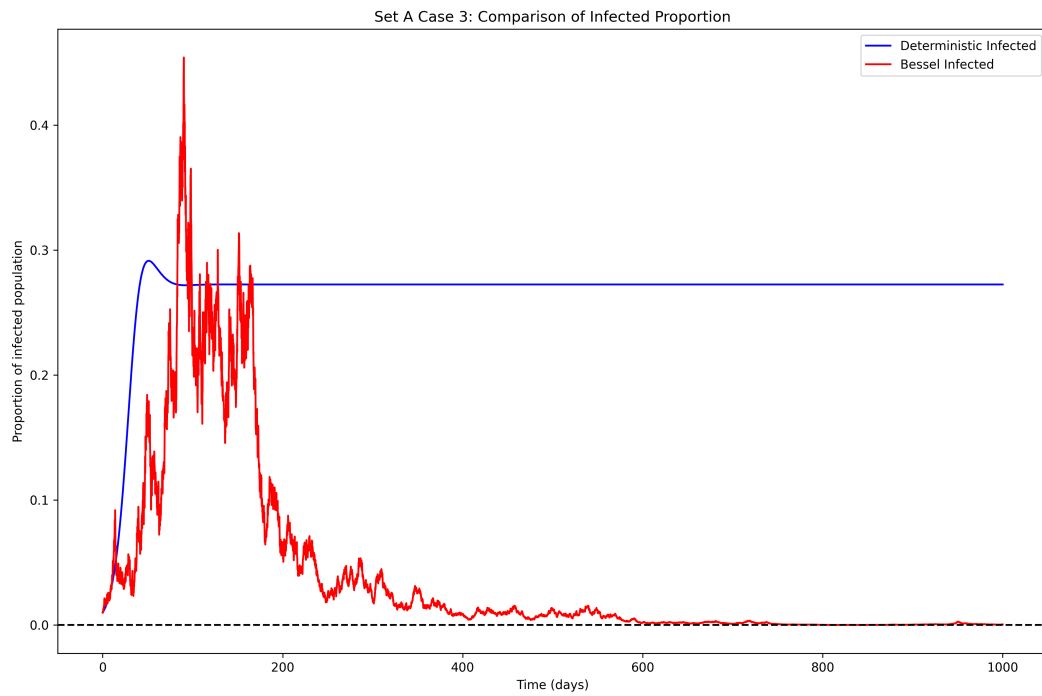


Figure A.43: A plot of the deterministic solution along with the simulated SIRS model with the Bessel transmission process. The data set is A and the case number is 3.

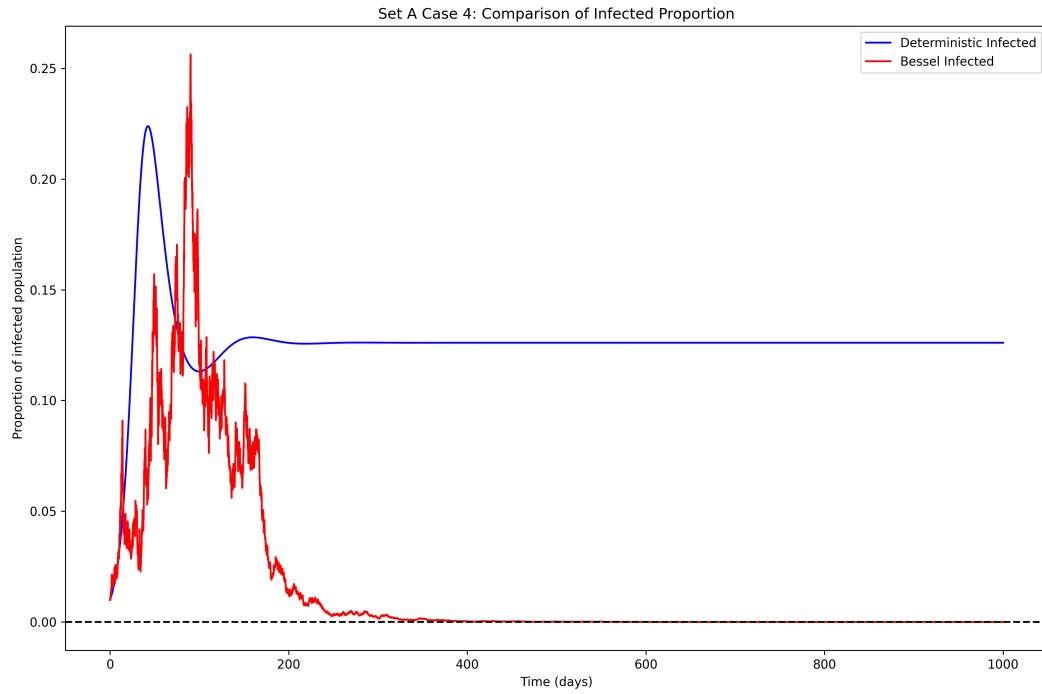


Figure A.44: A plot of the deterministic solution along with the simulated SIRS model with the Bessel transmission process. The data set is A and the case number is 4.

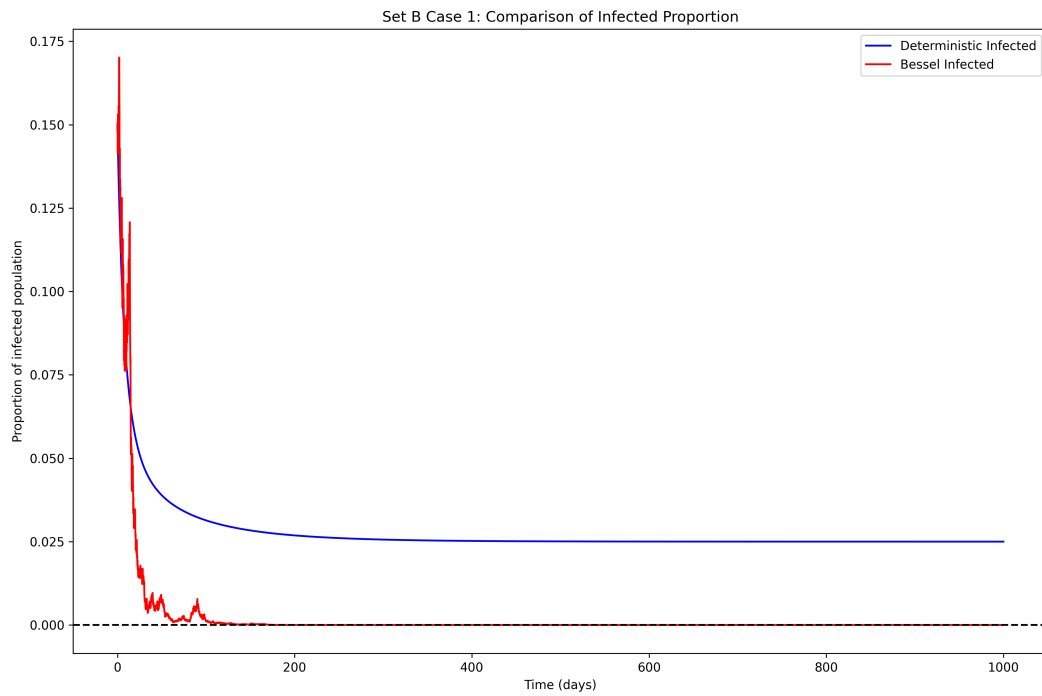


Figure A.45: A plot of the deterministic solution along with the simulated SIRS model with the Bessel transmission process. The data set is B and the case number is 1.

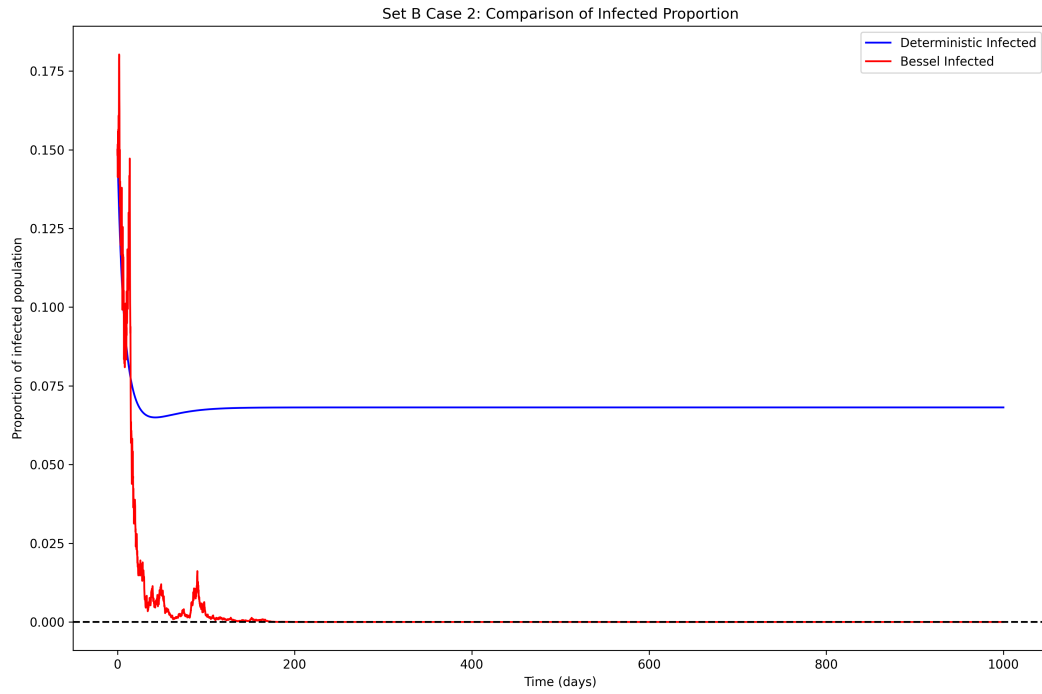


Figure A.46: A plot of the deterministic solution along with the simulated SIRS model with the Bessel transmission process. The data set is B and the case number is 2.

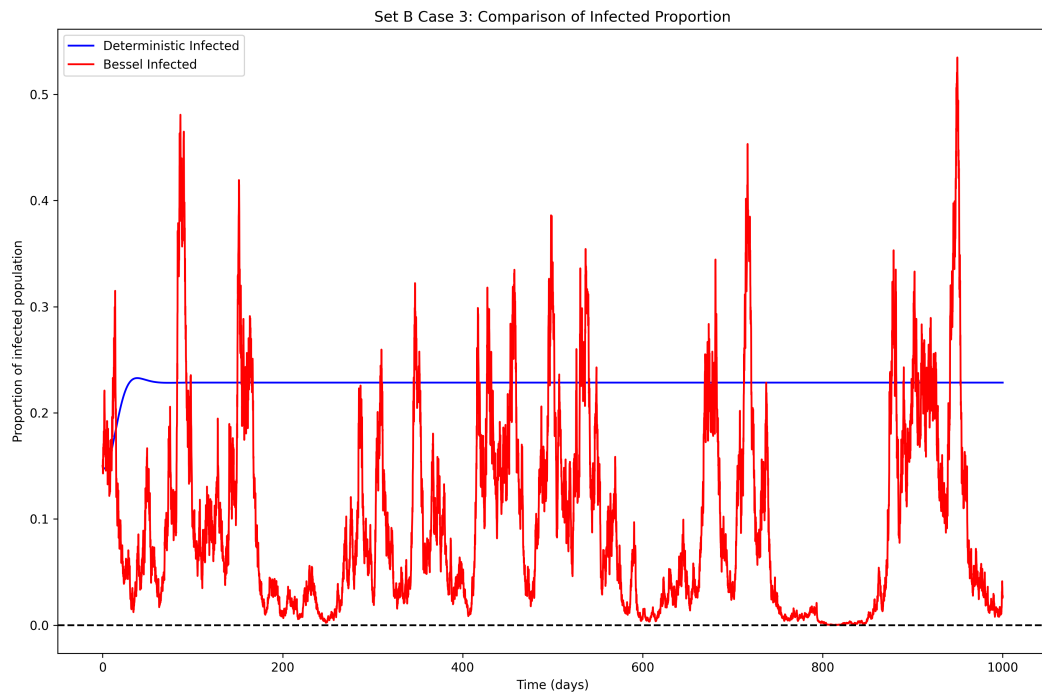


Figure A.47: A plot of the deterministic solution along with the simulated SIRS model with the Bessel transmission process. The data set is B and the case number is 3.

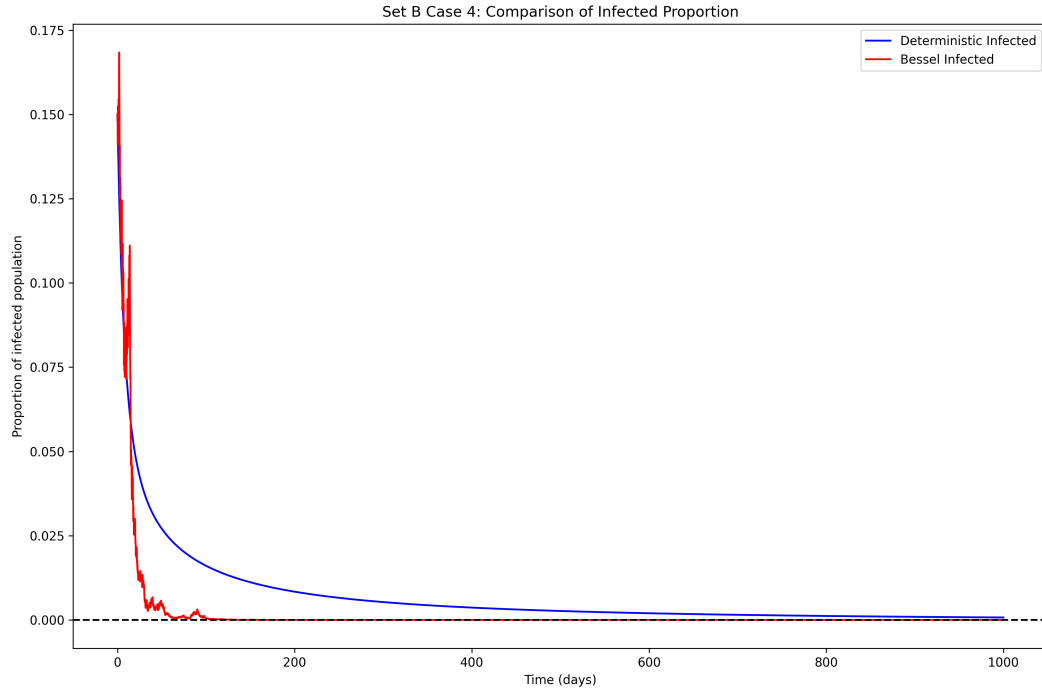


Figure A.48: A plot of the deterministic solution along with the simulated SIRS model with the Bessel transmission process. The data set is B and the case number is 4.

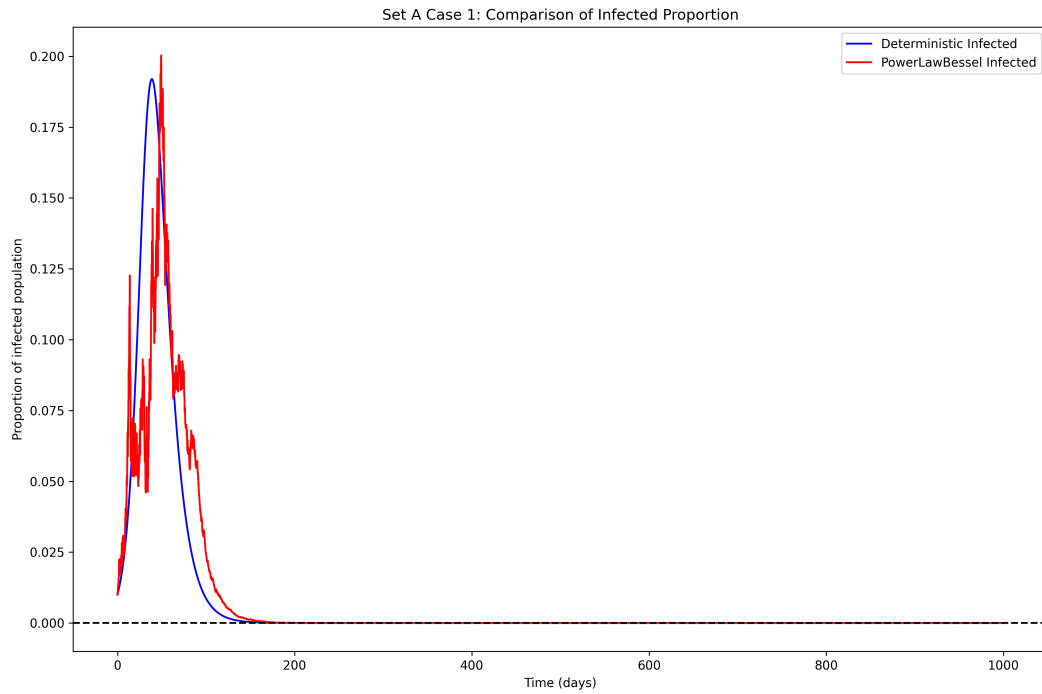


Figure A.49: A plot of the deterministic solution along with the simulated SIRS model with the PowerLawBessel transmission process. The data set is A and the case number is 1.

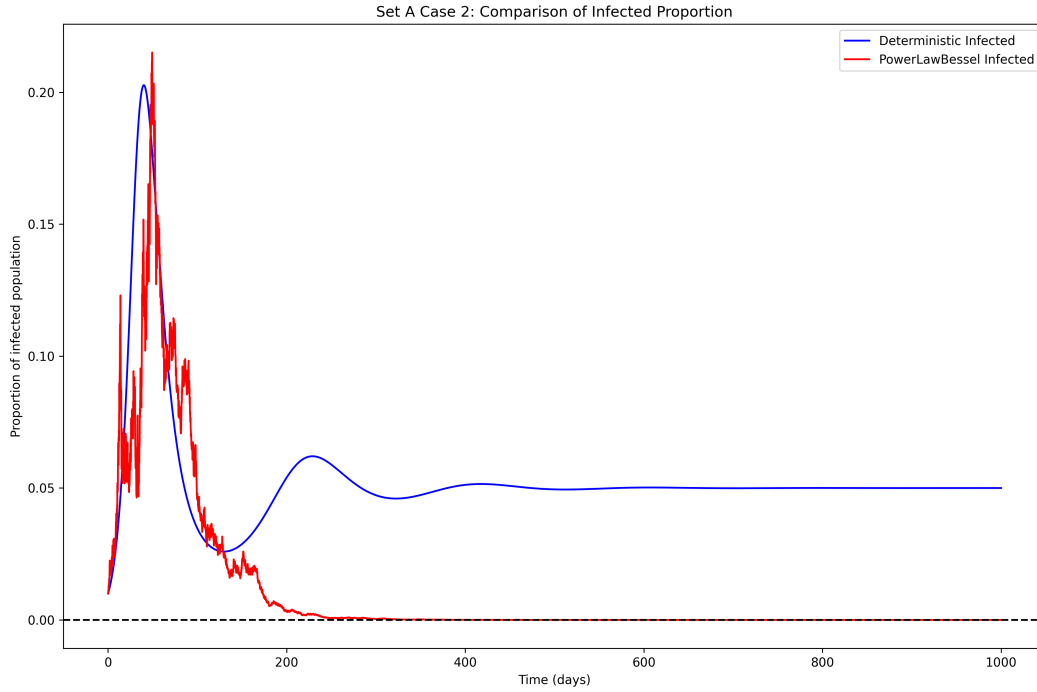


Figure A.50: A plot of the deterministic solution along with the simulated SIRS model with the PowerLawBessel transmission process. The data set is A and the case number is 2.

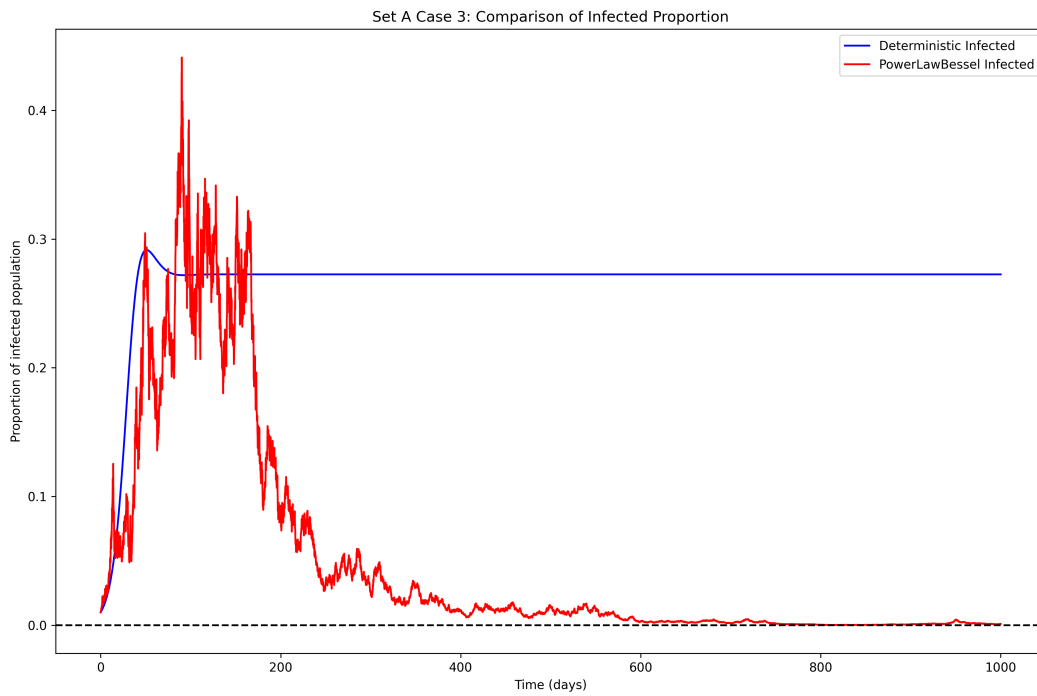


Figure A.51: A plot of the deterministic solution along with the simulated SIRS model with the PowerLawBessel transmission process. The data set is A and the case number is 3.

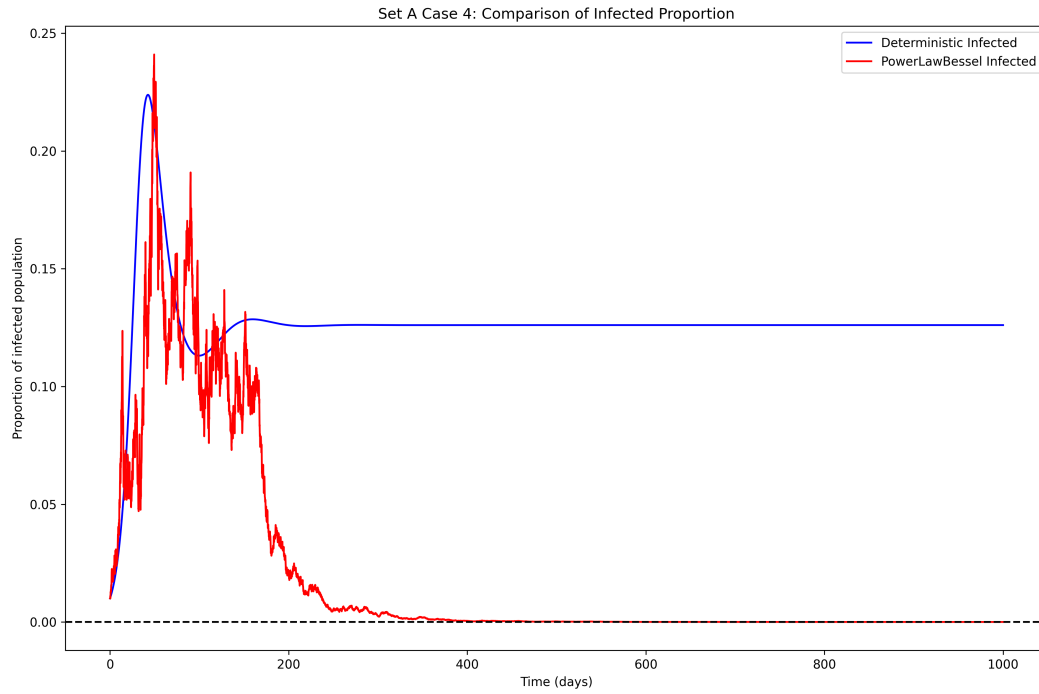


Figure A.52: A plot of the deterministic solution along with the simulated SIRS model with the PowerLawBessel transmission process. The data set is A and the case number is 4.

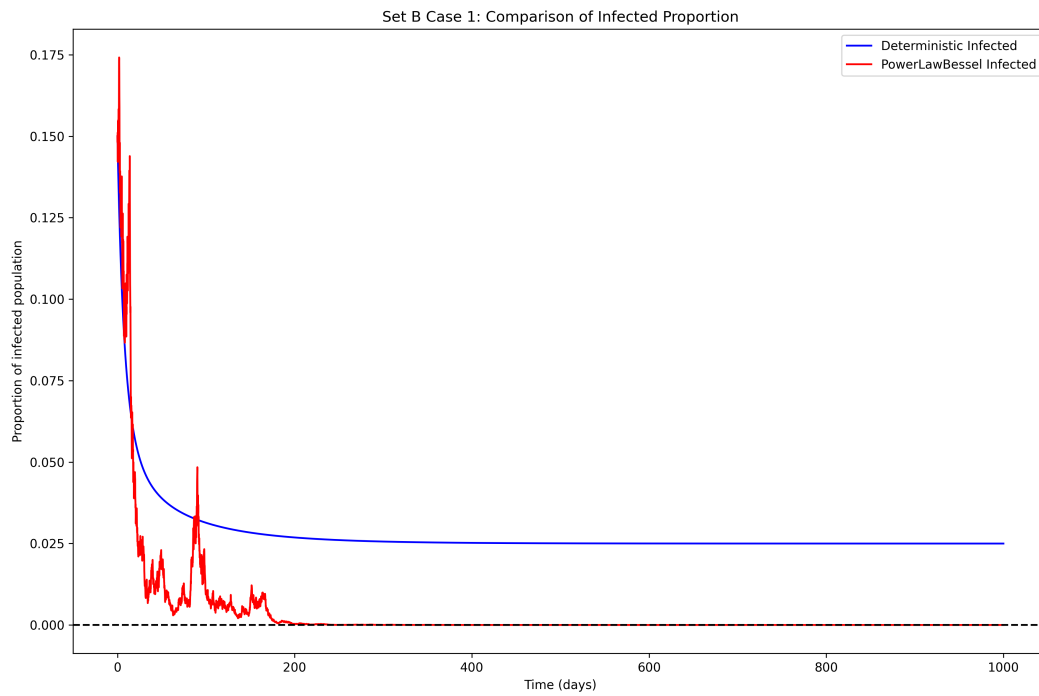


Figure A.53: A plot of the deterministic solution along with the simulated SIRS model with the PowerLawBessel transmission process. The data set is B and the case number is 1.

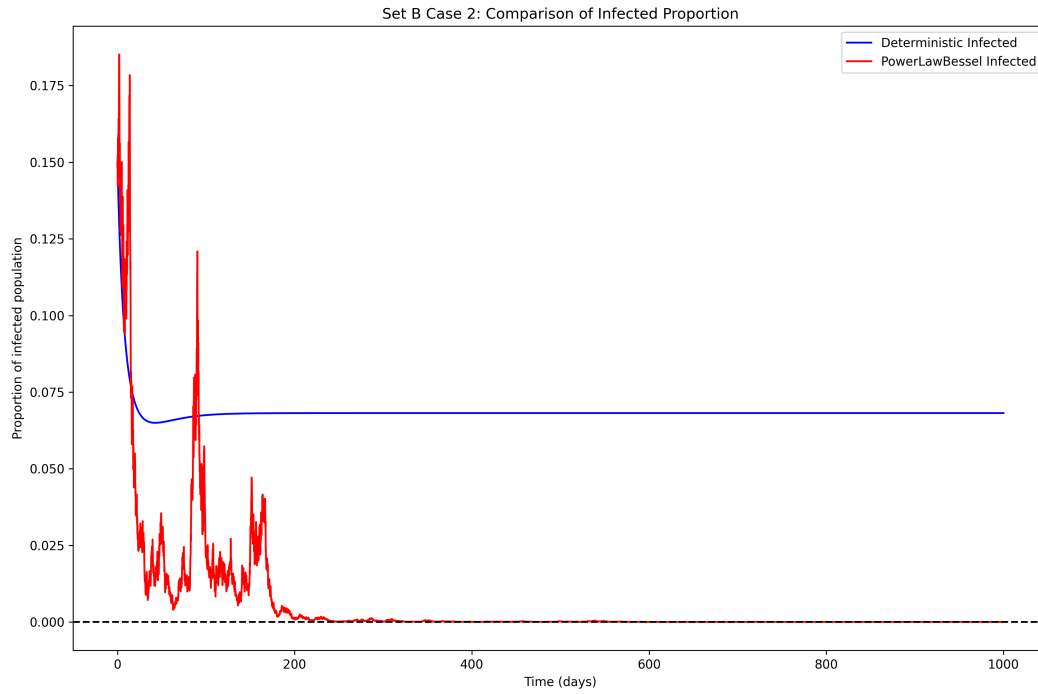


Figure A.54: A plot of the deterministic solution along with the simulated SIRS model with the PowerLawBessel transmission process. The data set is B and the case number is 2.

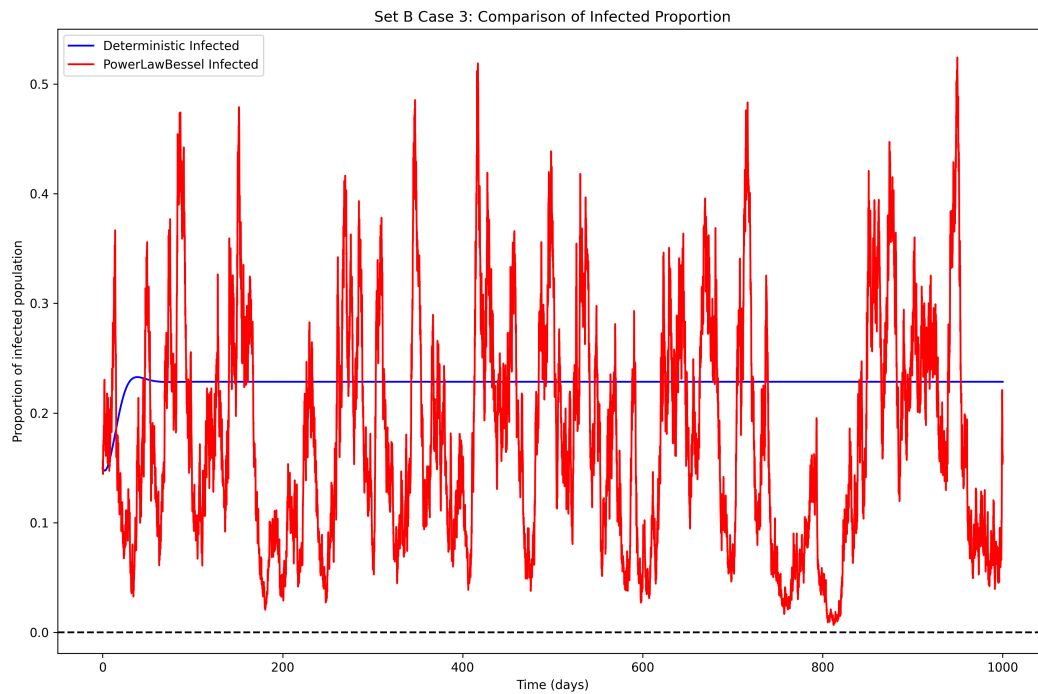


Figure A.55: A plot of the deterministic solution along with the simulated SIRS model with the PowerLawBessel transmission process. The data set is B and the case number is 3.

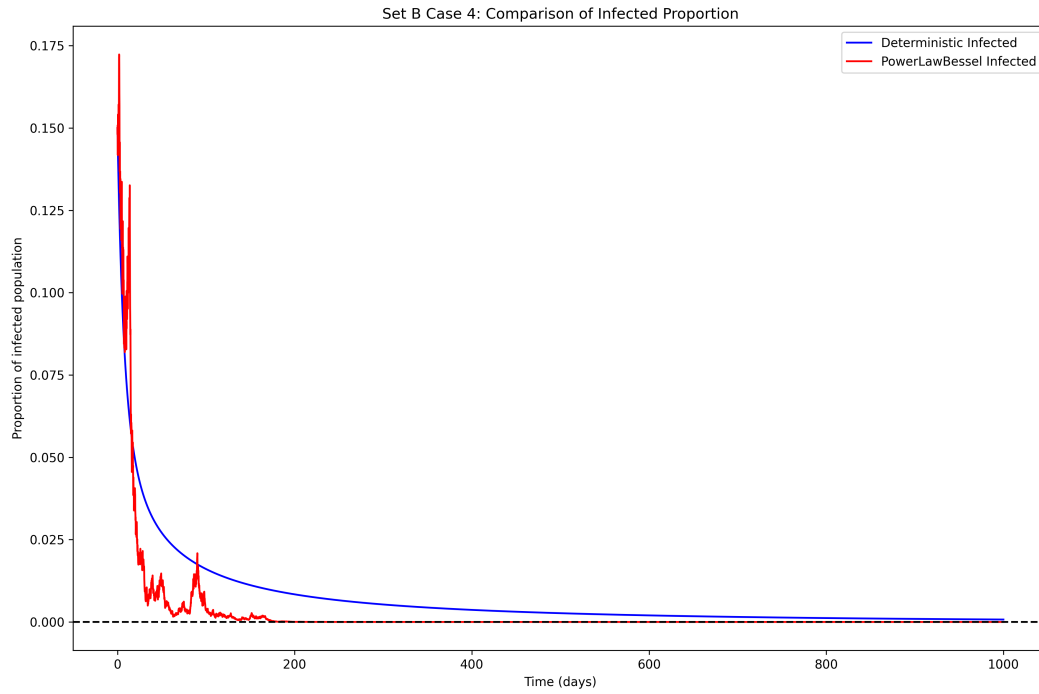


Figure A.56: A plot of the deterministic solution along with the simulated SIRS model with the PowerLawBessel transmission process. The data set is B and the case number is 4.

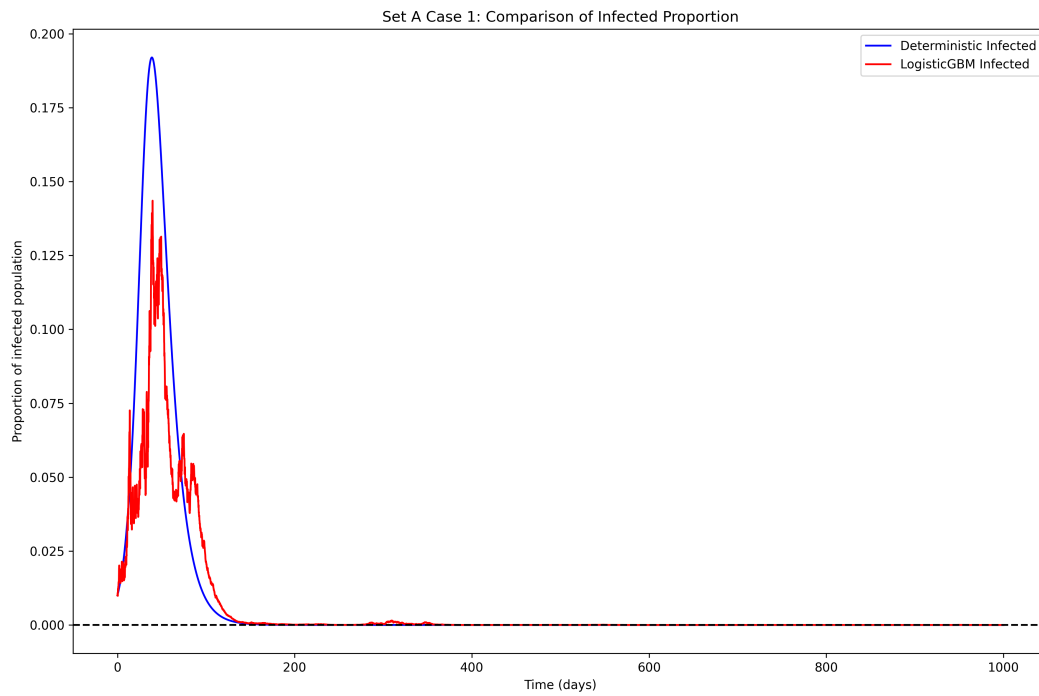


Figure A.57: A plot of the deterministic solution along with the simulated SIRS model with the LogisticGBM transmission process. The data set is A and the case number is 1.

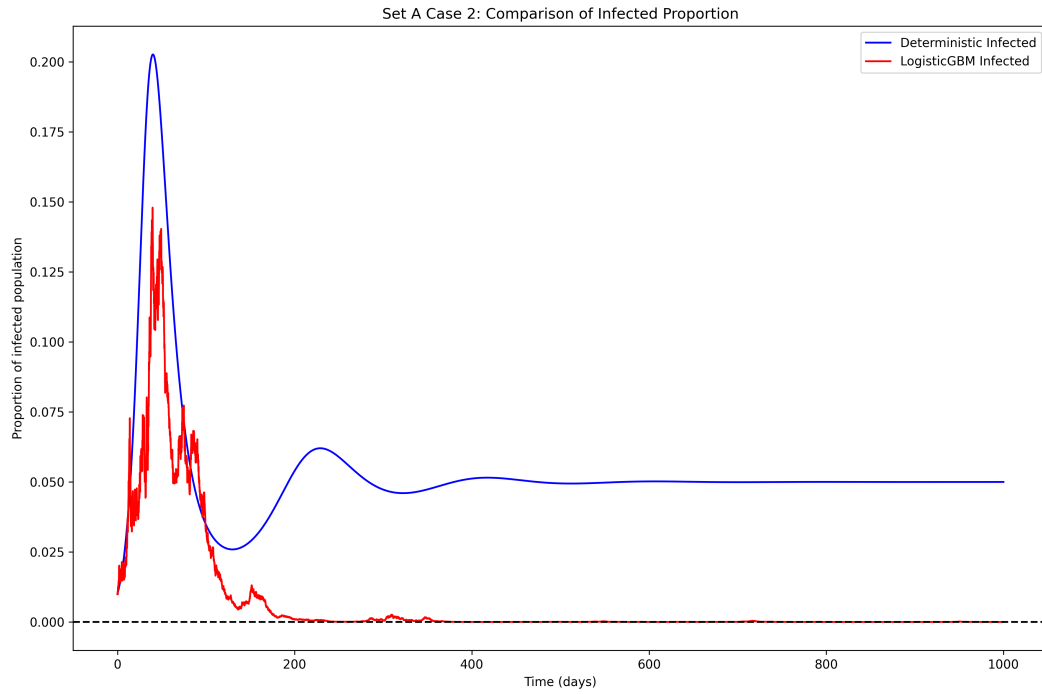


Figure A.58: A plot of the deterministic solution along with the simulated SIRS model with the LogisticGBM transmission process. The data set is A and the case number is 2.

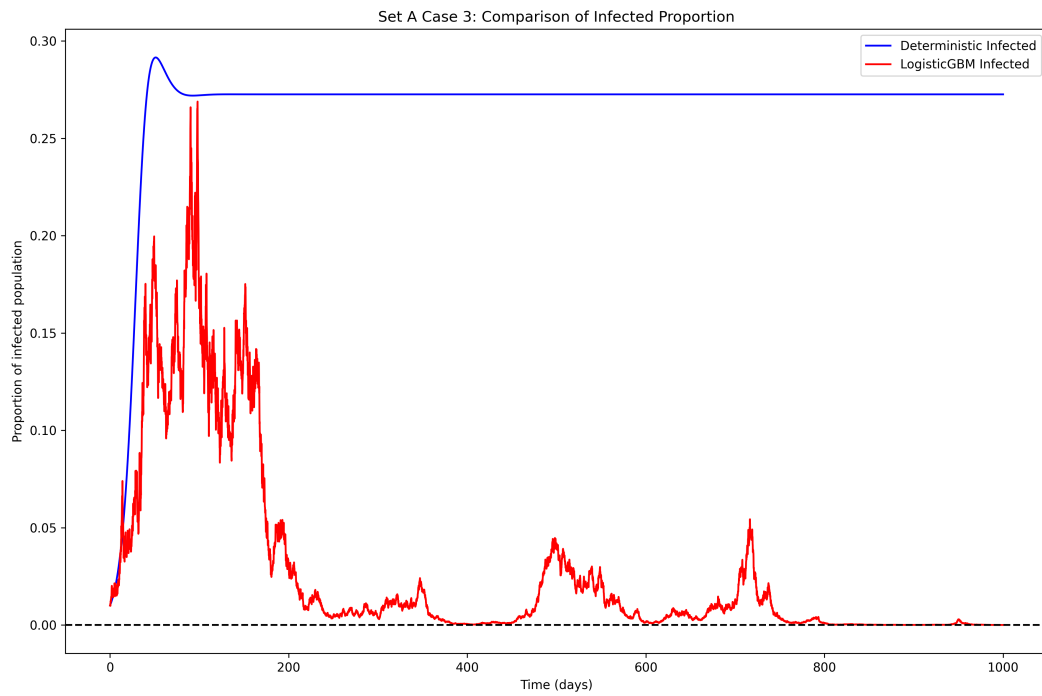


Figure A.59: A plot of the deterministic solution along with the simulated SIRS model with the LogisticGBM transmission process. The data set is A and the case number is 3.

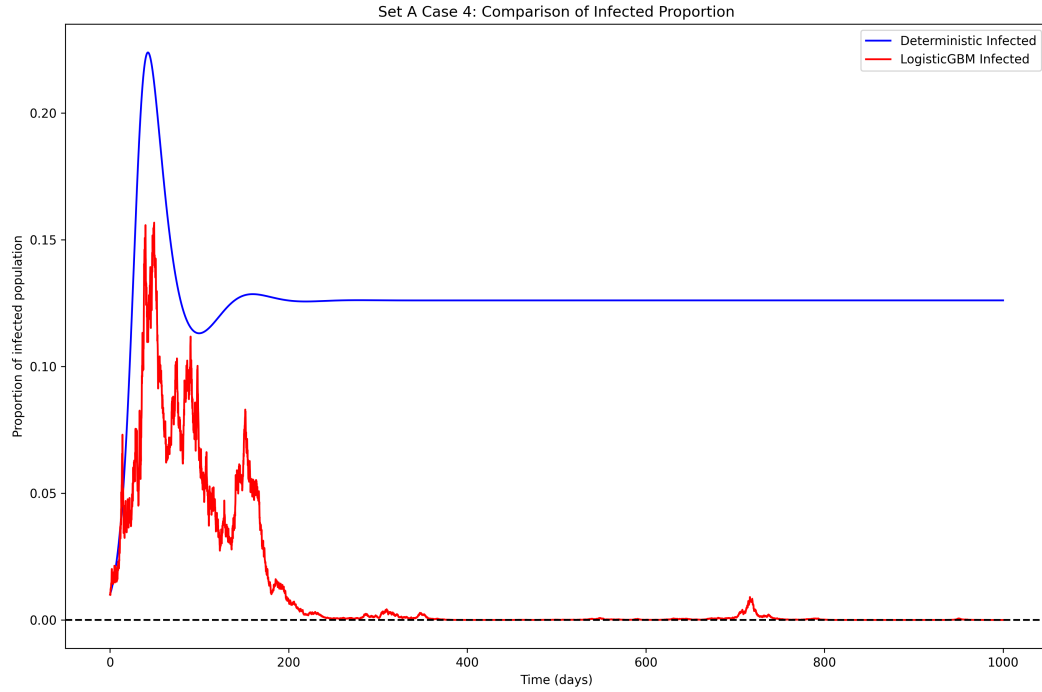


Figure A.60: A plot of the deterministic solution along with the simulated SIRS model with the LogisticGBM transmission process. The data set is A and the case number is 4.

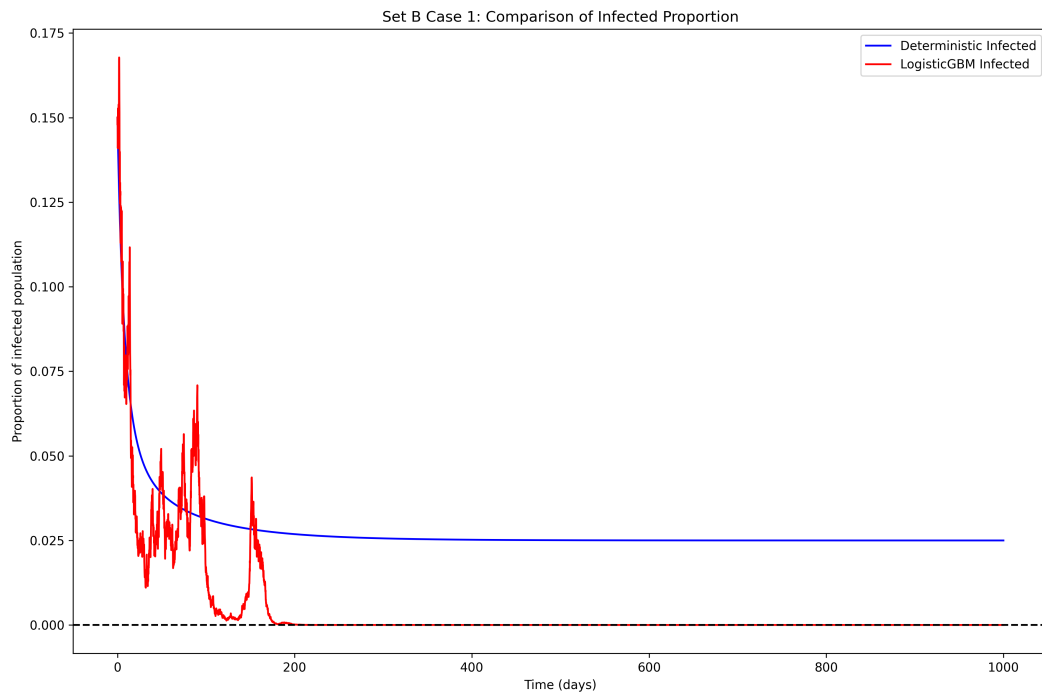


Figure A.61: A plot of the deterministic solution along with the simulated SIRS model with the LogisticGBM transmission process. The data set is B and the case number is 1.

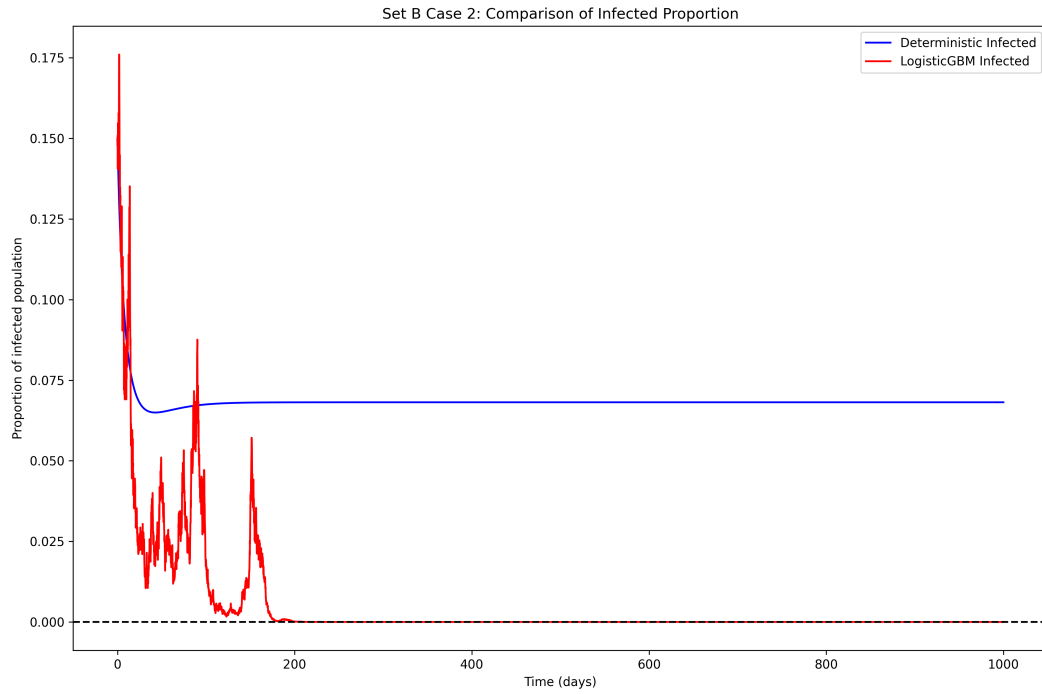


Figure A.62: A plot of the deterministic solution along with the simulated SIRS model with the LogisticGBM transmission process. The data set is B and the case number is 2.

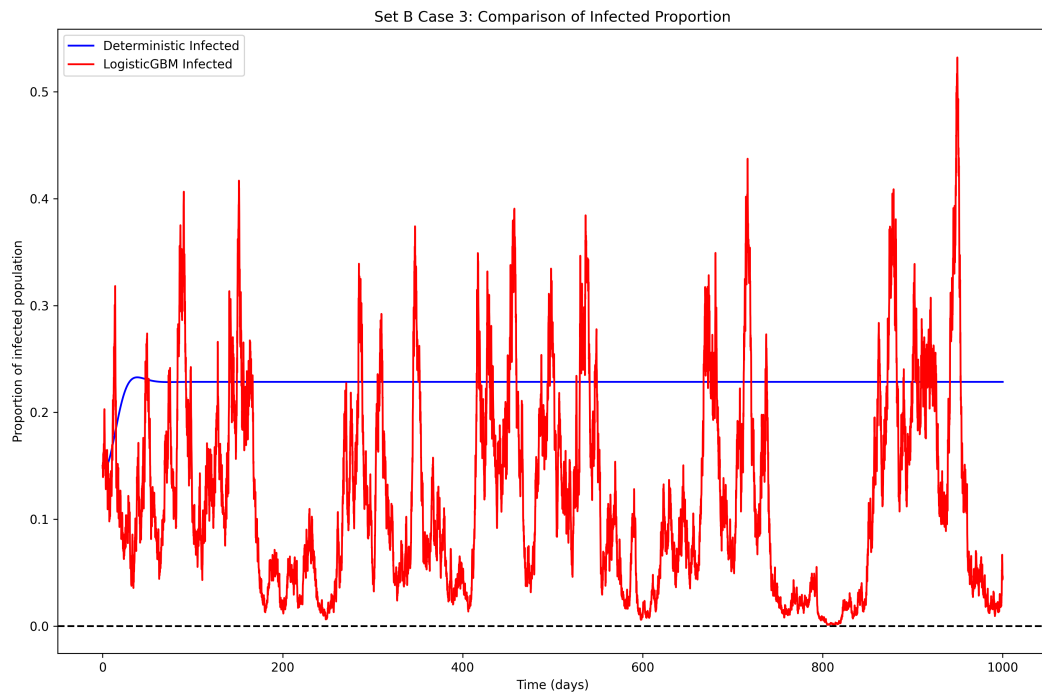


Figure A.63: A plot of the deterministic solution along with the simulated SIRS model with the LogisticGBM transmission process. The data set is B and the case number is 3.

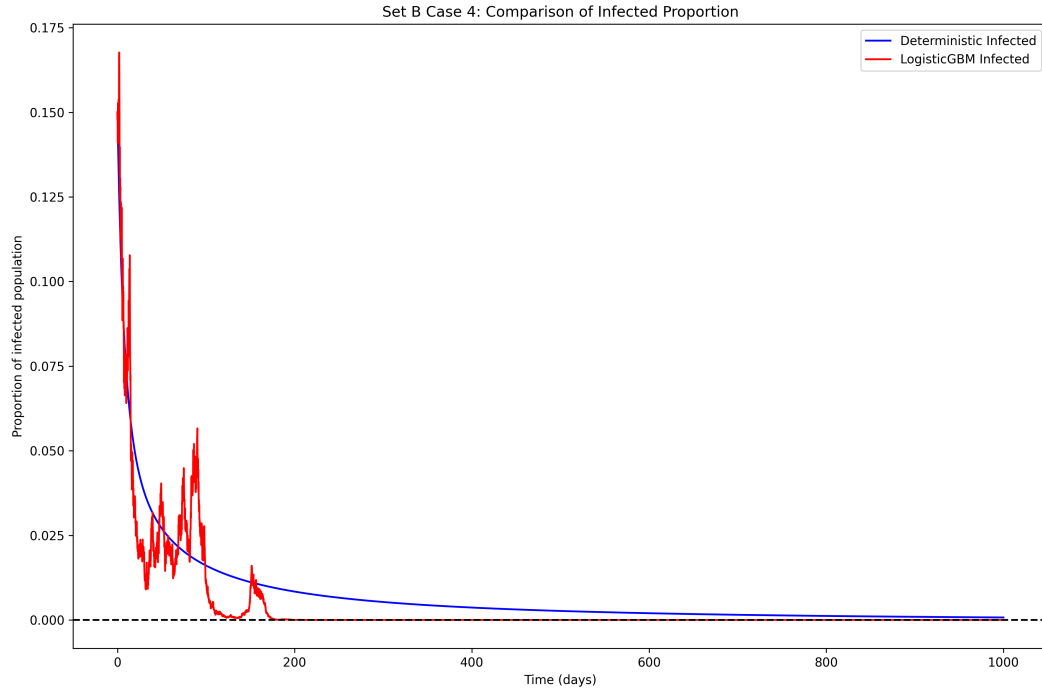


Figure A.64: A plot of the deterministic solution along with the simulated SIRS model with the LogisticGBM transmission process. The data set is B and the case number is 4.

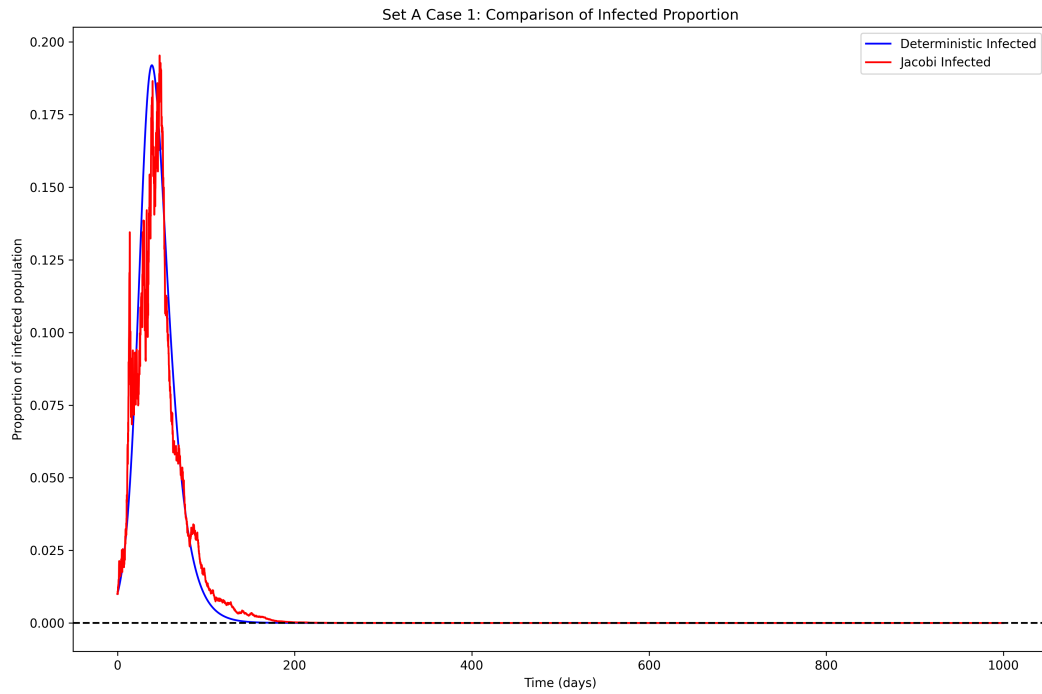


Figure A.65: A plot of the deterministic solution along with the simulated SIRS model with the Jacobi transmission process. The data set is A and the case number is 1.

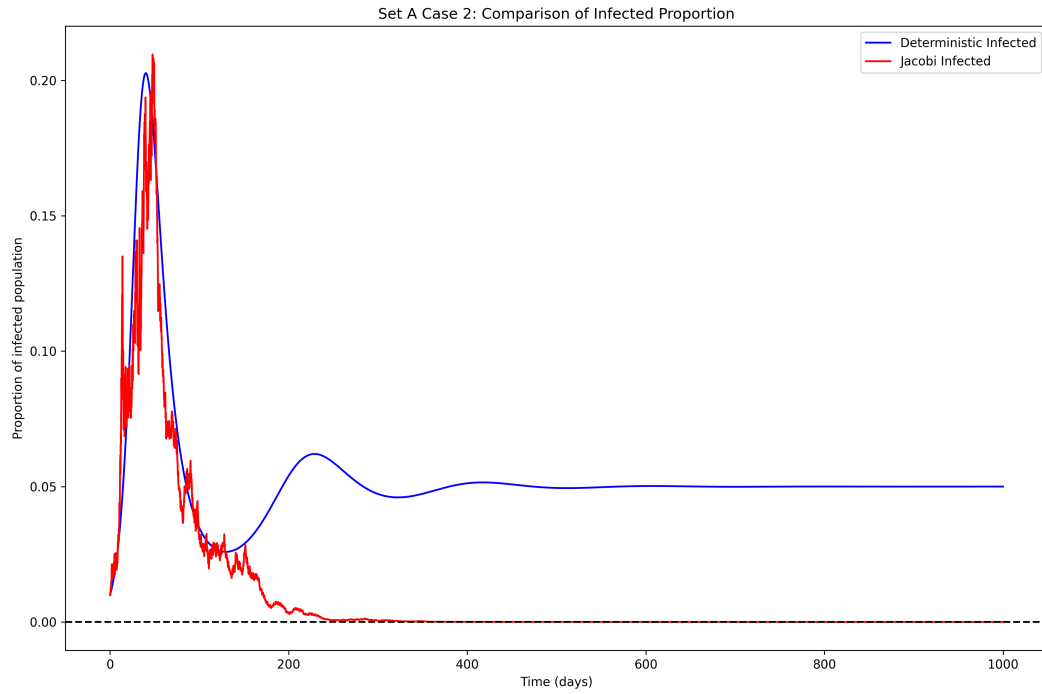


Figure A.66: A plot of the deterministic solution along with the simulated SIRS model with the Jacobi transmission process. The data set is A and the case number is 2.

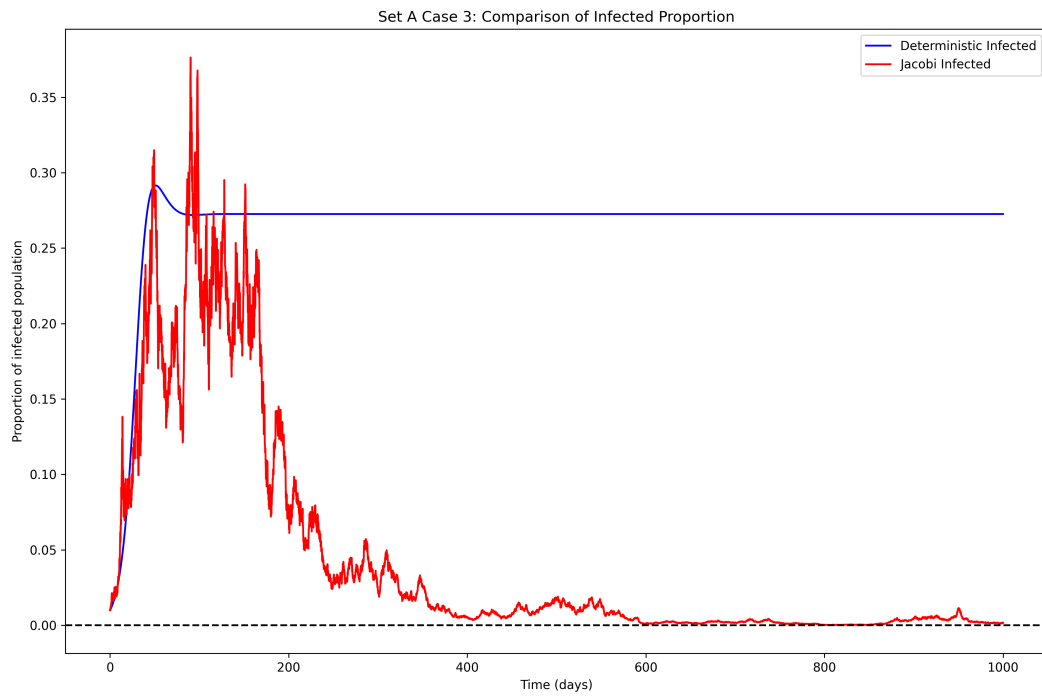


Figure A.67: A plot of the deterministic solution along with the simulated SIRS model with the Jacobi transmission process. The data set is A and the case number is 3.

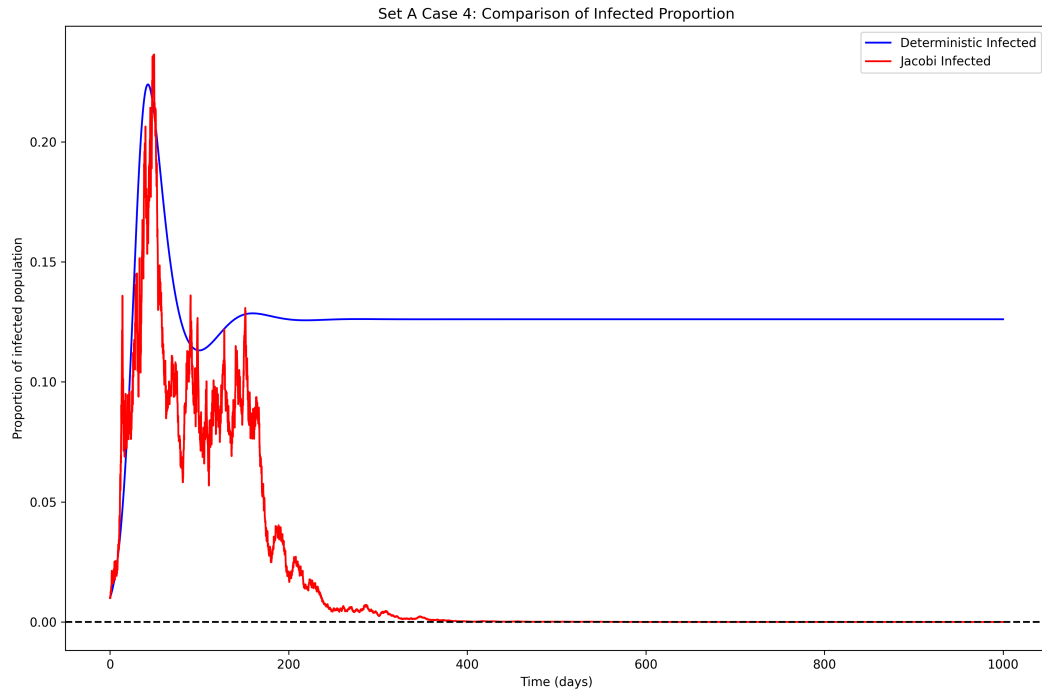


Figure A.68: A plot of the deterministic solution along with the simulated SIRS model with the Jacobi transmission process. The data set is A and the case number is 4.

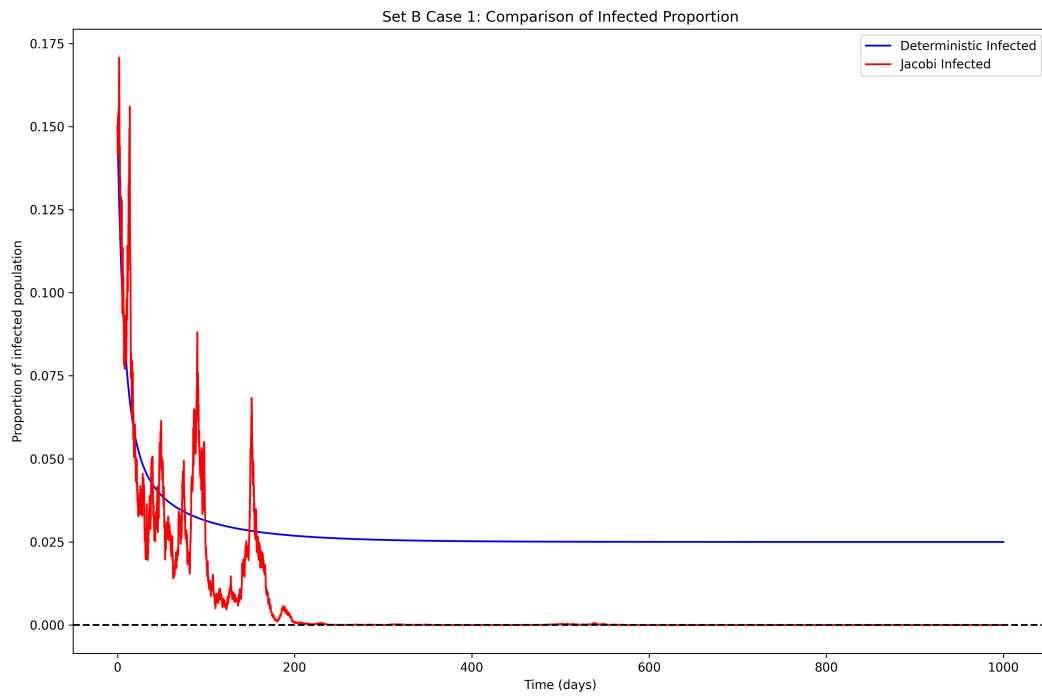


Figure A.69: A plot of the deterministic solution along with the simulated SIRS model with the Jacobi transmission process. The data set is B and the case number is 1.

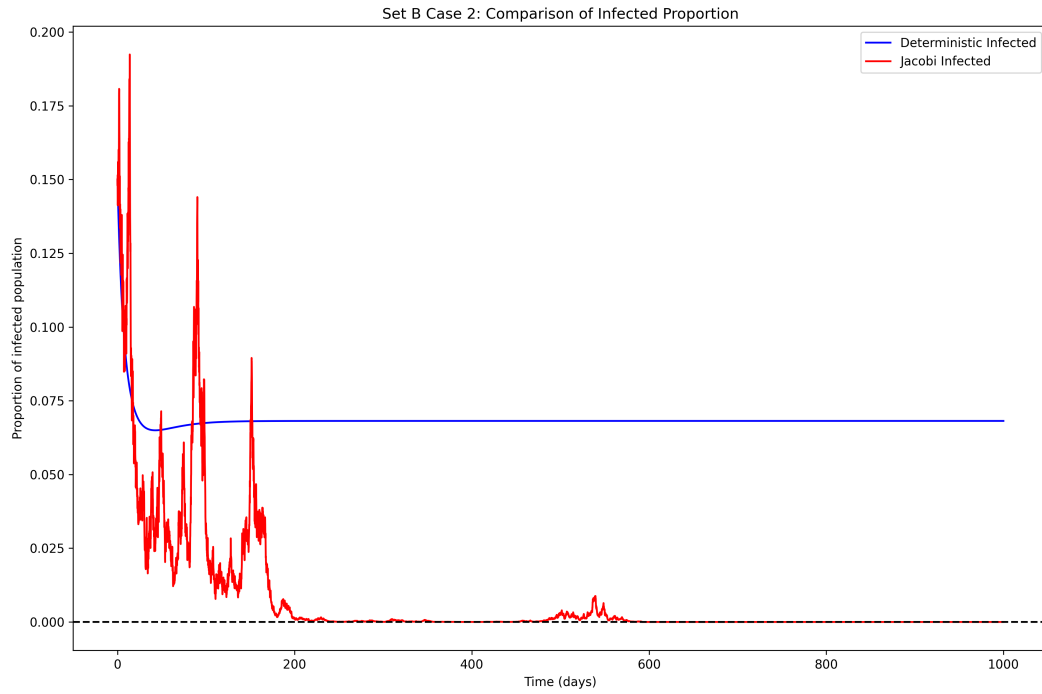


Figure A.70: A plot of the deterministic solution along with the simulated SIRS model with the Jacobi transmission process. The data set is B and the case number is 2.

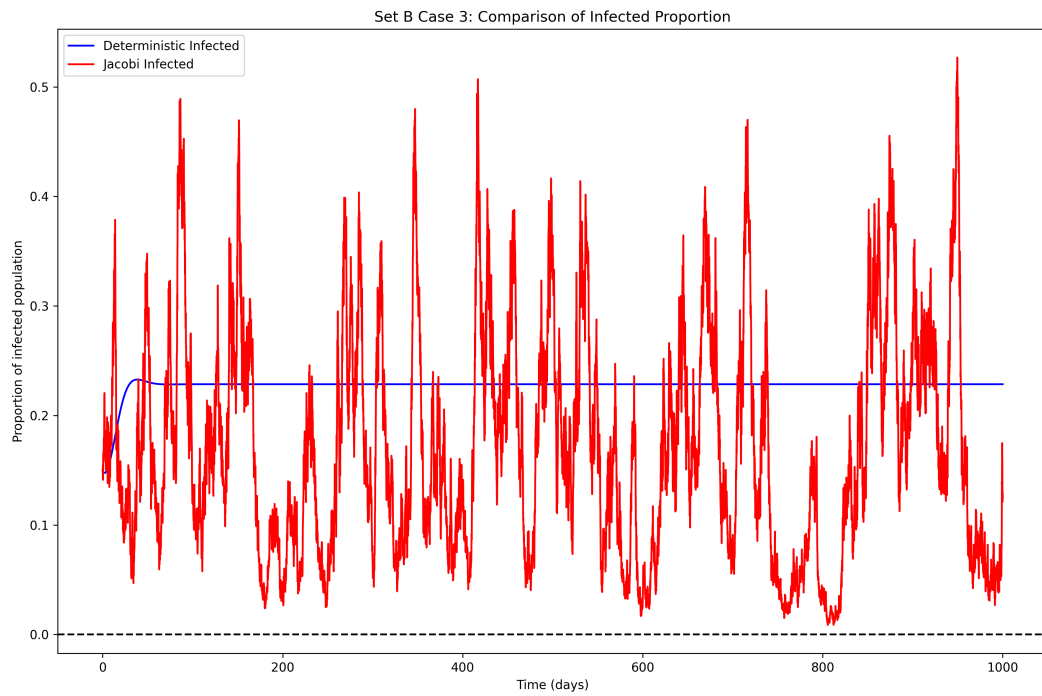


Figure A.71: A plot of the deterministic solution along with the simulated SIRS model with the Jacobi transmission process. The data set is B and the case number is 3.

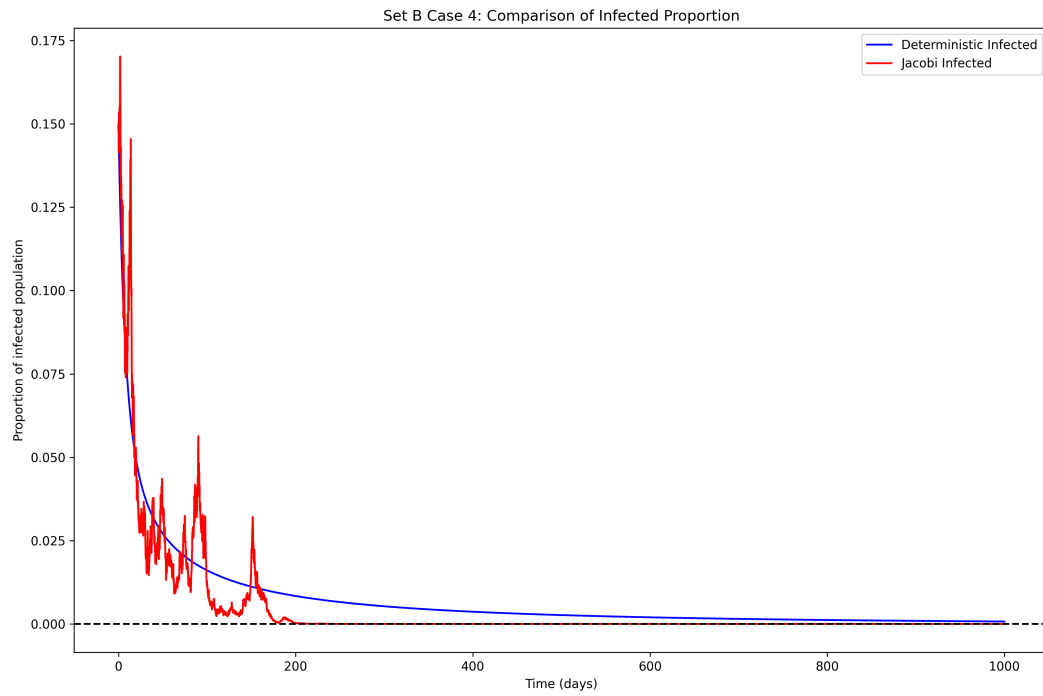


Figure A.72: A plot of the deterministic solution along with the simulated SIRS model with the Jacobi transmission process. The data set is B and the case number is 4.

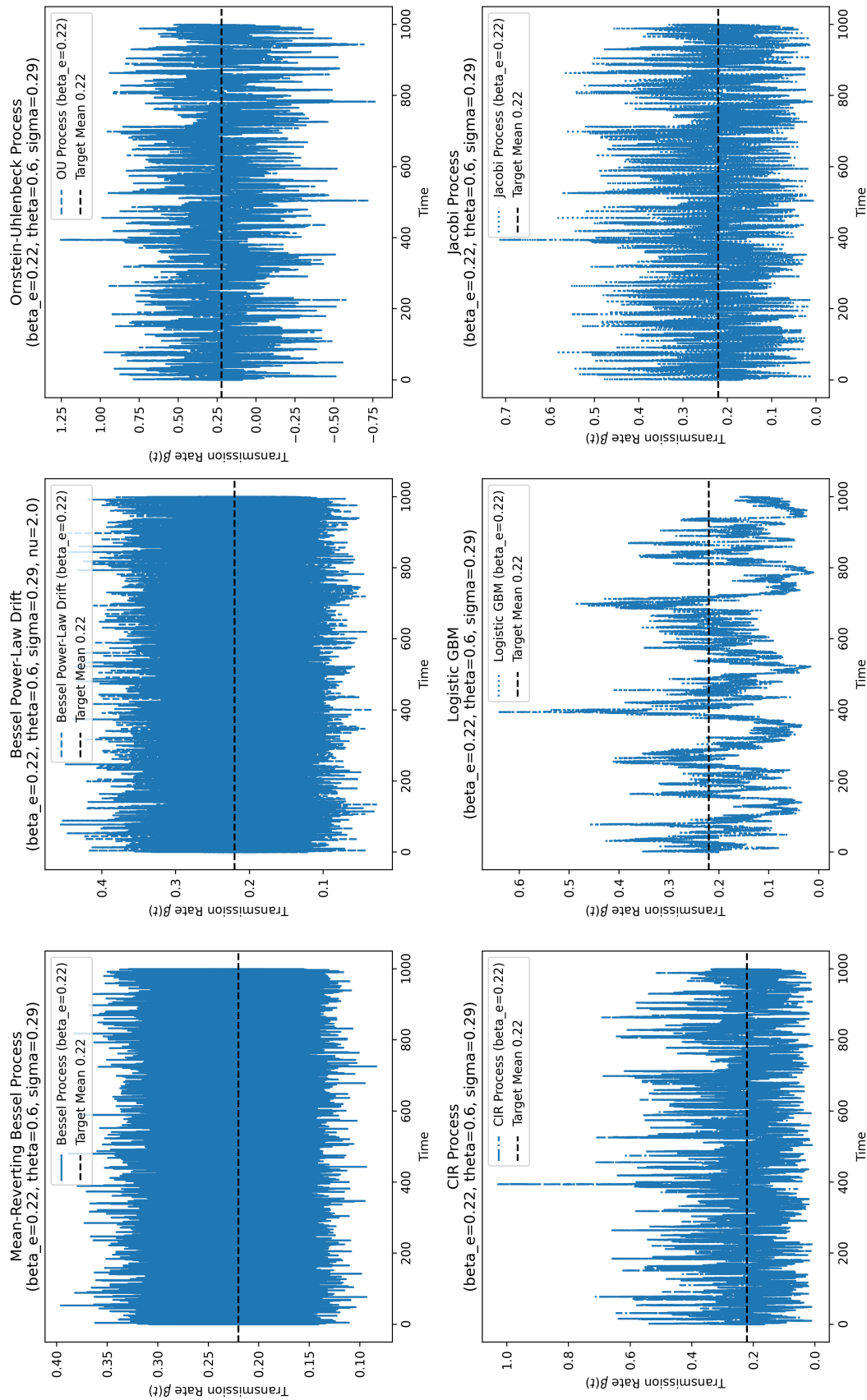


Figure A.73: Plots of the transmission processes for Set B, Case 2. Top row, from left to right: mean-reverting Bessel process, mean-reverting power law Bessel process, OU process; Bottom row, from left to right: CKLS process, Logistic GBM process, Jacobi process.

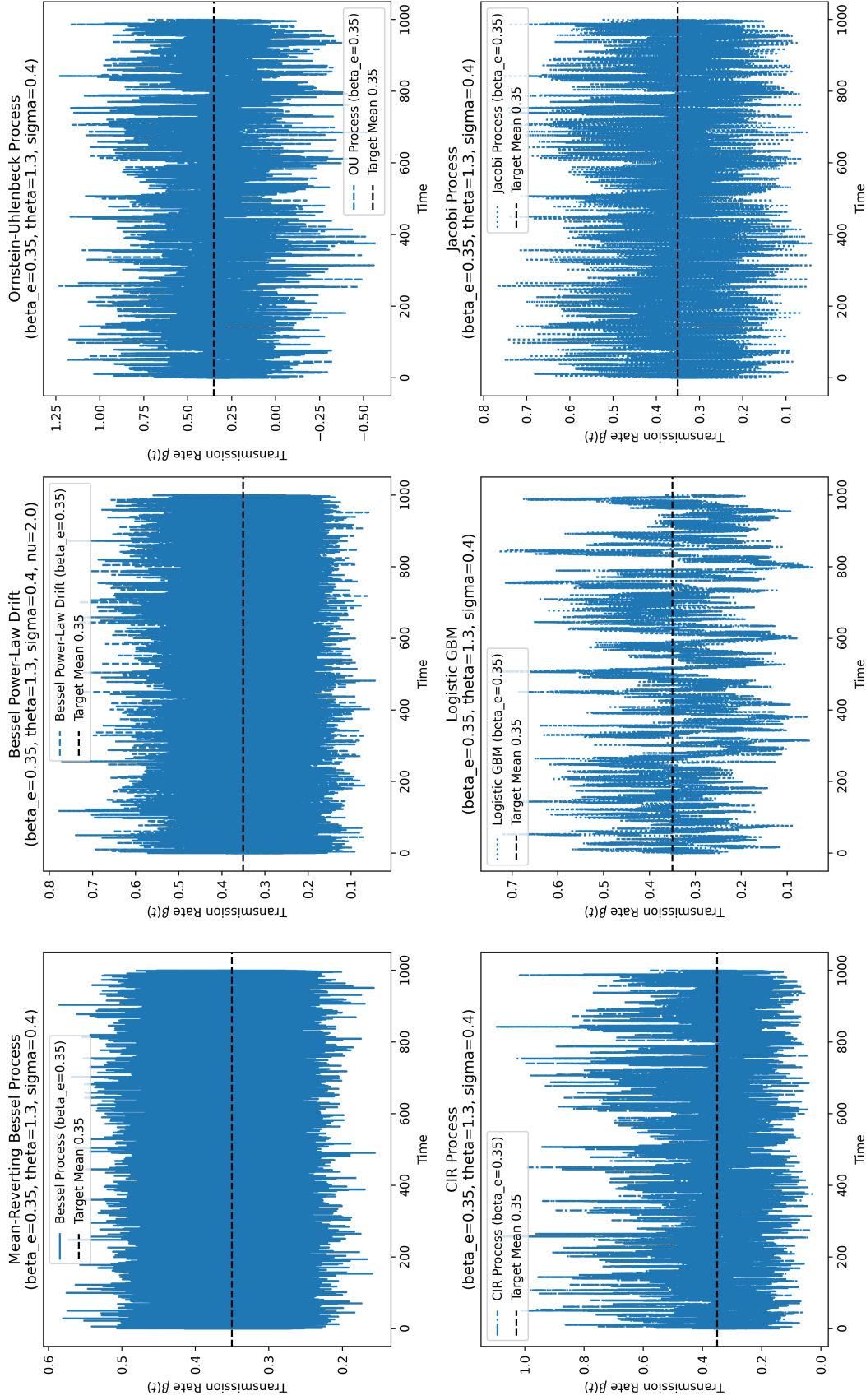


Figure A.74: Plots of the transmission processes for Set B, Case 3. Top row, from left to right: mean-reverting Bessel process, mean-reverting power law Bessel process, OU process; Bottom row, from left to right: CKLS process, Logistic GBM process, Jacobi process.

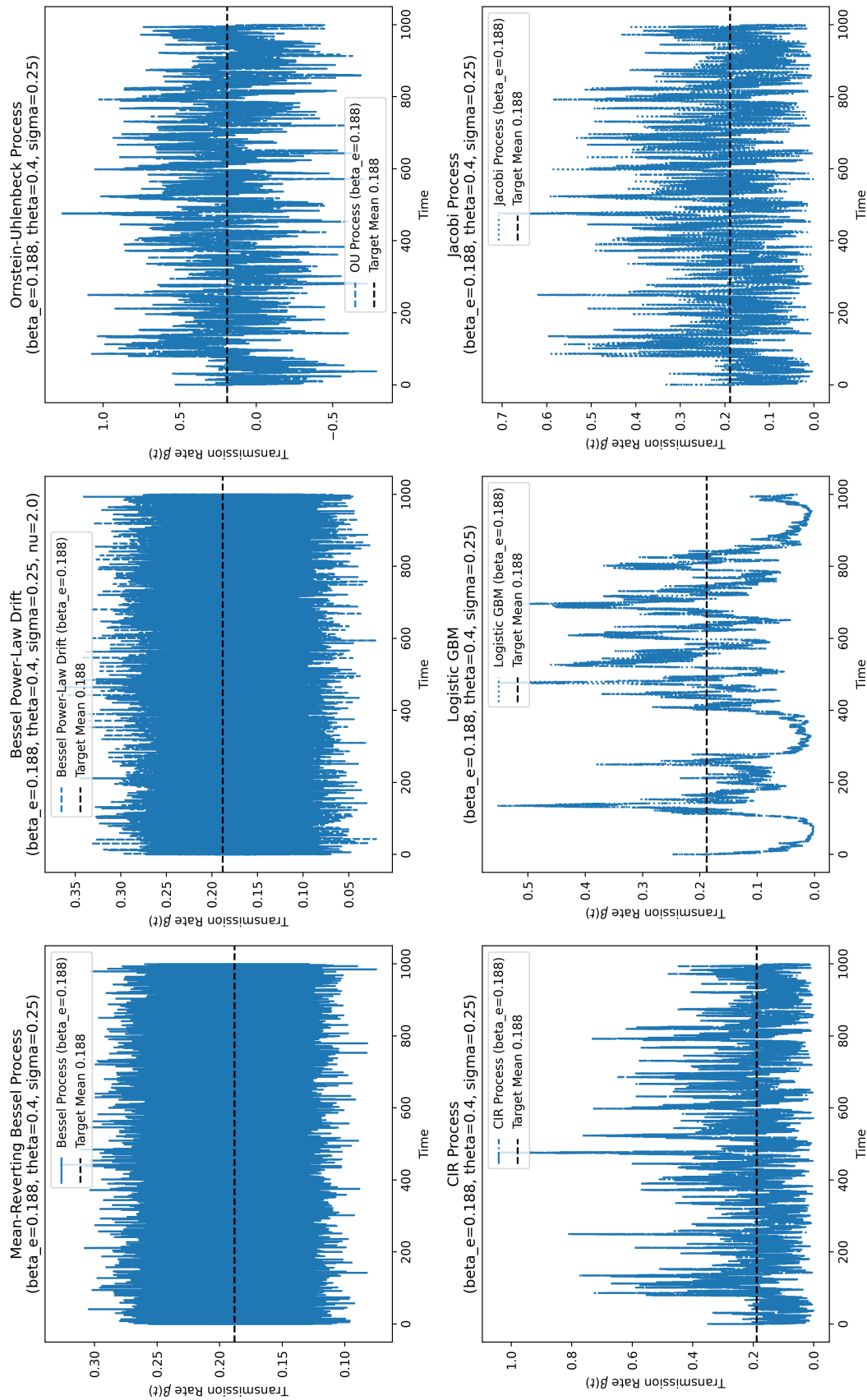


Figure A.75: Plots of the transmission processes for Set A. From left to right for Set B, Case 4. Top row, from left to right: mean-reverting Bessel process, mean-reverting power law Bessel process, OU process; Bottom row, from left to right: CKLS process, Logistic GBM process, Jacobi process.

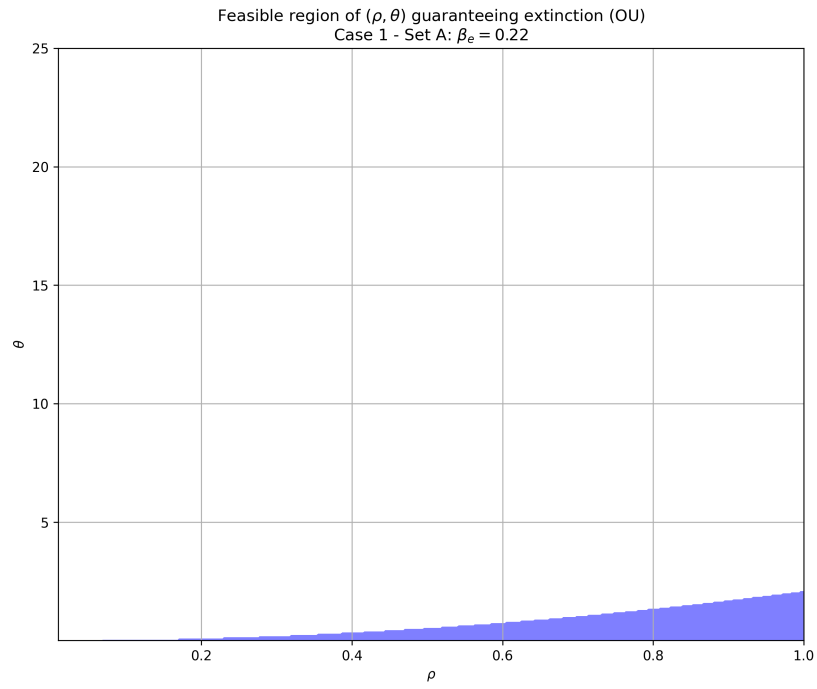


Figure A.76: Feasible region plot for the OU process in Set A, Case 1.

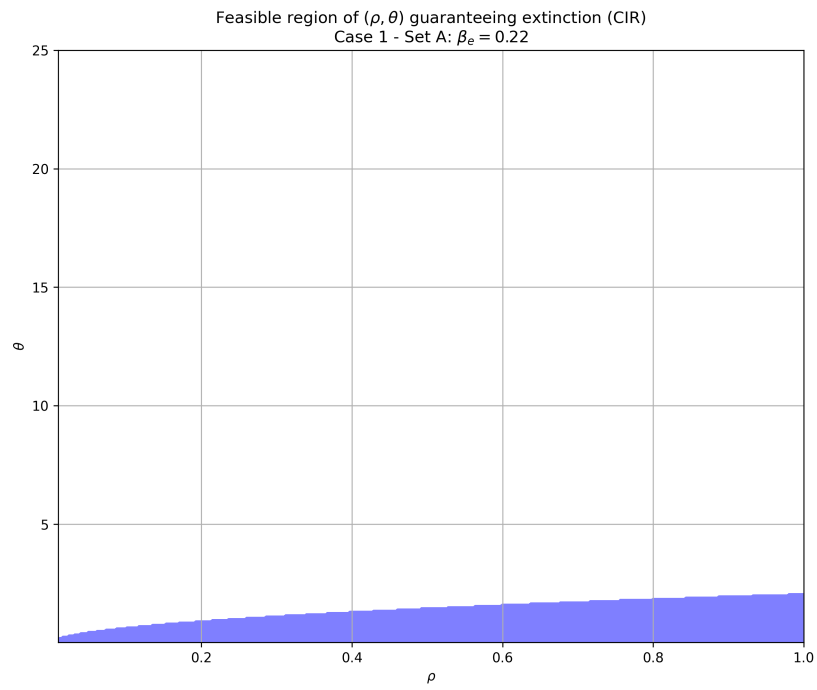


Figure A.77: Feasible region plot for the CIR process in Set A, Case 1.

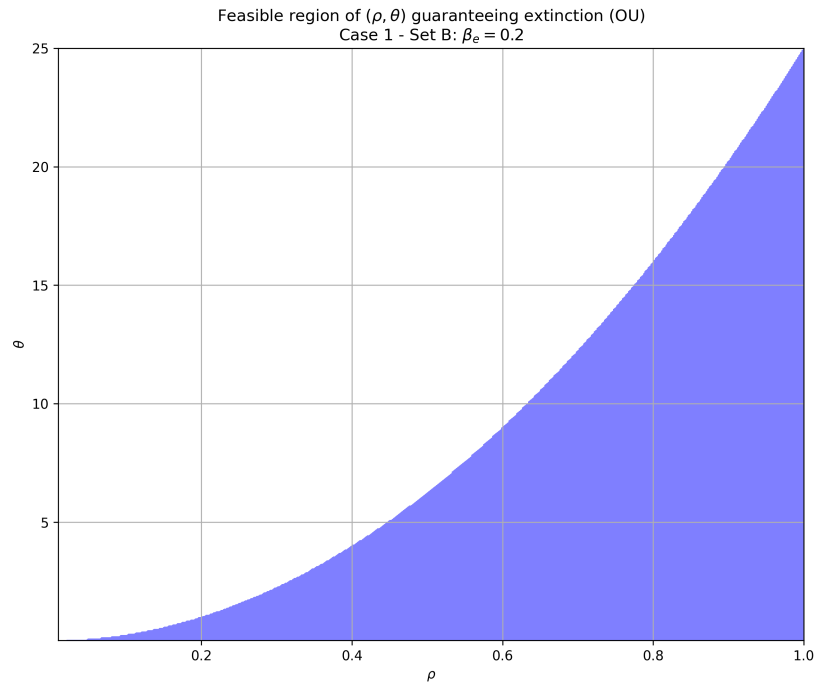


Figure A.78: Feasible region plot for the OU process in Set B, Case 1.

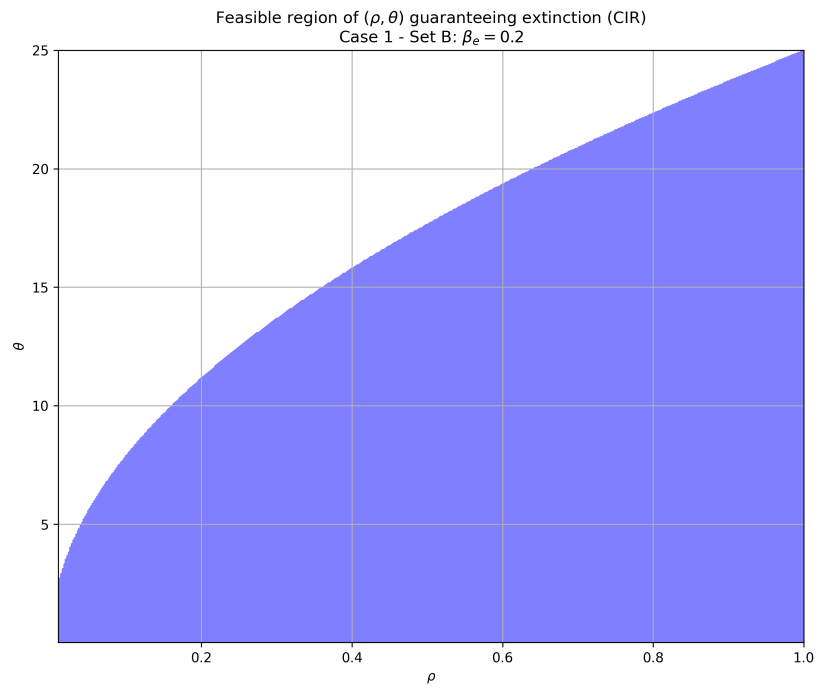


Figure A.79: Feasible region plot for the CIR process in Set B, Case 1.

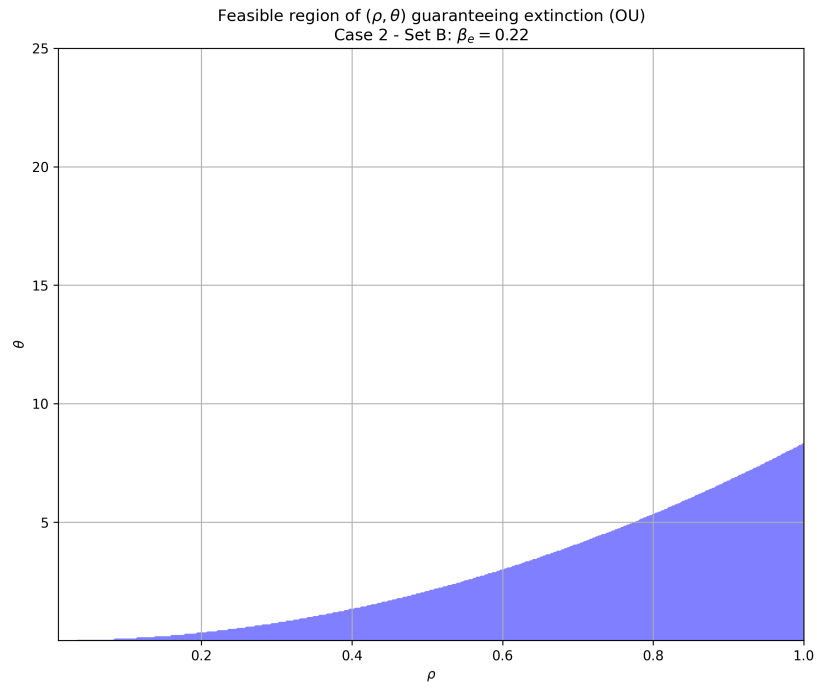


Figure A.80: Feasible region plot for the OU process in Set B, Case 2.

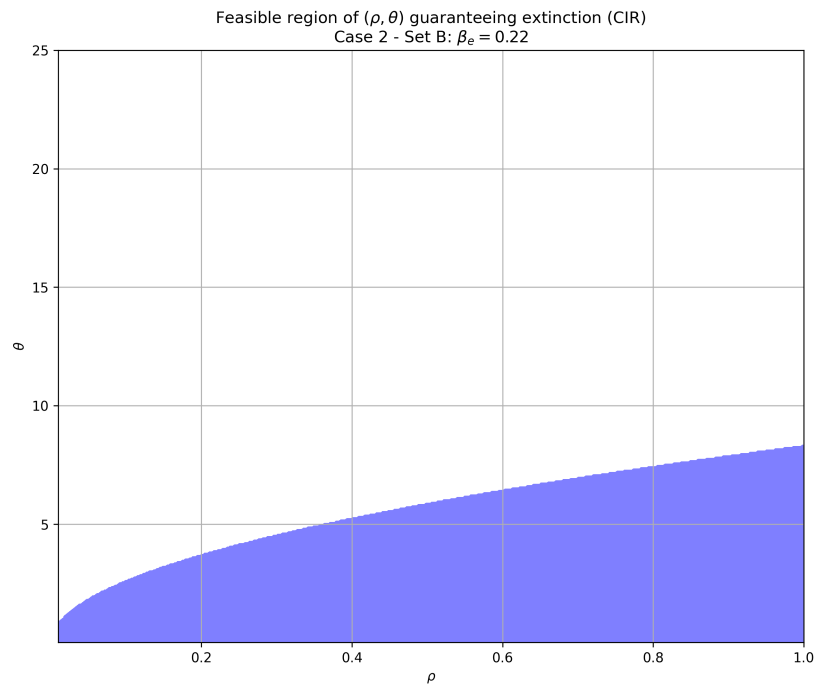


Figure A.81: Feasible region plot for the CIR process in Set B, Case 2.

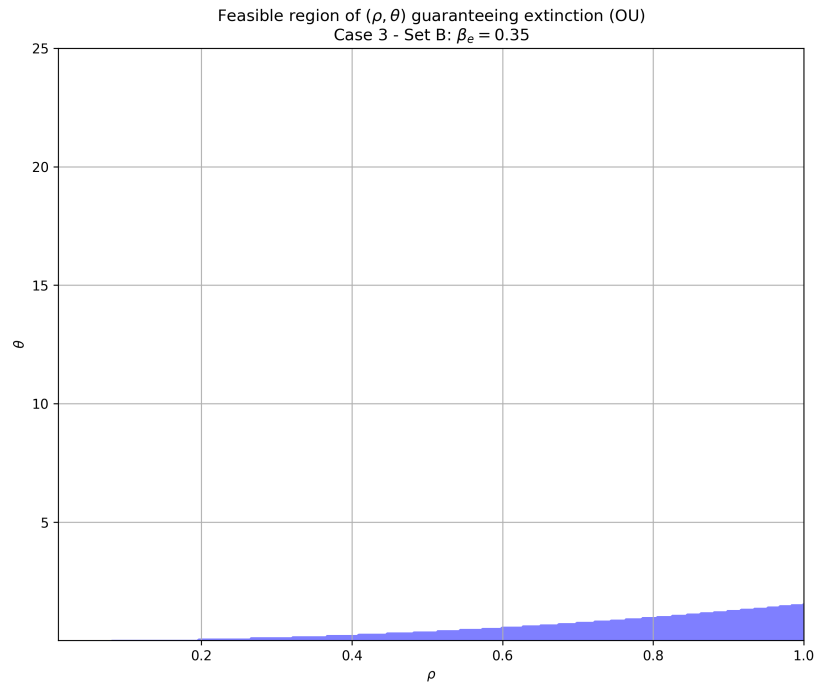


Figure A.82: Feasible region plot for the OU process in Set B, Case 4.

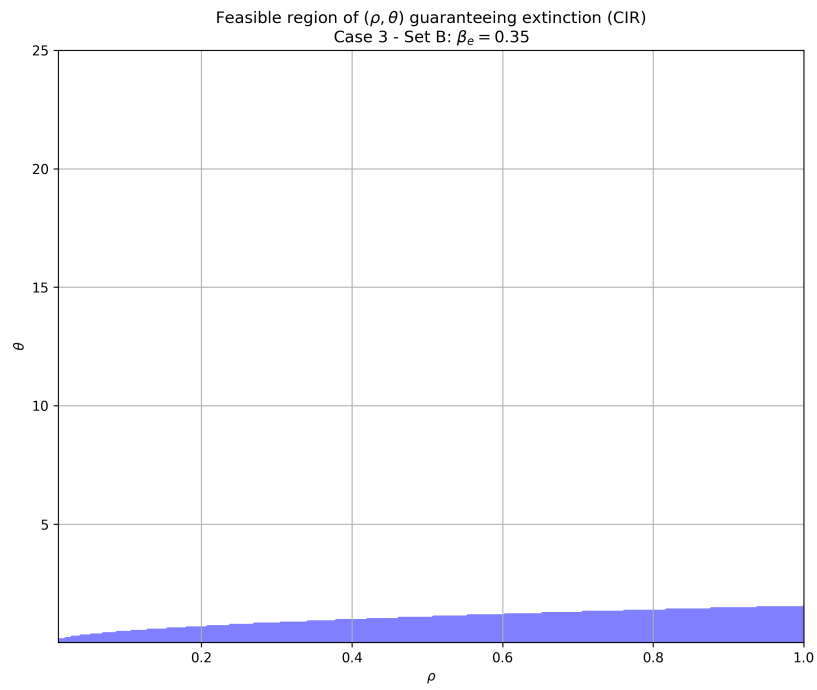


Figure A.83: Feasible region plot for the CIR process in Set B, Case 4.

A.2 Key mathematical concepts for Chapter 5

This section provides an introductory overview of several key mathematical concepts that underpin the stochastic analysis conducted in this work. These include the Fokker–Planck equation, Brownian motion, Itô calculus (including Itô’s Lemma and Itô isometry), martingales and their properties, and Doob’s h -transform. These concepts have deep historical roots and have been developed to address complex problems in stochastic processes, with wide-ranging applications in physics, finance, biology, and beyond.

A.2.1 Fokker–Planck equation

The Fokker–Planck equation [116], also known as the forward Kolmogorov equation, was first introduced by Adriaan Fokker and Max Planck in the early 20th century. It describes the time evolution of the probability density function of a stochastic process, providing a deterministic way to understand the behaviour of systems under random influences. In the context of the CKLS process, the Fokker–Planck equation helps us understand how the probability distribution of the transmission rate $\beta(t)$ changes over time.

Historically, the Fokker–Planck equation has been pivotal in the study of Brownian motion and has been applied in various fields, from statistical mechanics to financial mathematics. For a general stochastic process described by the stochastic differential equation (SDE):

$$dX_t = \mu(X_t, t) dt + \sigma(X_t, t) dB_t,$$

the corresponding Fokker–Planck equation is:

$$\frac{\partial p(x, t)}{\partial t} = -\frac{\partial}{\partial x}[\mu(x, t)p(x, t)] + \frac{1}{2} \frac{\partial^2}{\partial x^2}[\sigma^2(x, t)p(x, t)],$$

where $p(x, t)$ is the probability density function of X_t .

A.2.2 Brownian motion

Brownian motion, or Wiener process, is a cornerstone of modern stochastic processes [88]. Its discovery dates back to the 19th century when botanist Robert Brown observed the erratic motion of pollen particles in water. This phenomenon was later mathematically formalized by Albert Einstein and Norbert Wiener, leading to the development of the theory of Brownian motion.

Brownian motion is characterized by the following properties:

- $B(0) = 0$.
- $B(t)$ has independent increments.
- $B(t)$ has normally distributed increments with mean 0 and variance t .
- $B(t)$ has continuous paths.

The motivation behind studying Brownian motion lies in its ability to model random phenomena in various fields, such as physics, finance, and biology. In the context of the SIRS model, Brownian motion introduces randomness in the transmission rate $\beta(t)$, capturing the unpredictable nature of disease transmission in a population.

A.2.3 Itô calculus

Itô calculus [88], [117], developed by Japanese mathematician Kiyosi Itô in the 1940s, extends the classical calculus to stochastic processes. It is a fundamental tool for dealing with SDEs, allowing for the rigorous analysis of systems influenced by randomness. Itô calculus includes essential tools such as Itô's Lemma and Itô isometry, which have become standard techniques in stochastic analysis.

Itô's lemma

Itô's Lemma is the stochastic counterpart of the chain rule from classical calculus. It provides a way to compute the differential of a function of a stochastic process, which is crucial for analyzing SDEs. For a function $f(X_t, t)$ where X_t follows the SDE $dX_t = \mu(X_t, t) dt + \sigma(X_t, t) dB_t$, Itô's Lemma states:

$$df(X_t, t) = \left(\frac{\partial f}{\partial t} + \mu \frac{\partial f}{\partial X_t} + \frac{1}{2} \sigma^2 \frac{\partial^2 f}{\partial X_t^2} \right) dt + \sigma \frac{\partial f}{\partial X_t} dB_t.$$

The significance of Itô's Lemma lies in its ability to handle the non-linearities that arise in stochastic systems, making it indispensable in fields such as financial mathematics for option pricing, and in biology for modelling population dynamics.

Itô isometry

Itô isometry is a property that links the variance of a stochastic integral to the integral of the variance of the integrand. It plays a crucial role in simplifying the computation of variances

in stochastic processes. For a process X_t ,

$$\mathbb{E} \left[\left(\int_0^t X_s dB_s \right)^2 \right] = \mathbb{E} \left[\int_0^t X_s^2 ds \right].$$

This result is fundamental in the analysis of stochastic integrals and has applications in signal processing, filtering theory, and beyond.

A.2.4 Martingales and their properties

The concept of a martingale was introduced by French mathematician Paul Lévy in the 1930s as part of the broader development of probability theory. Martingales represent a class of stochastic processes that model fair games, where the expected future value of the process, given all past information, is equal to its current value [89].

Formally, a process M_t is a martingale if:

$$\mathbb{E}[M_{t+s} \mid \mathcal{F}_t] = M_t \quad \text{for all } s \geq 0,$$

where $\mathcal{F}_t$ is the filtration representing the information available up to time t .

Martingales have become a central concept in stochastic processes due to their rich mathematical structure and their applications in various domains, including finance (for modelling asset prices), and in stochastic control. Key properties of martingales include:

- Martingale property: $\mathbb{E}[M_t] = M_0$.
- Quadratic variation: For a martingale M_t , the quadratic variation $[M]_t$ tracks the accumulated variance.

A.2.5 Doob's h -transform and its applications to stochastic processes

Doob's h -transform is a mathematical technique in probability theory and stochastic processes used to modify a Markov process to condition it on a specific event or to change its behaviour to achieve a desired property. This transformation constructs a new Markov process from an existing one by using a harmonic function $h(x)$ associated with the generator of the original process. The transformed process allows us to study conditional behaviours, hitting times, and other probabilistic events.

Key concepts of Doob's h -transform are as follows:

- Markov processes and harmonic functions: A Markov process is a stochastic process that has the “memoryless” property: the future state of the process depends only on the current state, not on the sequence of events that preceded it. A function $h(x)$ is called harmonic with respect to a Markov process if it satisfies the generator equation $\mathcal{L}h(x) = 0$, where $\mathcal{L}$ is the infinitesimal generator of the Markov process.
- Infinitesimal generator $\mathcal{L}$: The infinitesimal generator $\mathcal{L}$ of a Markov process X_t is an operator that describes how functions of the process change infinitesimally over time. For a function $f(x)$, the generator $\mathcal{L}f(x)$ typically takes the form:

$$\mathcal{L}f(x) = b(x)f'(x) + \frac{1}{2}a(x)f''(x),$$

where $b(x)$ is the drift term and $a(x)$ is the diffusion term (often $a(x) = \sigma^2(x)$).

- Construction of the transformed process: Given a Markov process X_t with generator $\mathcal{L}$ and a positive harmonic function $h(x)$, Doob’s h -transform creates a new process X_t^h such that the new process behaves like the original process but is conditioned or modified according to $h(x)$. The SDE of the transformed process is:

$$dX_t^h = \left(b(X_t^h) + a(X_t^h) \frac{h'(X_t^h)}{h(X_t^h)} \right) dt + \sigma(X_t^h) dW_t,$$

where W_t is a standard Brownian motion.

Applications include: conditioning on events whereby the h -transform is used to model processes conditioned to avoid a set or to reach a target, and, hitting times and absorption which are useful for studying hitting probabilities, times to absorption, or conditioned diffusion processes in physics, finance, and other fields.

Example: Brownian motion conditioned to stay positive

Consider a standard Brownian motion B_t such that we want to condition it to stay positive (i.e., $B_t > 0$ for all $t > 0$). That is, want to construct a process that behaves like Brownian motion but never crosses or touches zero.

To condition Brownian motion to stay positive, we choose the harmonic function $h(x) = x$ for $x > 0$. This function is linear and increases with x , reflecting a bias away from zero.

The generator for Brownian motion B_t is:

$$\mathcal{L}f(x) = \frac{1}{2}f''(x).$$

Using Doob's h -transform with $h(x) = x$, the transformed generator becomes:

$$\mathcal{L}^h f(x) = \frac{1}{2} f''(x) + \frac{f'(x)}{x}.$$

The new SDE for the transformed process X_t^h is:

$$dX_t^h = dB_t + \frac{1}{X_t^h} dt.$$

This is a diffusion process with an additional drift term $\frac{1}{X_t^h}$ that becomes significant as X_t^h approaches zero, effectively pushing the process away from zero. Using an appropriate harmonic function, a similar argument is made for the derivation of the mean-reverting Bessel processes.

A.3 Additional figures for Chapter 6

Prediction for 2018-2019 Season by Linear Regression

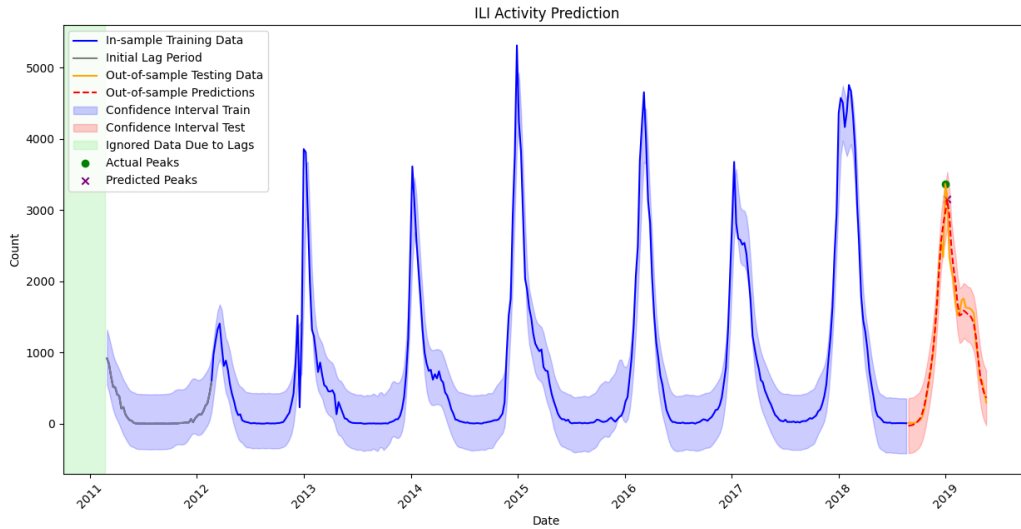


Figure A.84: Prediction for 2018-2019 Season by Linear Regression without Southern Hemisphere Covariate.

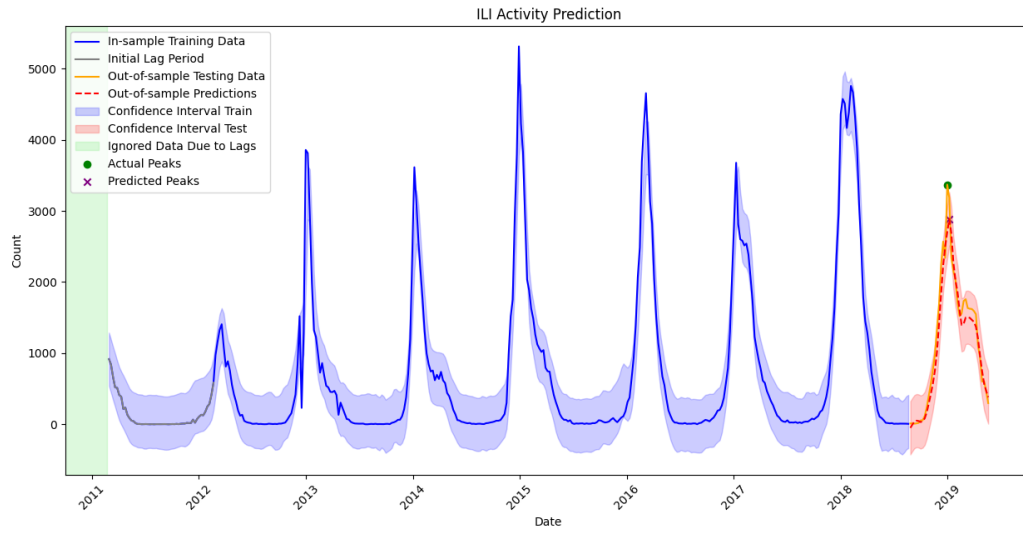


Figure A.85: Prediction for 2018-2019 Season by Linear Regression including Southern Hemisphere Covariate.

2018-2019 Season (Last 10% Training Focus)

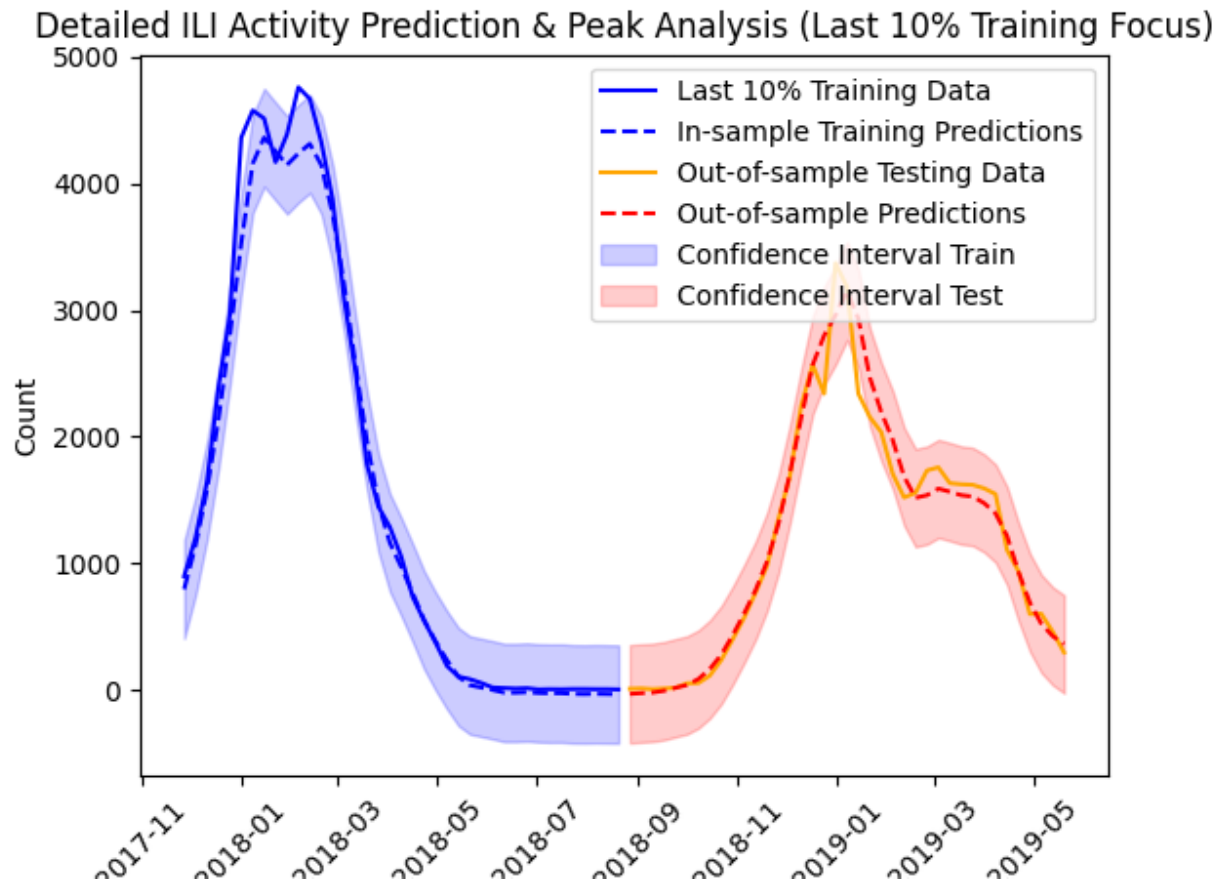


Figure A.86: 2018-2019 Season Prediction Focused on Last 10% of Training Data without Southern Hemisphere Covariate.

Detailed ILI Activity Prediction & Peak Analysis (Last 10% Training Focus)

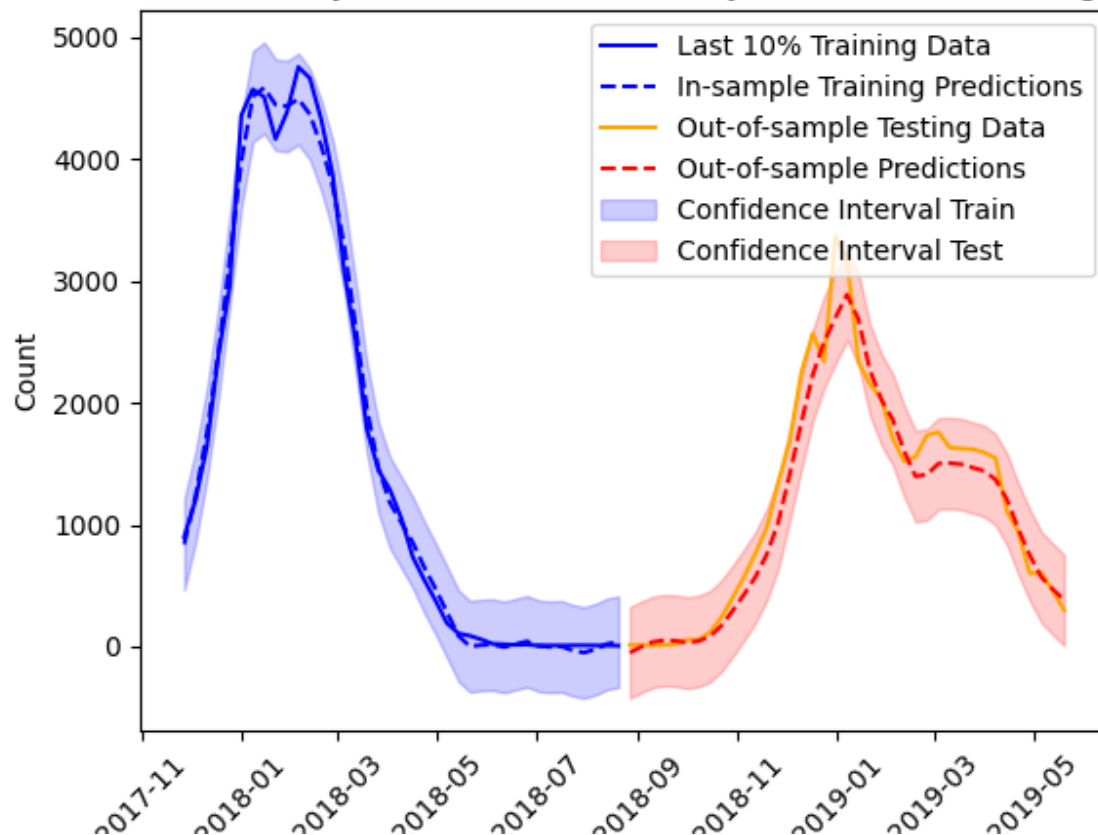


Figure A.87: 2018-2019 Season Prediction Focused on Last 10% of Training Data including Southern Hemisphere Covariate.

2018-2019 Season (Test Set Focus)

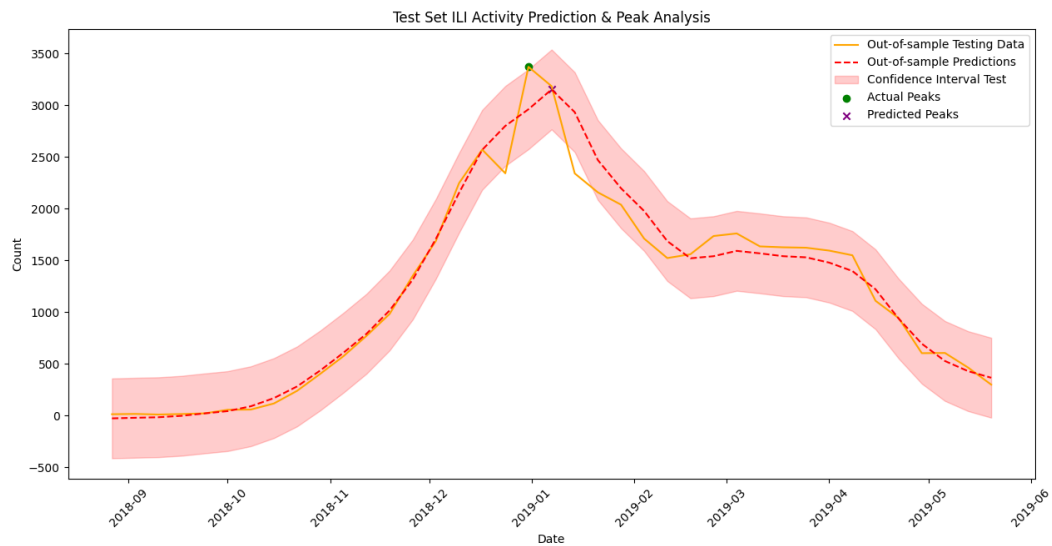


Figure A.88: 2018-2019 Season Prediction on Test Set without Southern Hemisphere Co-variate.

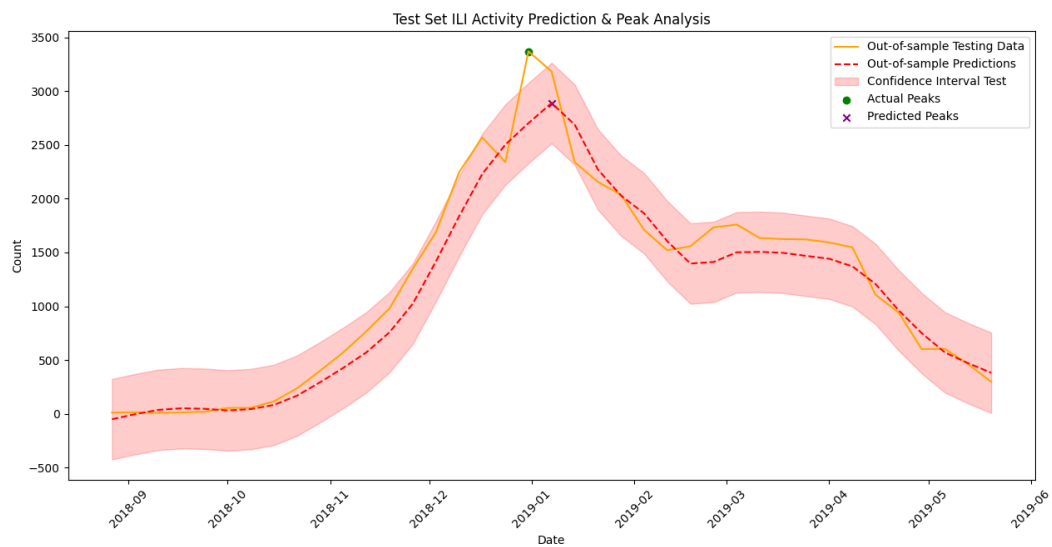


Figure A.89: 2018-2019 Season Prediction on Test Set including Southern Hemisphere Co-variate.

Prediction for 2017-2018 Season by Linear Regression

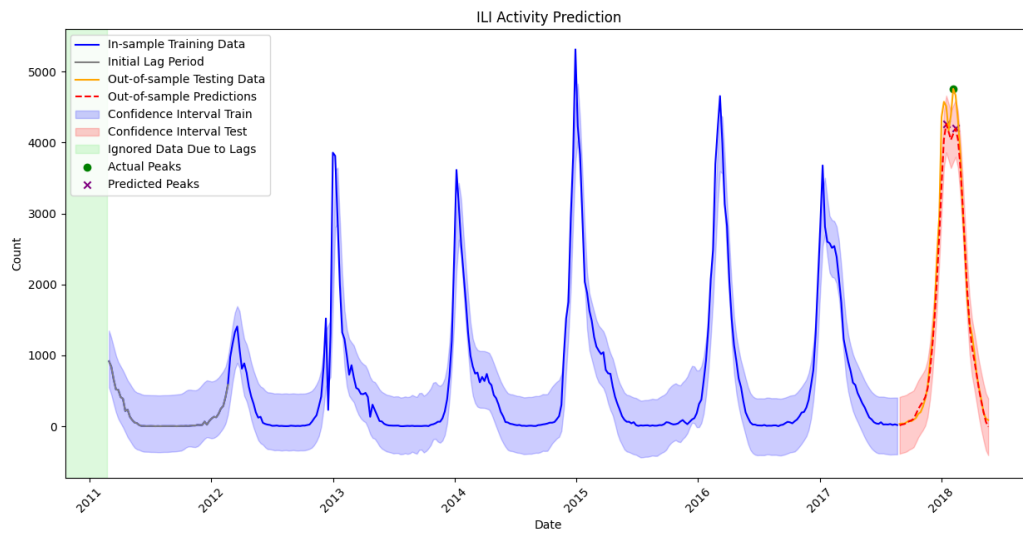


Figure A.90: Prediction for 2017-2018 Season by Linear Regression without Southern Hemisphere Covariate.

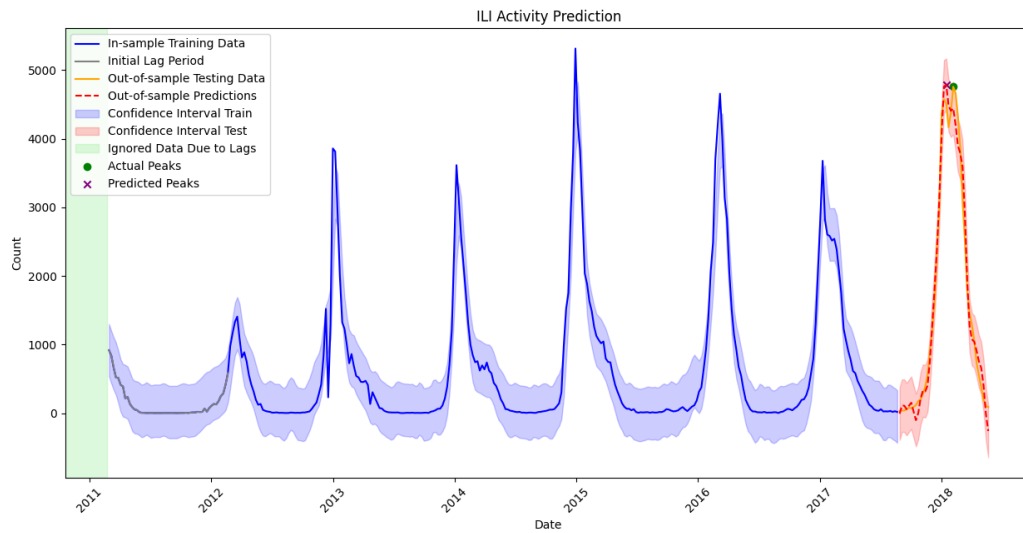


Figure A.91: Prediction for 2017-2018 Season by Linear Regression including Southern Hemisphere Covariate.

2017-2018 Season (Last 10% Training Focus)

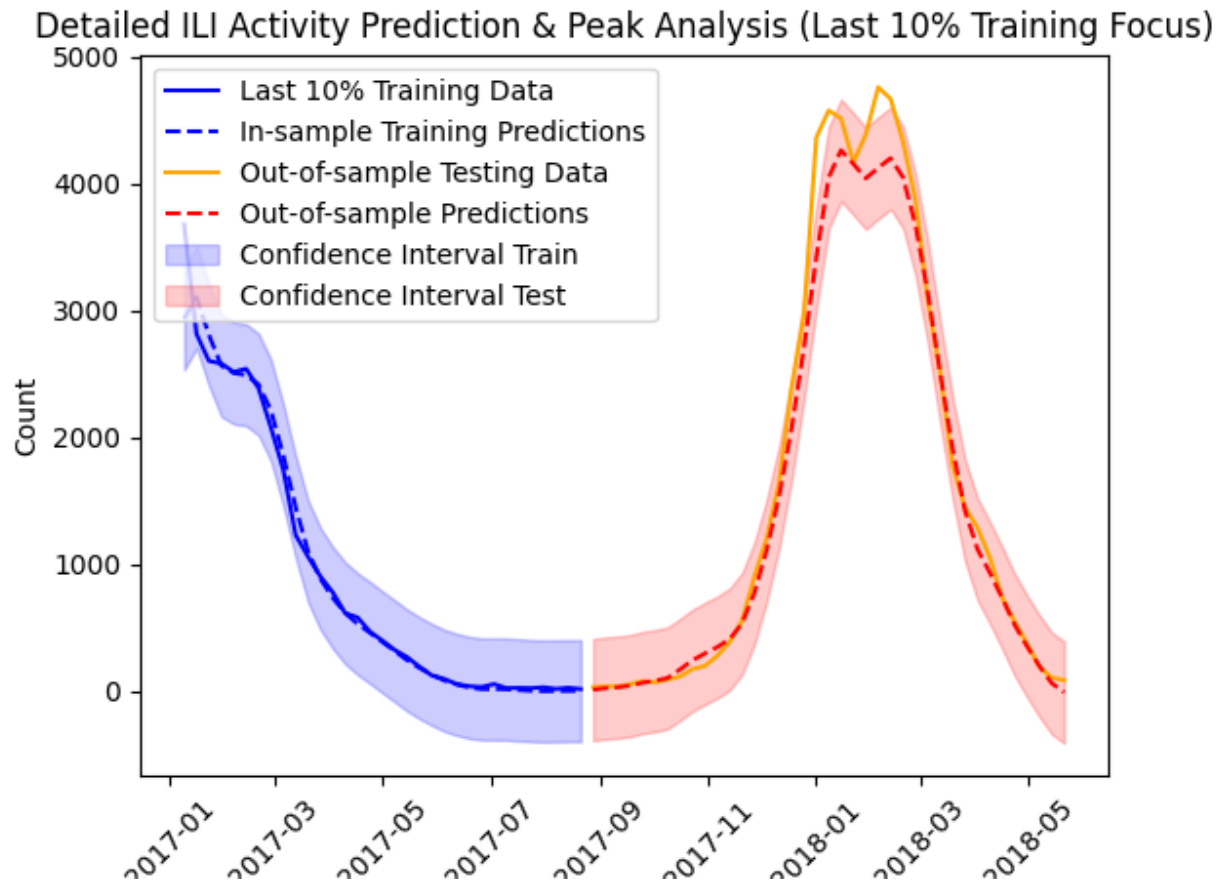


Figure A.92: 2017-2018 Season Prediction Focused on Last 10% of Training Data without Southern Hemisphere Covariate.

Detailed ILI Activity Prediction & Peak Analysis (Last 10% Training Focus)

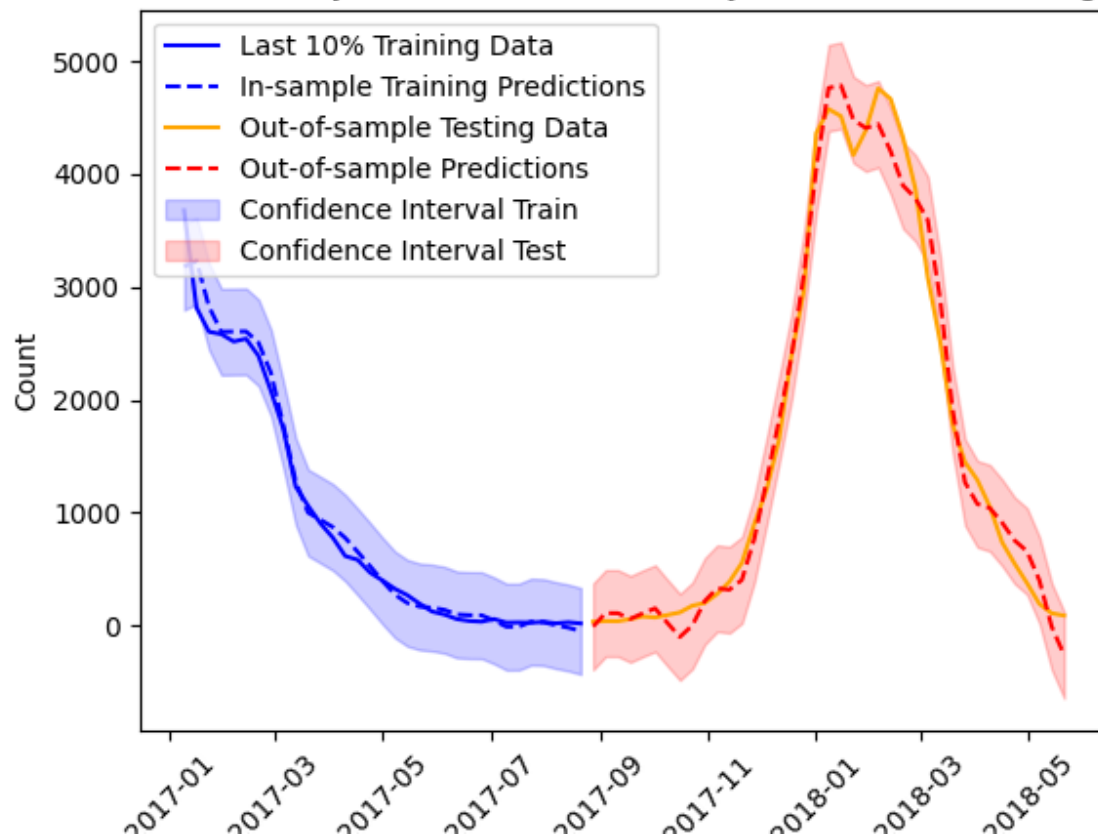


Figure A.93: 2017-2018 Season Prediction Focused on Last 10% of Training Data including Southern Hemisphere Covariate.

2017-2018 Season (Test Set Focus)

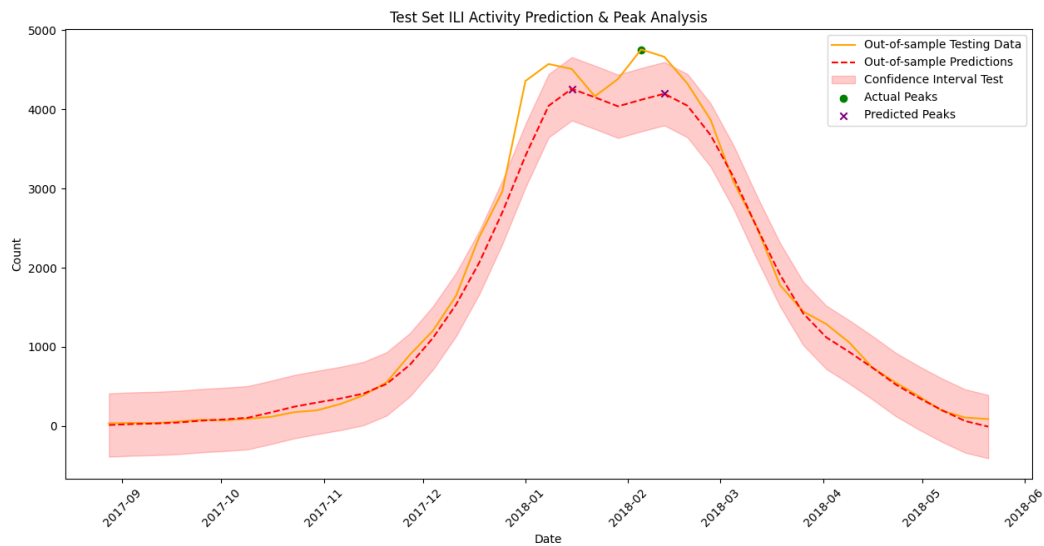


Figure A.94: 2017-2018 Season Prediction on Test Set without Southern Hemisphere Co-variate.

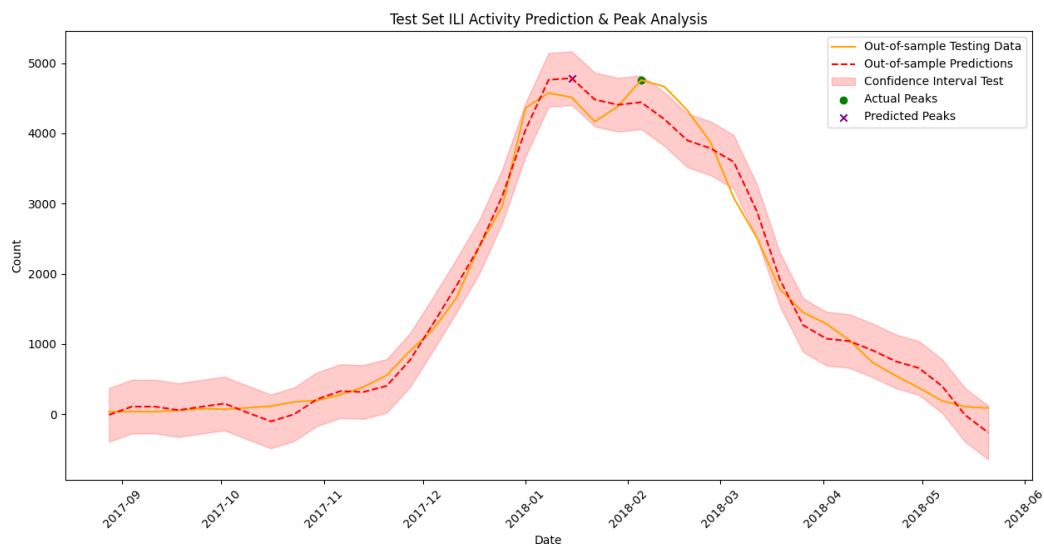


Figure A.95: 2017-2018 Season Prediction on Test Set including Southern Hemisphere Co-variate.

Prediction for 2016-2017 Season by Linear Regression

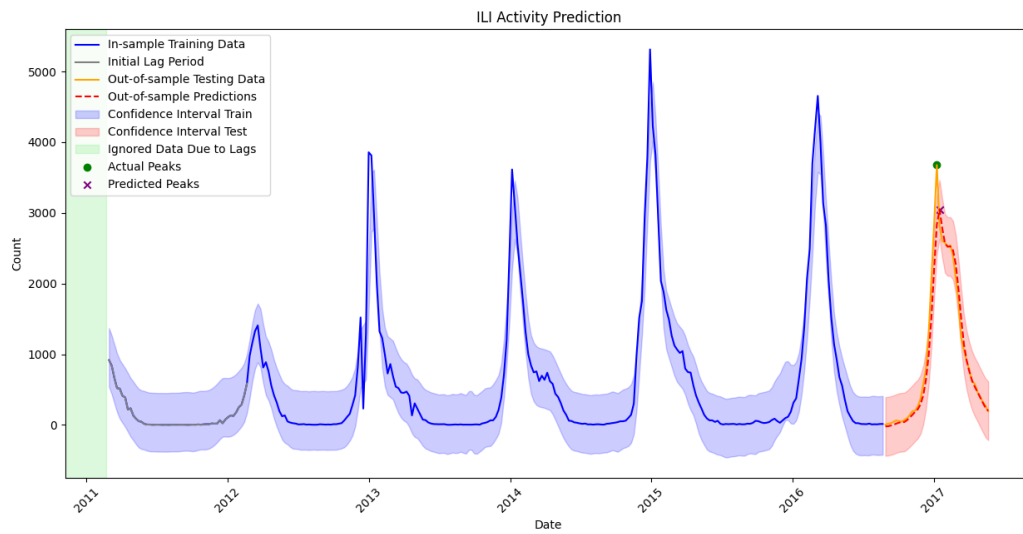


Figure A.96: Prediction for 2016-2017 Season by Linear Regression without Southern Hemisphere Covariate.

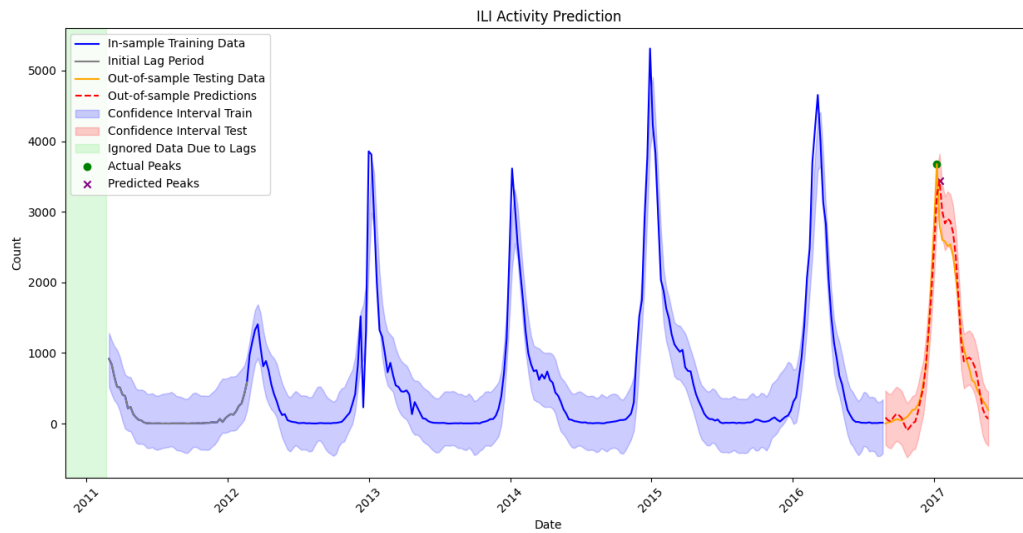


Figure A.97: Prediction for 2016-2017 Season by Linear Regression including Southern Hemisphere Covariate.

2016-2017 Season (Last 10% Training Focus)

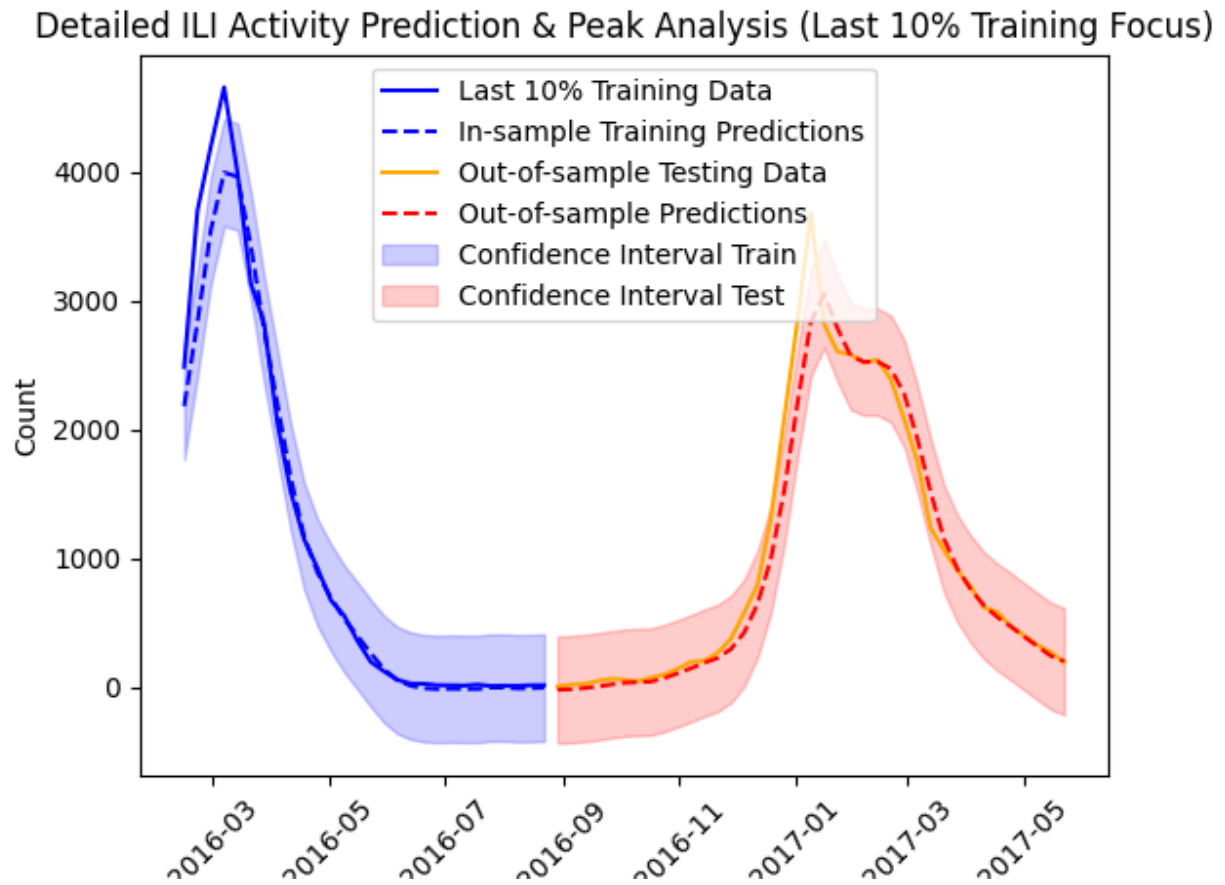


Figure A.98: 2016-2017 Season Prediction Focused on Last 10% of Training Data without Southern Hemisphere Covariate.

Detailed ILI Activity Prediction & Peak Analysis (Last 10% Training Focus)

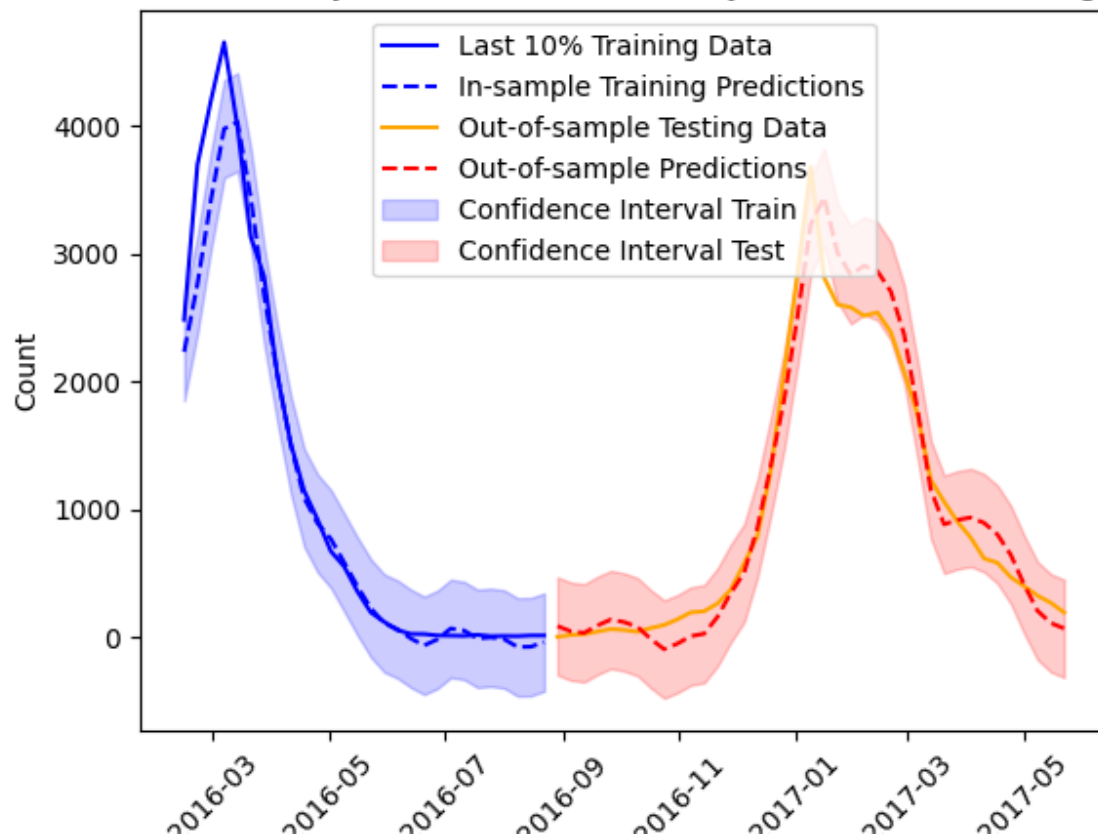


Figure A.99: 2016-2017 Season Prediction Focused on Last 10% of Training Data including Southern Hemisphere Covariate.

2016-2017 Season (Test Set Focus)

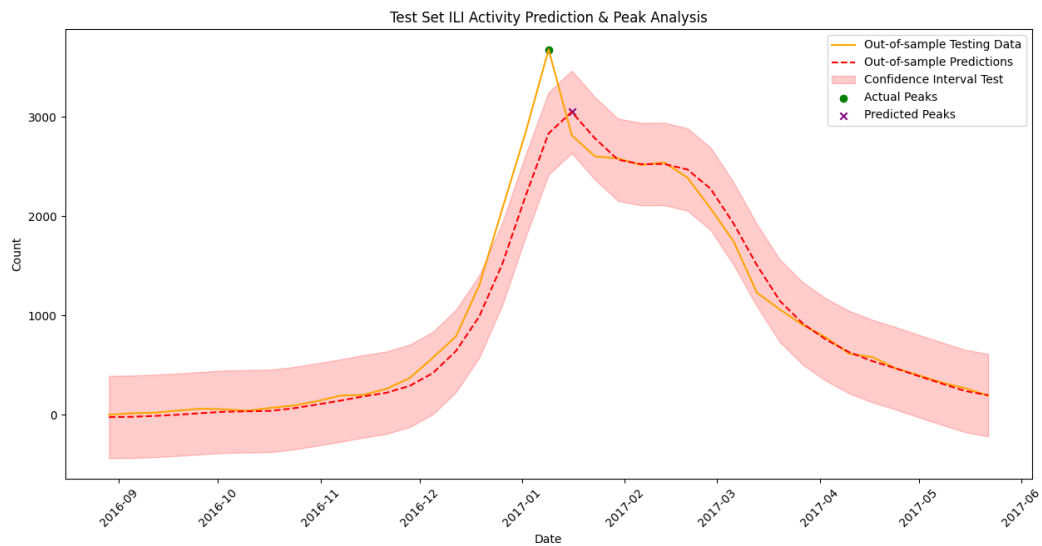


Figure A.100: 2016-2017 Season Prediction on Test Set without Southern Hemisphere Covariate.

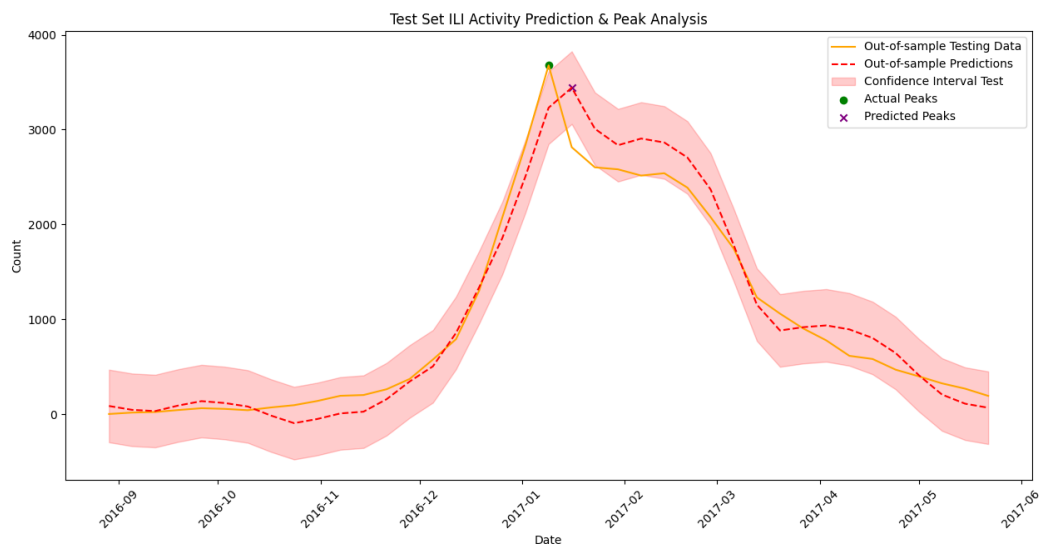


Figure A.101: 2016-2017 Season Prediction on Test Set including Southern Hemisphere Covariate.

Prediction for 2015-2016 Season by Linear Regression

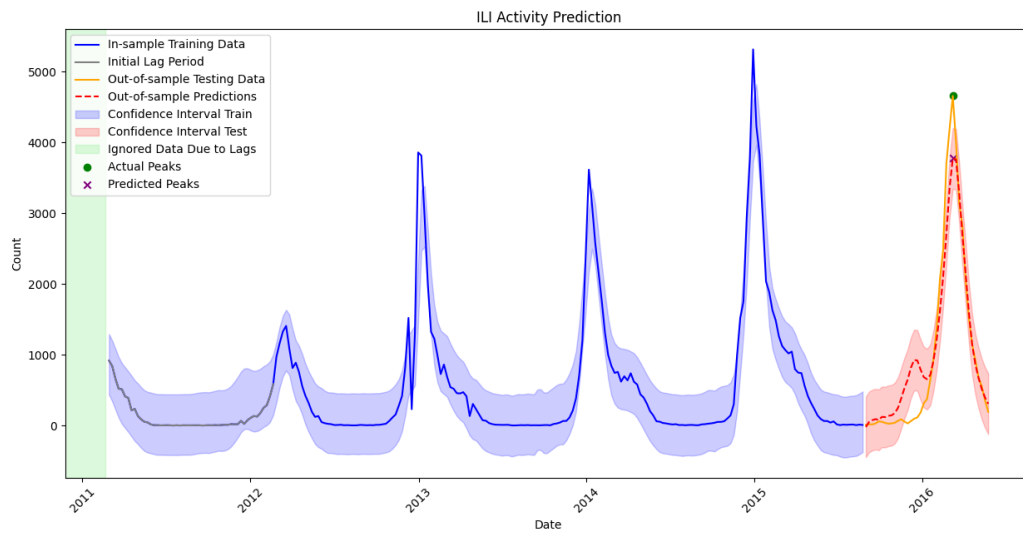


Figure A.102: Prediction for 2015-2016 Season by Linear Regression without Southern Hemisphere Covariate.

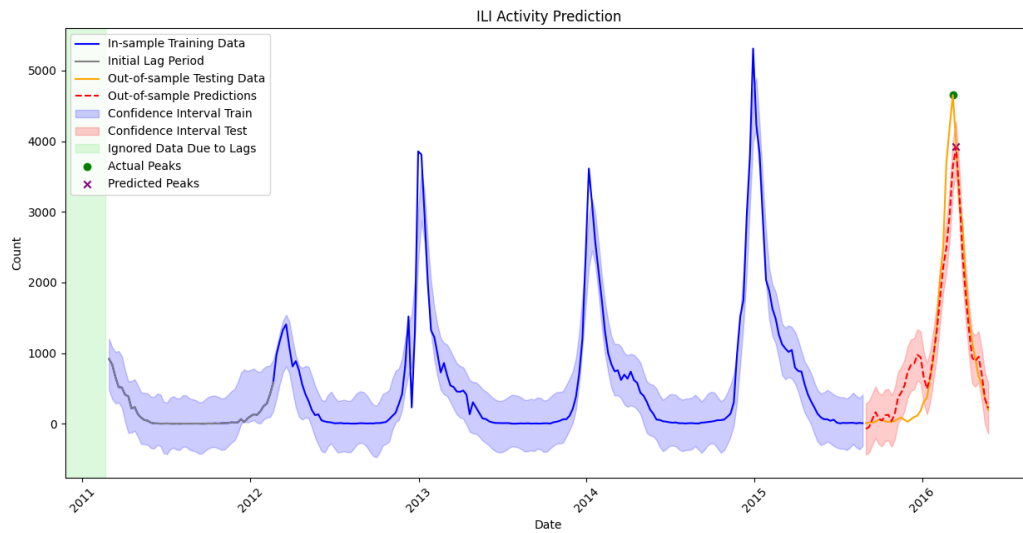


Figure A.103: Prediction for 2015-2016 Season by Linear Regression including Southern Hemisphere Covariate.

2015-2016 Season (Last 10% Training Focus)

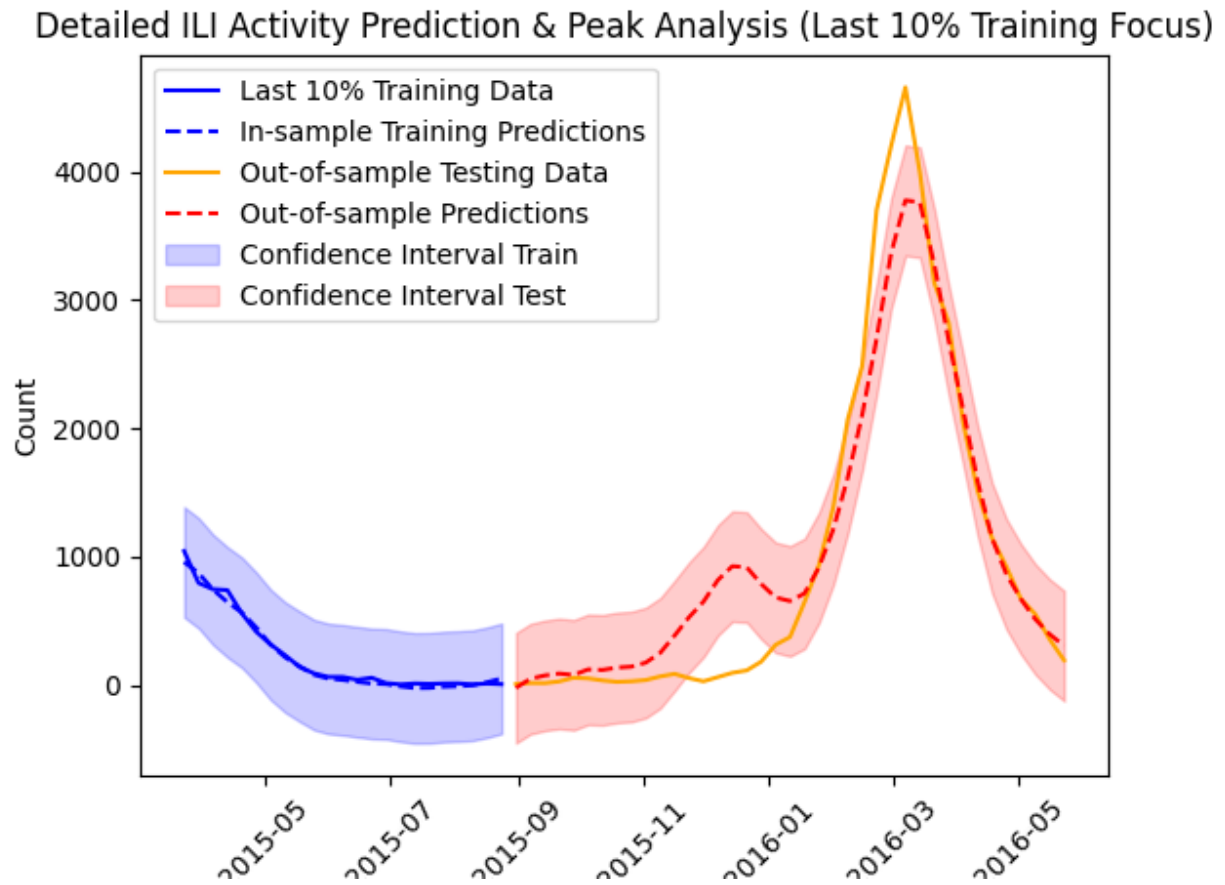


Figure A.104: 2015-2016 Season Prediction Focused on Last 10% of Training Data without Southern Hemisphere Covariate.

Detailed ILI Activity Prediction & Peak Analysis (Last 10% Training Focus)

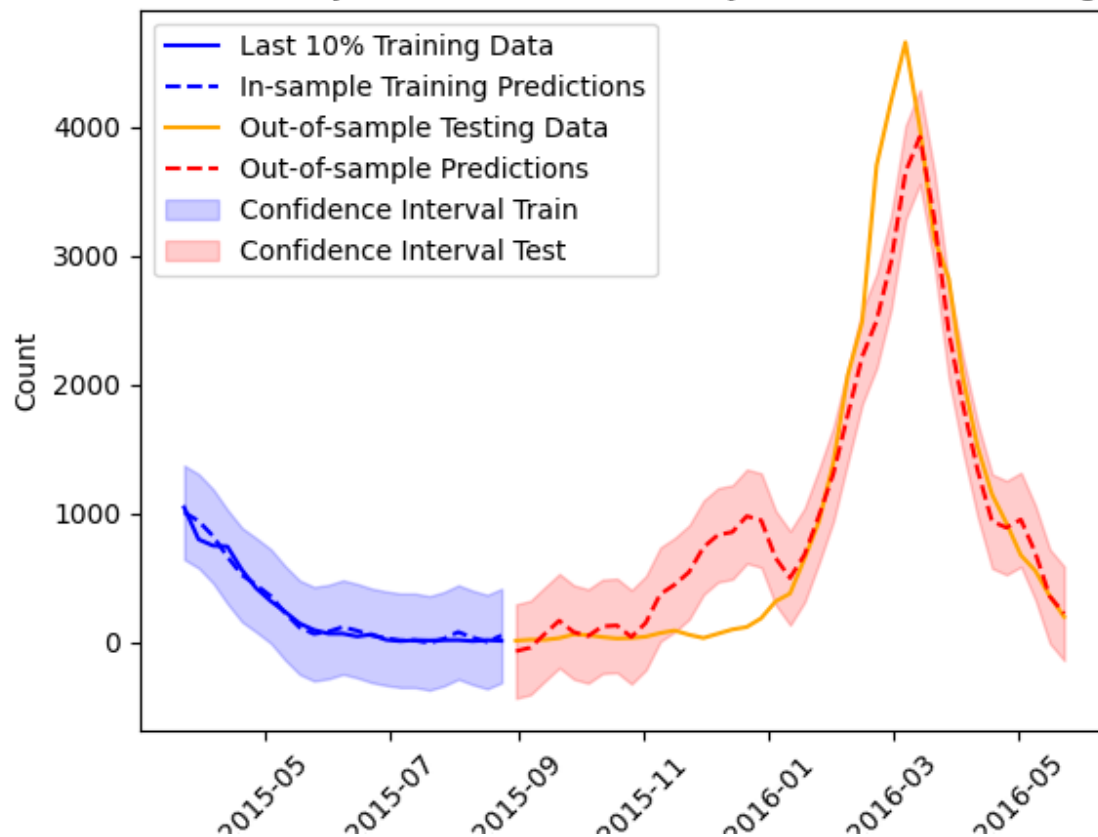


Figure A.105: 2015-2016 Season Prediction Focused on Last 10% of Training Data including Southern Hemisphere Covariate.

2015-2016 Season (Test Set Focus)

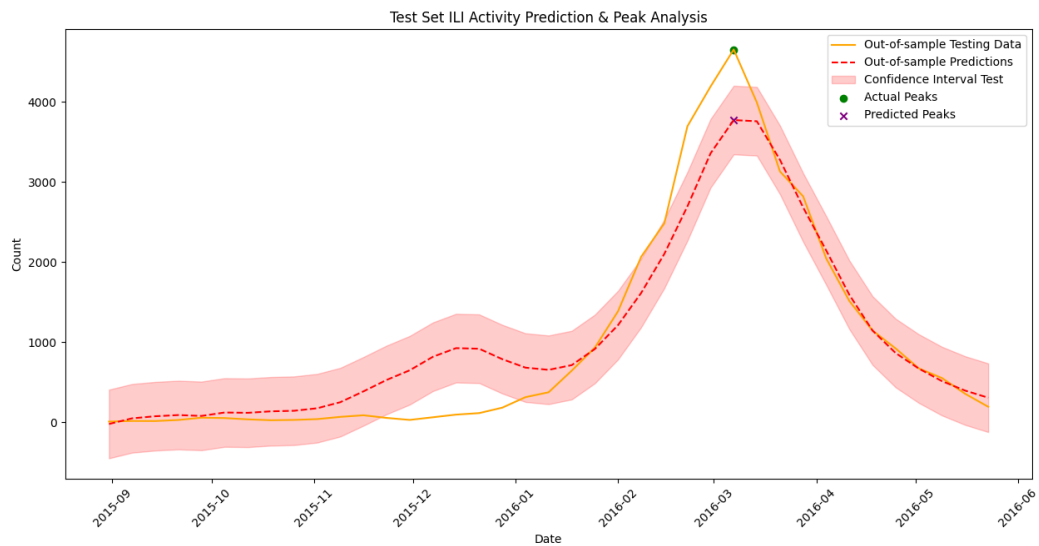


Figure A.106: 2015-2016 Season Prediction on Test Set without Southern Hemisphere Covariate.

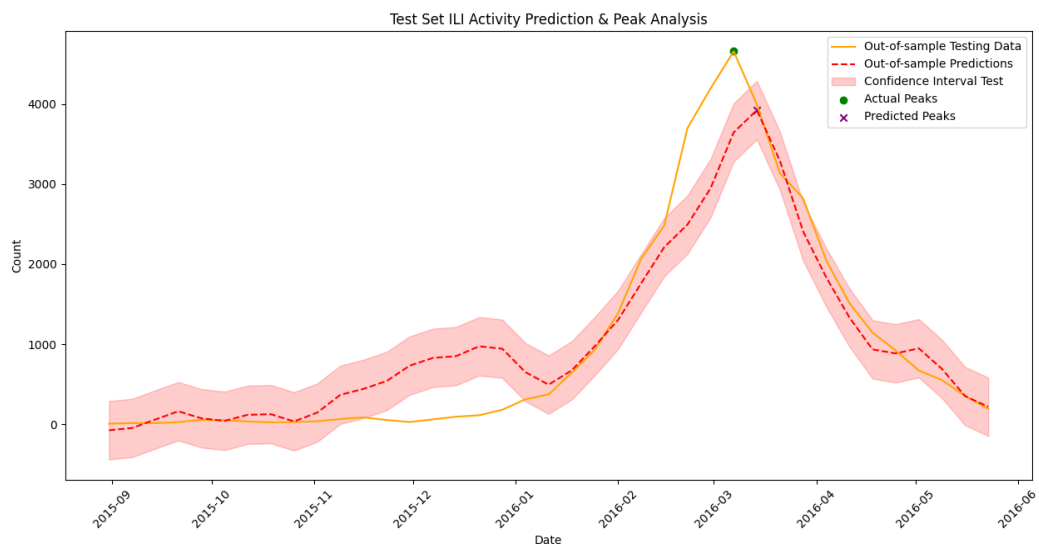


Figure A.107: 2015-2016 Season Prediction on Test Set including Southern Hemisphere Covariate.

Prediction for 2014-2015 Season by Linear Regression

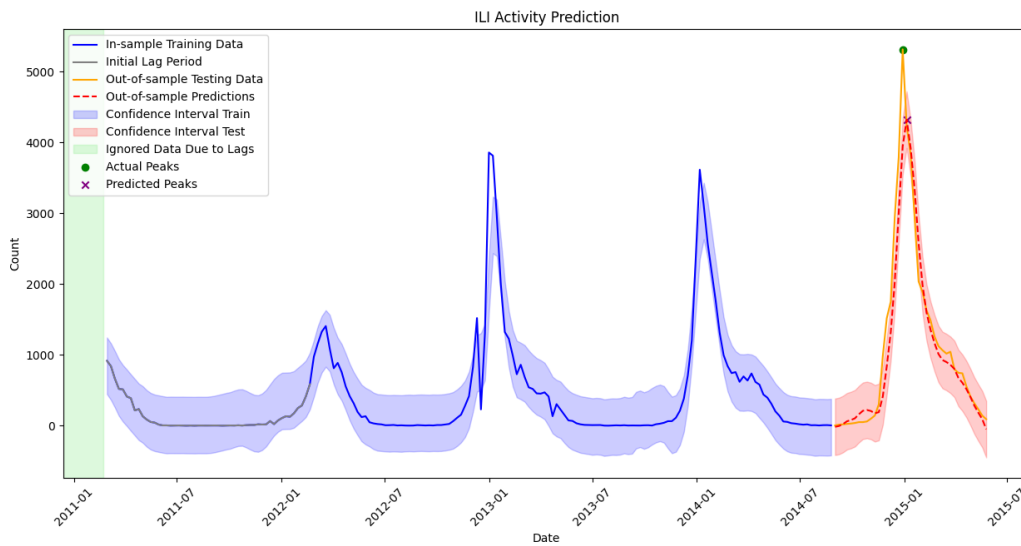


Figure A.108: Prediction for 2014-2015 Season by Linear Regression without Southern Hemisphere Covariate.

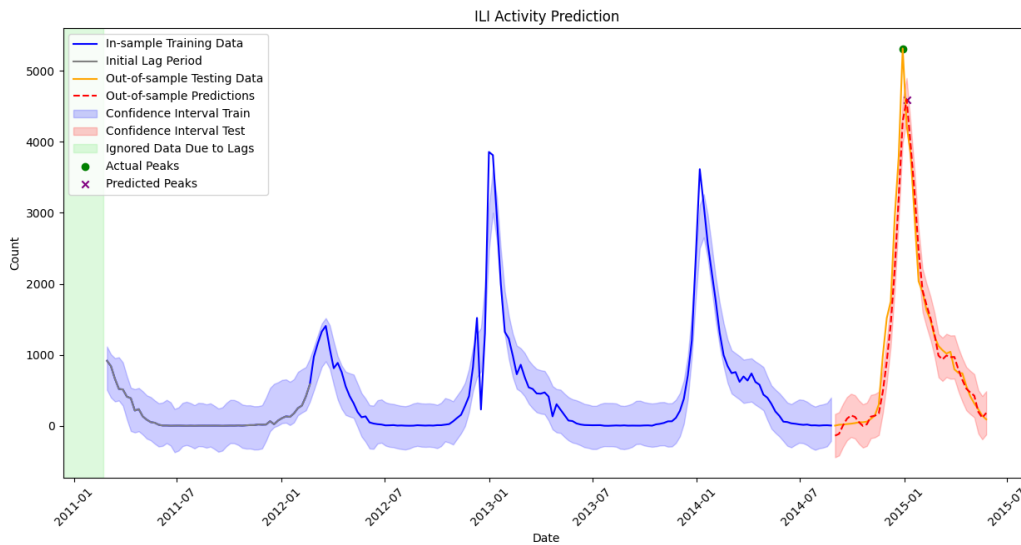


Figure A.109: Prediction for 2014-2015 Season by Linear Regression including Southern Hemisphere Covariate.

2014-2015 Season (Last 10% Training Focus)

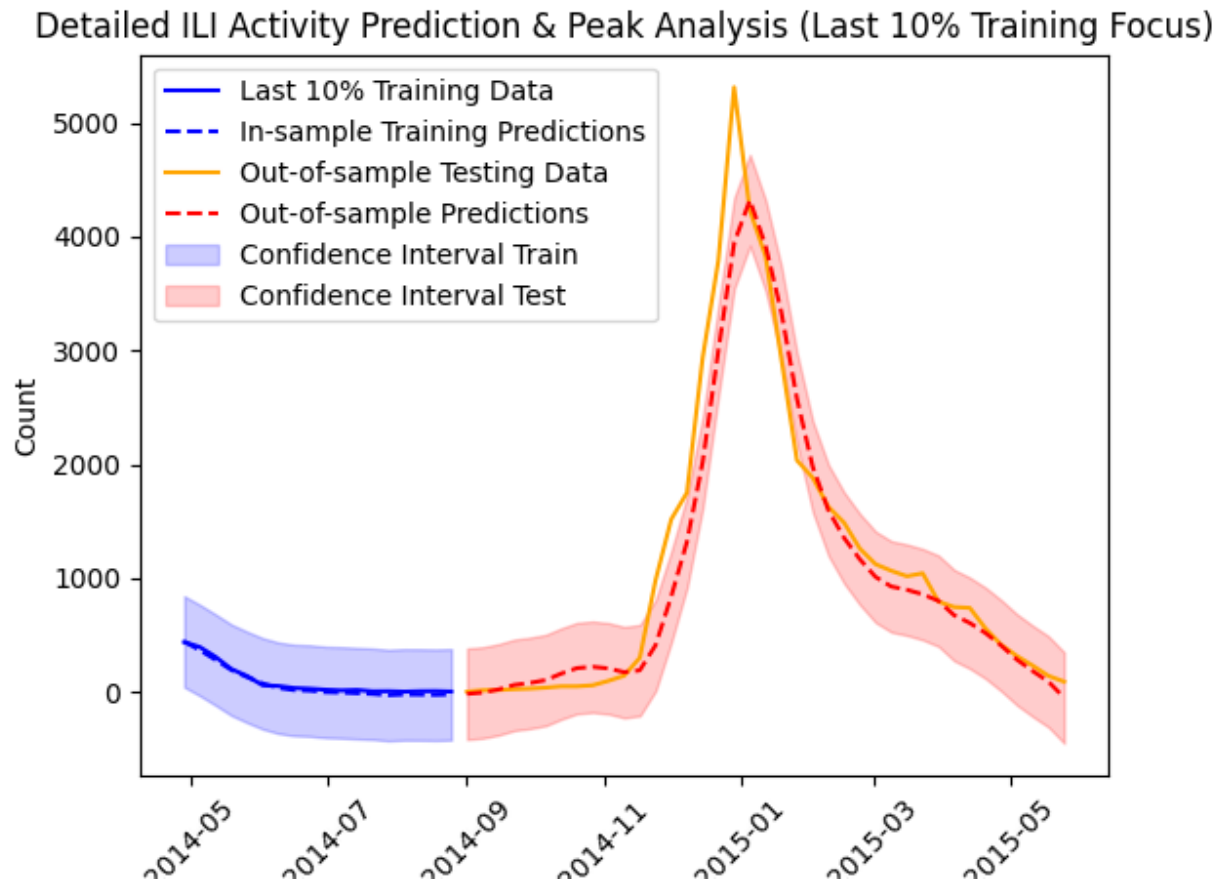


Figure A.110: 2014-2015 Season Prediction Focused on Last 10% of Training Data without Southern Hemisphere Covariate.

Detailed ILI Activity Prediction & Peak Analysis (Last 10% Training Focus)

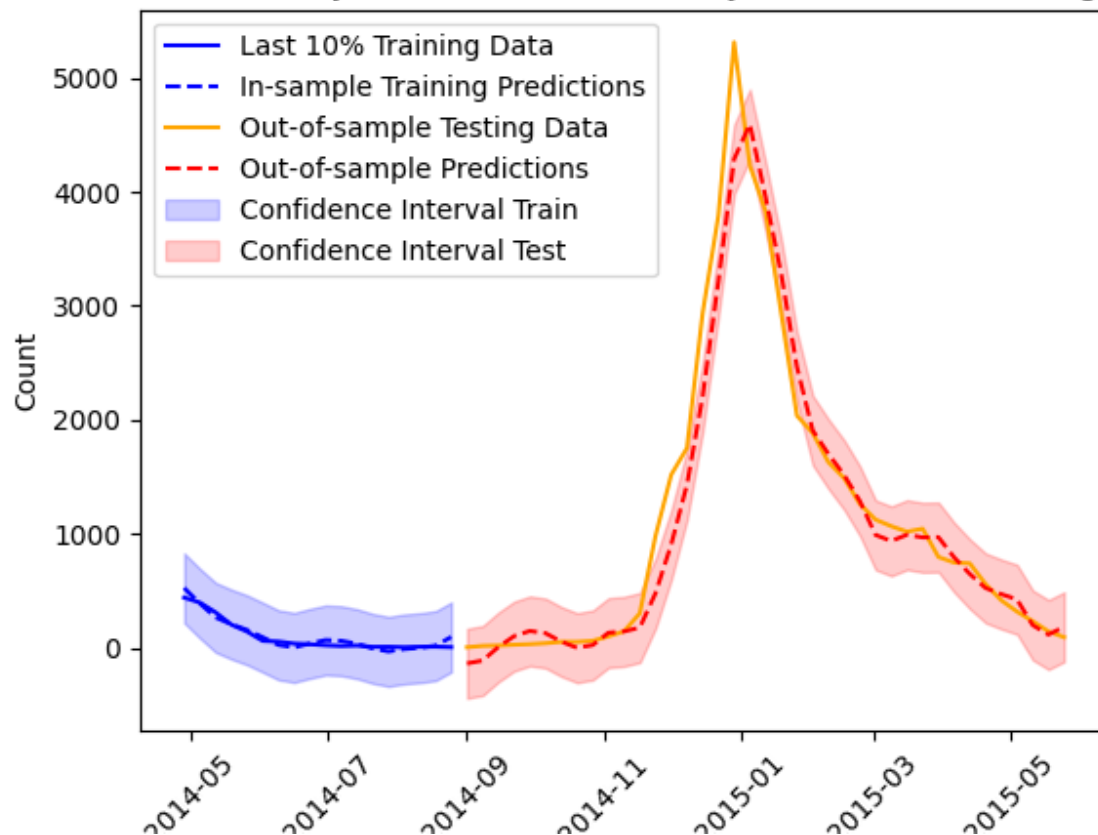


Figure A.111: 2014-2015 Season Prediction Focused on Last 10% of Training Data including Southern Hemisphere Covariate.

2014-2015 Season (Test Set Focus)

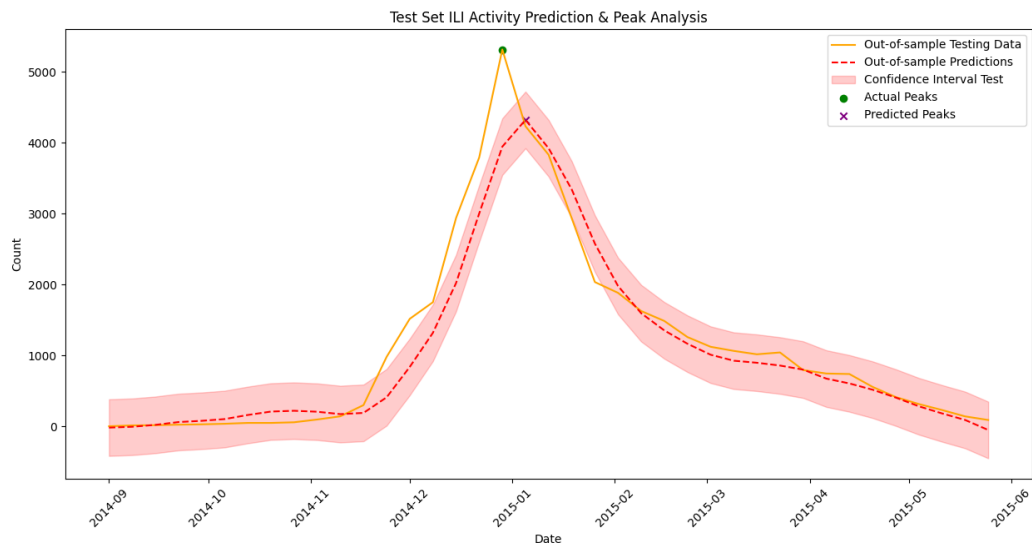


Figure A.112: 2014-2015 Season Prediction on Test Set without Southern Hemisphere Covariate.

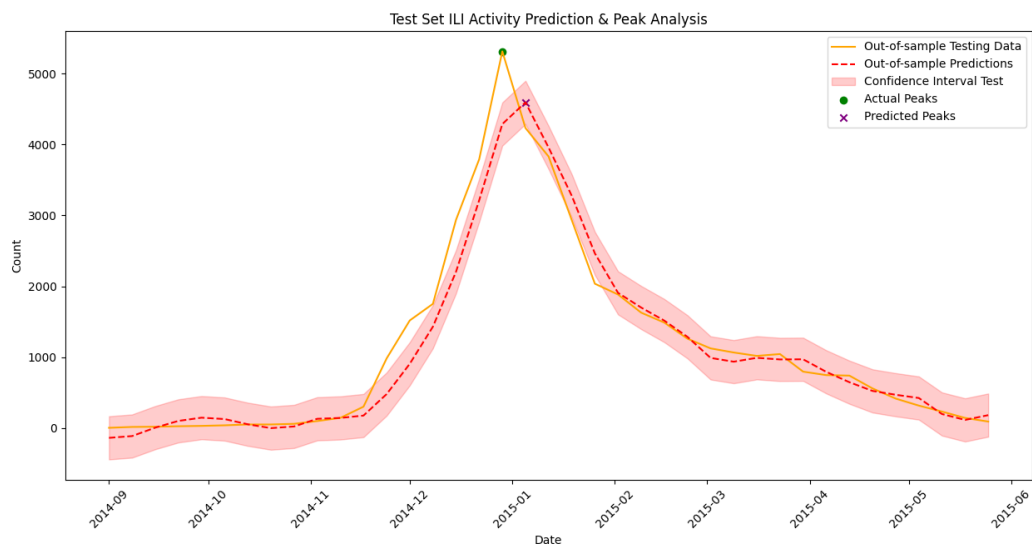


Figure A.113: 2014-2015 Season Prediction on Test Set including Southern Hemisphere Covariate.

Prediction for 2013-2014 Season by Linear Regression

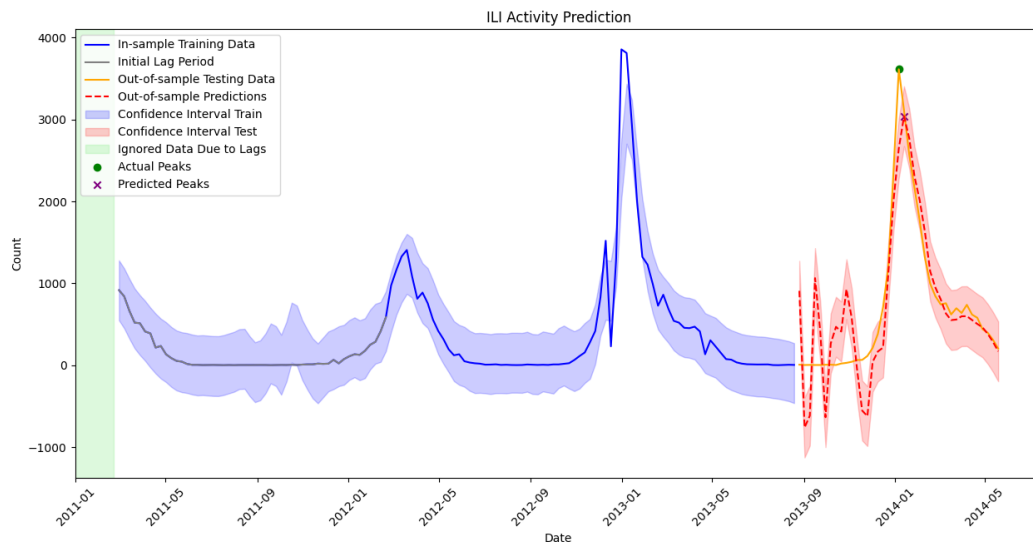


Figure A.114: Prediction for 2013-2014 Season by Linear Regression without Southern Hemisphere Covariate.

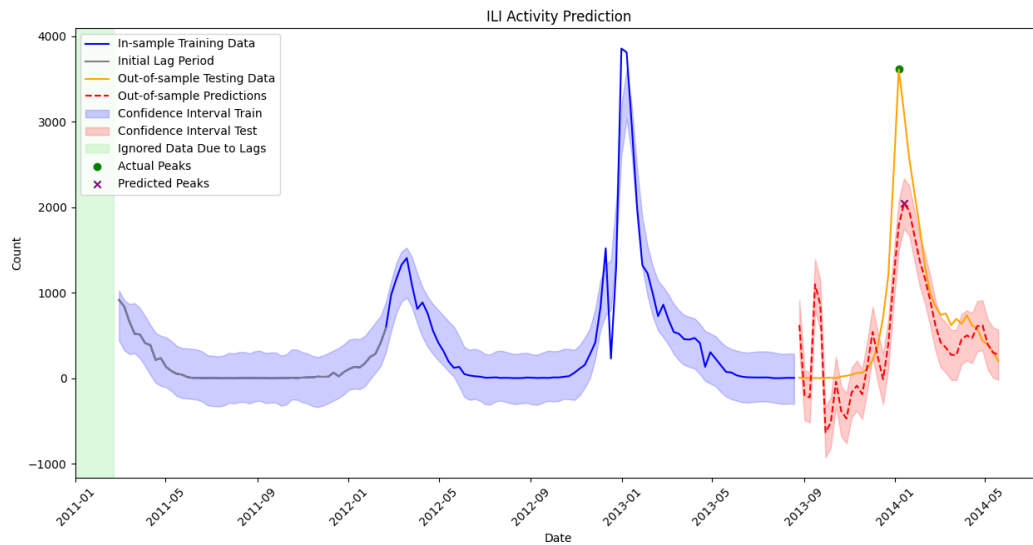


Figure A.115: Prediction for 2013-2014 Season by Linear Regression including Southern Hemisphere Covariate.

2013-2014 Season (Last 10% Training Focus)

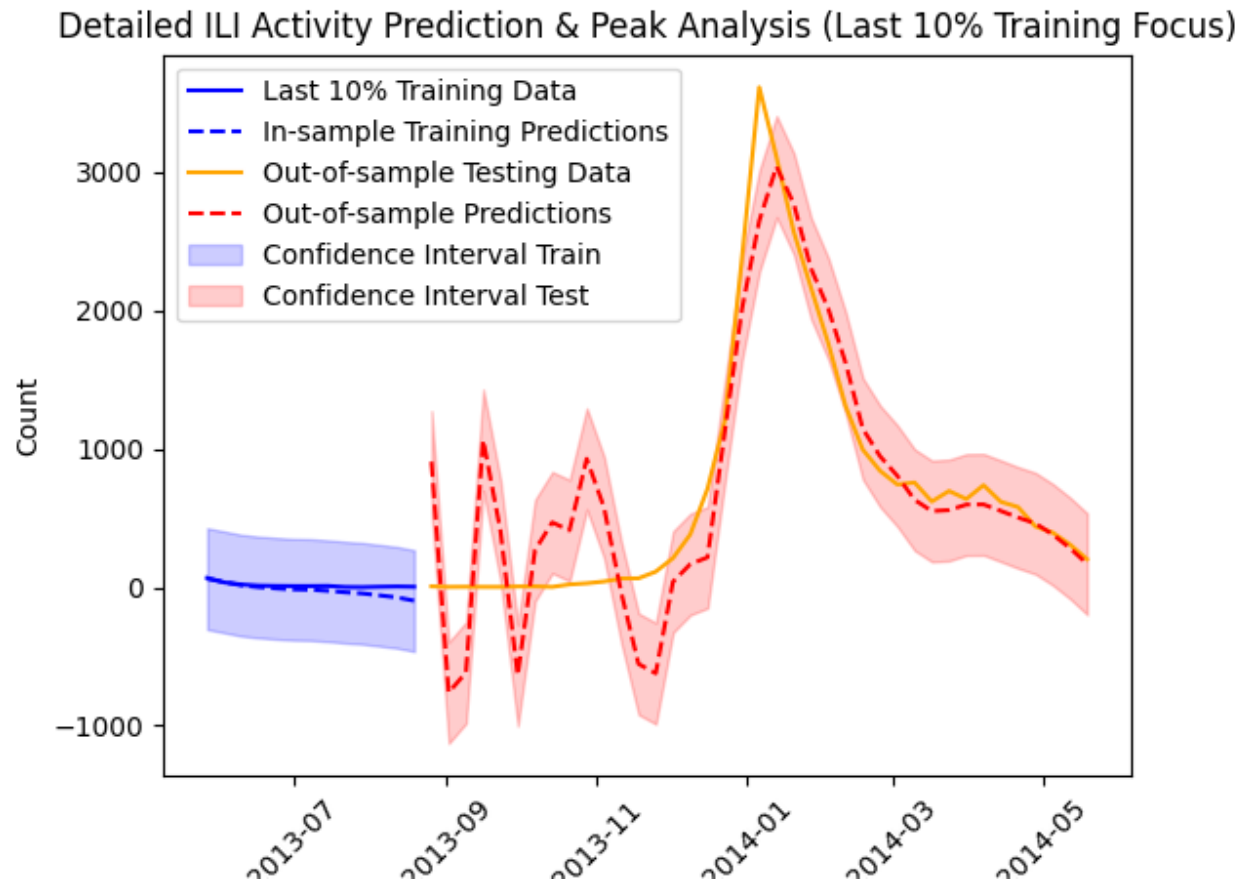


Figure A.116: 2013-2014 Season Prediction Focused on Last 10% of Training Data without Southern Hemisphere Covariate.

Detailed ILI Activity Prediction & Peak Analysis (Last 10% Training Focus)

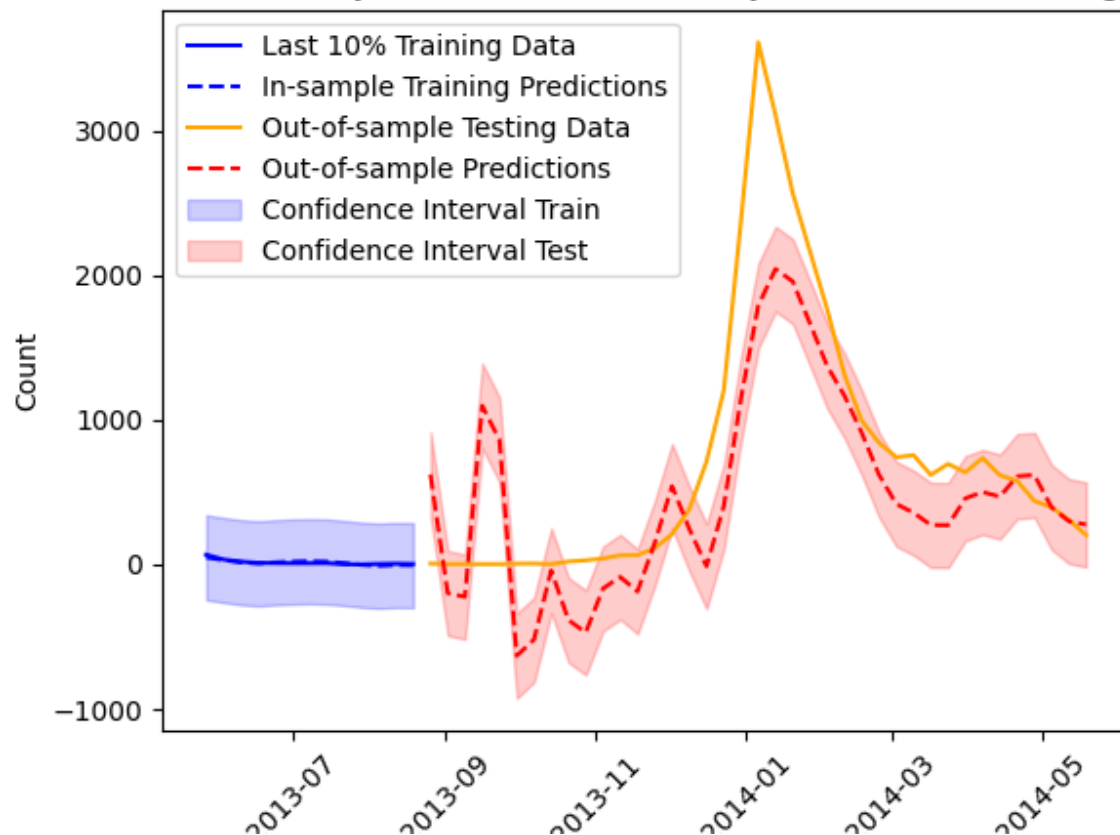


Figure A.117: 2013-2014 Season Prediction Focused on Last 10% of Training Data including Southern Hemisphere Covariate.

2013-2014 Season (Test Set Focus)

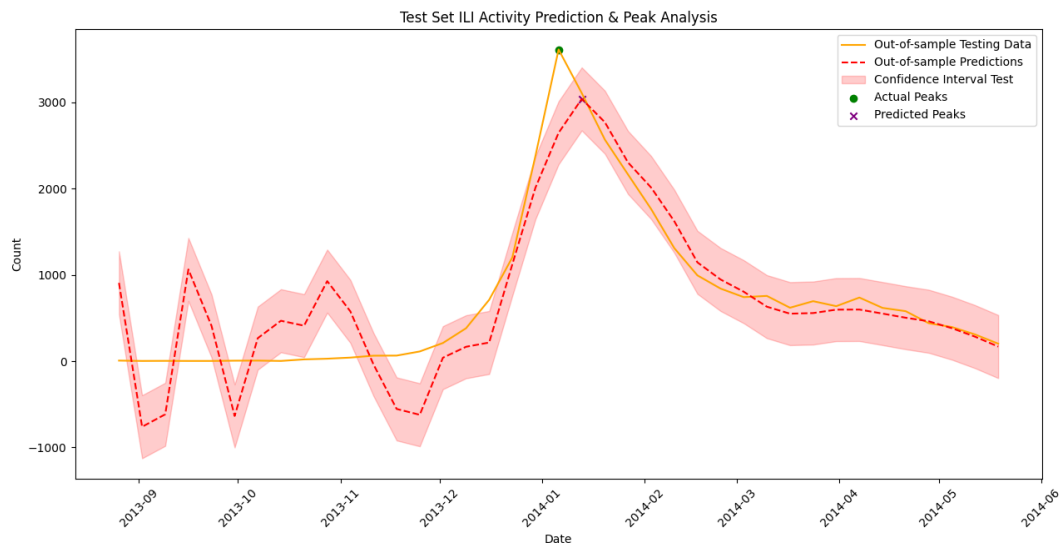


Figure A.118: 2013-2014 Season Prediction on Test Set without Southern Hemisphere Covariate.

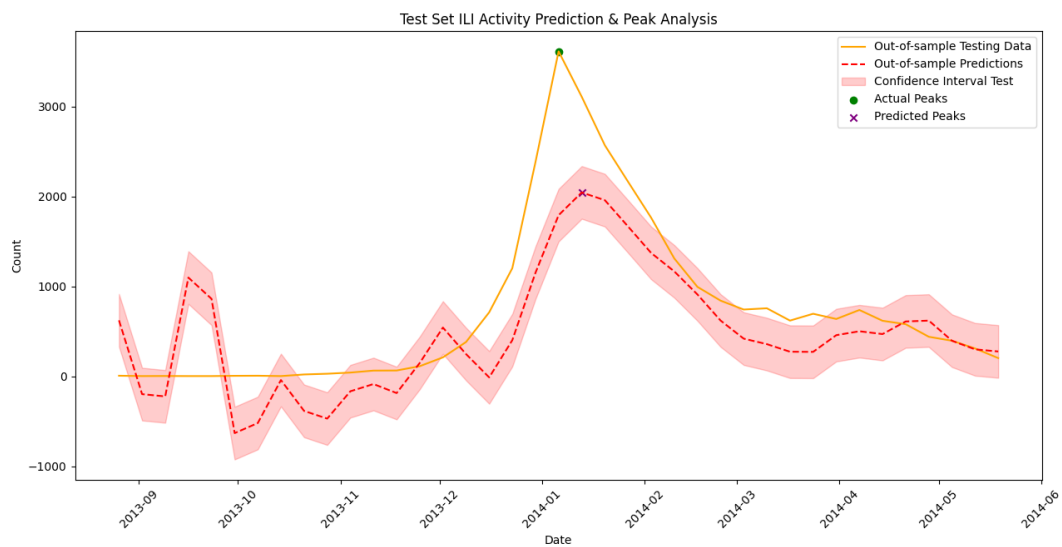


Figure A.119: 2013-2014 Season Prediction on Test Set including Southern Hemisphere Covariate.

References

- [1] R. Prado, M. A. R. Ferreira, and M. West, *Time Series: Modeling, Computation, and Inference, Second Edition*. CRC Press, 2021.
- [2] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and its Applications: With R examples, Fourth Edition*. Springer, 2017.
- [3] S. Sarkka, *Bayesian Filtering and Smoothing*. Cambridge University Press, 2013.
- [4] A. Papoulis, *Probability, Random Variables, and Stochastic Processes, Second Edition*. McGraw-Hill, 1994.
- [5] A. C. Harvey, *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, 1989.
- [6] J. J. Commandeur and S. J. Koopman, *An Introduction to State Space Time Series Analysis*. Oxford University Press, 2007.
- [7] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis, Third Edition*. CRC Press, 2013.
- [8] G. Chowell, “Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts,” *Infectious Disease Modelling*, vol. 2, no. 3, pp. 379–398, 2017.
- [9] G. Chowell, A. Bleichrodt, S. Dahal, *et al.*, “Growthpredict: A toolbox and tutorial-based primer for fitting and forecasting growth trajectories using phenomenological growth models,” *Scientific Reports*, vol. 14, no. 1630, 2024.
- [10] P. Yan and G. Chowell, *Quantitative Methods for Investigating Infectious Disease Outbreaks*. Springer, 2020.
- [11] J. Durbin and S. J. Koopman, *Time Series Analysis by State Space Methods, Second Edition*. Oxford University Press, 2012.
- [12] O. Cappe, E. Moulines, and T. Ryden, *Inference in Hidden Markov Models*. Springer, 2005.

- [13] H. Kunsch, *State space and hidden Markov models*. In Barndorff-Nielsen O., Cox D. and Kluppelberg C. (Eds.), *Complex stochastic systems*. Chapman and Hall/CRC, 2001, pp. 109–173.
- [14] G. Petris, S. Petrone, and P. Campagnoli, *Dynamic Linear Models with R*. Springer, 2009.
- [15] M. West and J. Harrison, *Bayesian Forecasting and Dynamic Models, Second Edition*. Springer, 1997.
- [16] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. Academic Press, 1970.
- [17] N. Chopin and O. Papaspiliopoulos, *An Introduction to Sequential Monte Carlo*. Springer, 2020.
- [18] G. H. Givens and J. A. Hoeting, *Computational Statistics, Second Edition*. Wiley, 2012.
- [19] H. E. Rauch, F. Tung, and C. T. Striebel, “Maximum likelihood estimates of linear dynamic systems,” *AIAA J*, vol. 3, no. 8, pp. 1445–1450, 1965.
- [20] A. Pankratz, *Forecasting with Dynamic Regression Models*. Wiley Series in Probability and Statistics, 1991.
- [21] Ontario Agency for Health Protection and Promotion (Public Health Ontario), *Ontario covid-19 data tool*, 2022. URL: <https://www.publichealthontario.ca/en/data-and-analysis/infectious-disease/covid-19-data-surveillance/covid-19-data-tool?tab=trends>.
- [22] Ontario Agency for Health Protection and Promotion (Public Health Ontario), *Covid-19 wastewater surveillance in ontario*, 2022. URL: <https://www.publichealthontario.ca/en/Data-and-Analysis/Infectious-Disease/COVID-19-Data-Surveillance/Wastewater>.
- [23] S. Haykin, *Adaptive Filter Theory, Fourth Edition*. Prentice Hall, 2001.
- [24] R. H. Shumway and D. S. Stoffer, “An approach to time series smoothing and forecasting using the em algorithm,” *Journal of Time Series Analysis*, vol. 81, no. 3, pp. 253–264, 1982.
- [25] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. John Wiley, 1958.
- [26] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*. Chapman and Hall/CRC Press, 1999.

- [27] A. C. Harvey, “Analysis and generalisation of a multivariate exponential smoothing model,” *Management Science*, vol. 32, pp. 374–380, 1986.
- [28] J. M. Quintana and M. West, “An analysis of international exchange rates using multivariate dlm,” *The Statistician*, vol. 36, no. ... Pp. 275–281, 1987.
- [29] D. Barber, A. Cemgil, and S. Chiappa, *Bayesian Time Series Models*. Cambridge University Press, 2011.
- [30] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Transactions of the ASME – Journal of Basic Engineering (Series D)*, vol. 82, pp. 35–45, 1960.
- [31] R. J. Meinhold and N. D. Singpurwalla, “Understanding the kalman filter,” *The American Statistician*, vol. 37, no. 2, 1983.
- [32] V. Dukic, H. Lopes, and N. Polson, “Tracking epidemics with google flu trends data and a state-space seir model,” *J Am Stat Assoc*, vol. 107, no. 500, pp. 1410–1426, 2012.
- [33] D. Osthus, K. Hickmann, P. Caragea, D. Higdon, and S. Del Valle, “Forecasting seasonal influenza with a state-space sir model,” *Ann Appl Stat*, vol. 11, no. 1, pp. 202–224, 2017.
- [34] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, *Handbook of Markov Chain Monte Carlo*. CRC Press, 2011.
- [35] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo methods in practice*. Springer, 2001.
- [36] J. Liu and R. Chen, “Sequential monte carlo methods for dynamic systems,” *Journal of the American Statistical Association*, vol. 443, no. 93, pp. 1032–1044, 1998.
- [37] H. Migon, D. Gamerman, H. Lopez, and M. Ferreira, “Bayesian dynamic models,” in *Handbook of Statistics*, D. Day and C. Rao, Eds., vol. 19, Elsevier B.V., 2005, pp. 553–588.
- [38] C. M. Carvalho, N. G. Polson, and J. G. Scott, “The horseshoe estimator for sparse signals,” *Biometrika*, vol. 97, no. 2, pp. 465–480, 2010.
- [39] D. Gamerman and H. F. Lopes, *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference, Second Edition*. CRC Press, 2006.
- [40] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, *Novel approach to nonlinear/non-gaussian bayesian state estimation*, 1993. URL: [doi:10.1049/ip-f-2.1993.0015](https://doi.org/10.1049/ip-f-2.1993.0015).

- [41] M. K. Pitt and N. Shephard, “Filtering via simulation: Auxiliary particle filters,” *Journal of the American Statistical Association*, vol. 94, no. 446, pp. 590–599, 1999.
- [42] A. Doucet, N. De Freitas, K. Murphy, and S. Russell, “Rao–blackwellised particle filtering for dynamic bayesian networks,” *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, pp. 176–183, 2000.
- [43] R. Chen and J. Liu, “Mixture kalman filter,” *Journal of the Royal Statistical Society, Series B*, vol. 62, pp. 493–508, 2000.
- [44] C. P. Robert and G. O. Roberts, “Rao–blackwellization in the mcmc era,” *To appear in The International Statistical Review*, 2021. URL: <https://doi.org/10.48550/arXiv.2101.01011>.
- [45] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Transactions of the ASME–Journal of Basic Engineering*, vol. Series D, no. 82, pp. 35–45, 1960.
- [46] A. Kong, J. S. Liu, and W. H. Wong, “Sequential imputations and bayesian missing data problems,” *Journal of the American Statistical Association*, vol. 89, no. 425, pp. 278–288, 1994.
- [47] N. Chopin, “Central limit theorem for sequential monte carlo and its application to bayesian inference,” *Annals of Statistics*, vol. 32, no. 6, pp. 2385–2411, 2004.
- [48] P. Del Moral, *Feynman-Kac Formulae*. Springer, 2004.
- [49] J. Liu and M. West, “Combined parameter and state estimation in simulation-based filtering,” in *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. de Freitas, and N. Gordon, Eds., Springer, 2001.
- [50] R. P. Adams and D. J. C. MacKay, *Bayesian online changepoint detection*, 2007. URL: <https://arxiv.org/abs/0710.3742>.
- [51] M. West, “On scale mixtures of normal distributions,” *Biometrika*, pp. 646–648, 1987.
- [52] T. C. Hsiang, “A bayesian view on ridge regression,” *The Statistician*, vol. 24, no. 4, p. 267, 1975.
- [53] N. Shephard, “Partial non-gaussian state space,” *Biometrika*, vol. 81, no. 1, pp. 115–131, 1994.
- [54] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [55] T. Park and G. Casella, “The bayesian lasso,” *Journal of the American Statistical Association*, vol. 103, no. 482, pp. 681–686, 2008.

- [56] S. Van Erp, D. L. Oberski, and J. Mulder, “Shrinkage priors for bayesian penalized regression,” *Journal of Mathematical Psychology*, vol. 89, pp. 31–50, 2019.
- [57] Y. Sugita and Y. Okamoto, “Replica-exchange molecular dynamics method for protein folding simulation,” *Chemical Physics Letters*, vol. 314, no. 1–2, pp. 141–151, 1999. DOI: [10.1016/S0009-2614\(99\)01123-9](https://doi.org/10.1016/S0009-2614(99)01123-9).
- [58] R. Killick and I. A. Eckley, “Changepoint: An r package for changepoint analysis,” *Journal of Statistical Software*, vol. 58, no. 3, pp. 1–19, 2014.
- [59] C. Erdman and J. W. Emerson, “Bcp: An r package for performing a bayesian analysis of change point problems,” *Journal of Statistical Software*, vol. 23, pp. 1–13, 2008.
- [60] J. Liu, B. Bellows, X. Hu, J. Wu, Z. Zhou, C. Soteris, and L. Wang, “A new time-varying coefficient regression approach for analyzing infectious disease data,” *Scientific Reports*, vol. 13, 2023.
- [61] L. M. A. Bettencourt, R. M. Ribeiro, G. Chowell, T. Lant, and C. Castillo-Chavez, “Towards real time epidemiology: Data assimilation, modeling and anomaly detection of health surveillance data streams,” in *Intelligence and Security Informatics: Biosurveillance*, ser. Lecture Notes in Computer Science, D. Zeng and et al., Eds., vol. 4506, Springer, Berlin, Heidelberg, 2007, pp. 79–90.
- [62] E. Kürüm, J. L. Warren, C. Schuck-Paim, R. Lustig, J. A. Lewnard, R. Fuentes, C. A. W. Bruhn, R. J. Taylor, L. Simonsen, and D. M. Weinberger, “Bayesian model averaging with change points to assess the impact of vaccination and public health interventions,” *Epidemiology*, vol. 28, no. 6, pp. 889–897, 2017.
- [63] F. Brauer, C. Castillo-Chavez, and Z. Feng, *Mathematical Models in Epidemiology*. Springer, 2019.
- [64] O. Diekmann and J. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases*. Wiley, 2000.
- [65] H. Hethcote, “The mathematics of infectious diseases,” *SIAM Rev.*, vol. 42, no. 4, pp. 599–653, 2000.
- [66] P. Van den Driessche and J. Watmough, “Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission,” *Mathematical Biosciences*, vol. 180, pp. 29–48, 2002.
- [67] J. Keller, T. Zhou, A. Kaplan, G. Anderson, and W. Zhou, “Tracking the transmission dynamics of covid-19 with a time-varying coefficient state-space model,” *Statistics in Medicine*, vol. 41, no. 15, pp. 2745–2767, 2022.

- [68] R. Winkelmann, *Econometric Analysis of Count Data, Fifth Edition*. Springer, 2008.
- [69] M. Zhou, L. Li, D. Dunson, and L. Carin, “Lognormal and gamma mixed binomial regression,” in *Proceedings of the 29th International Conference on Machine Learning*, 2012.
- [70] G. S. Kimeldorf and G. Wahba, “A correspondence between bayesian estimation on stochastic processes and smoothing splines,” *The Annals of Mathematical Statistics*, vol. 41, no. 2, pp. 495–502, 1970.
- [71] G. Riutort-Mayol, P.-C. Bürkner, R. A. Andersen, A. Solin, and A. Vehtari, *Practical hilbert space approximate bayesian gaussian processes for probabilistic programming*, 2020. URL: <https://arxiv.org/abs/2004.11408>.
- [72] X. He, E. Lau, P. Wu, *et al.*, “Temporal dynamics in viral shedding and transmissibility of covid-19,” *Nat Med*, vol. 26, no. 5, pp. 672–675, 2020.
- [73] Stan development team, *Rstan: The r interface to stan, r package version 2.21.3*, 2021. URL: <https://mc-stan.org/>.
- [74] Stan development team, *Stan modeling language users guide and reference manual, version 2.29*, 2022. URL: <https://mc-stan.org/>.
- [75] Ontario Data Catalogue, Ministry of Health, *Covid-19 vaccine data in ontario*, 2022. URL: <https://data.ontario.ca/dataset/covid-19-vaccine-data-in-ontario>.
- [76] Google COVID-19 Community Mobility Reports, *Google covid-19 community mobility reports*, 2022. URL: <https://www.google.com/covid19/mobility/>.
- [77] B. Althouse, E. Wenger, J. Miller, *et al.*, “Superspreading events in the transmission dynamics of sars-cov-2: Opportunities for interventions and control,” *PLoS Biol*, vol. 18, no. 11, e3000897, 2020.
- [78] M. Bashir, B. Ma, B. Komal, M. Bashir, D. Tan, and M. Bashir, “Correlation between climate indicators and covid-19 pandemic in new york, usa,” *Sci Total Environ*, vol. 728, p. 138 835, 2020.
- [79] A. Frontera, L. Cianfanelli, K. Vlachos, G. Landoni, and G. Cremona, “Severe air pollution links to higher mortality in covid-19 patients: The double-hit hypothesis,” *J Infect*, vol. 81, no. 2, pp. 255–259, 2020.
- [80] R. Anderson and R. May, *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, 1991.

- [81] K. C. Chan, G. A. Karolyi, F. A. Longstaff, and A. B. Sanders, “An empirical comparison of alternative models of the short-term interest rate,” *The Journal of Finance*, vol. 47, pp. 1209–1227, 1992.
- [82] A. Laaribi, B. Boukanjime, M. El Khalifi, D. Bouggar, and M. El Fatini, “A generalized stochastic sirs epidemic model incorporating mean-reverting ornstein–uhlenbeck process,” *Physica A: Statistical Mechanics and its Applications*, vol. 615, p. 128 609, 2023.
- [83] O. N. Bjornstad, B. F. Finkenstadt, and B. T. Grenfell, “Dynamics of measles epidemics: Estimating scaling of transmission rates using a time series sir model,” *Ecological Monographs*, vol. 72, no. 2, pp. 169–184, 2002.
- [84] A. Gray, D. Greenhalgh, L. Hu, X. Mao, and J. Pan, “A stochastic differential equation sis epidemic model,” *SIAM Journal on Applied Mathematics*, vol. 71, no. 3, pp. 876–902, 2011.
- [85] W. Wang, Y. Cai, Z. Ding, and Z. Gui, “A stochastic differential equation sis epidemic model incorporating ornstein–uhlenbeck process,” *Physica A: Statistical Mechanics and its Applications*, vol. 509, pp. 921–936, 2018.
- [86] A. Bertoune, M. El Fatini, M. El Khalifi, and A. Laaribi, “A stochastic hepatitis b model with infinite lévy jumps,” *Physica A: Statistical Mechanics and its Applications*, vol. 624, p. 128 830, 2023.
- [87] J. C. Cox, J. E. Ingersoll, and S. A. Ross, “A theory of the term structure of interest rates,” *Econometrica*, vol. 53, no. 2, pp. 385–407, 1985.
- [88] I. Karatzas and S. E. Shreve, *Brownian motion and stochastic calculus* (Graduate Texts in Mathematics), 2nd. Springer-Verlag, 1991, vol. 113.
- [89] O. Kallenberg, *Foundations of Modern Probability*, 2nd. Springer, 2002.
- [90] O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz, “On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations,” *Journal of Mathematical Biology*, vol. 28, no. 4, pp. 365–382, 1990.
- [91] K. Dietz, “The estimation of the basic reproduction number for infectious diseases,” *Statistical Methods in Medical Research*, vol. 9, no. 4, pp. 365–382, 2000.
- [92] P. van den Driessche and J. Watmough, “Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission,” *Mathematical Biosciences*, vol. 180, pp. 29–48, 2002.

- [93] Y. Mishura, K. Ralchenko, and O. Dehtiar, “Parameter estimation in ckls model by continuous observations,” *Statistics & Probability Letters*, vol. 182, p. 109 297, 2022.
- [94] Y. Hu, G. Lan, and C. Zhang, “The explicit solution and precise distribution of ckls model under girsanov transform,” *International Journal of Statistics and Probability*, vol. 4, no. 1, pp. 1–15, 2015, ISSN: 1927-7032.
- [95] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, 3rd. Springer, 1999.
- [96] L. C. G. Rogers and D. Williams, *Diffusions, Markov Processes and Martingales, Volume 1: Foundations* (Cambridge Mathematical Library). Cambridge University Press, 2000.
- [97] L. C. G. Rogers and D. Williams, *Diffusions, Markov Processes and Martingales, Volume 2: Itô Calculus* (Cambridge Mathematical Library). Cambridge University Press, 2000.
- [98] S. Sarkka and A. Solin, *Applied Stochastic Differential Equations*. Cambridge University Press, 2019.
- [99] M. Turelli, “Random environments and stochastic calculus,” *The American Naturalist*, vol. 111, no. 982, pp. 465–478, 1977.
- [100] X. Mao, *Stochastic Differential Equations and Applications*, 2nd. Woodhead Publishing, 2013.
- [101] C. Gouriéroux and P. Valéry, *Estimation of a jacobi process*, Presented at the 2004 conference of the Société Canadienne de Science Économique (SCSE), 2004. URL: http://www.scse.ca/scse/congres2004/articles/Valery_Gourieroux.pdf.
- [102] Y. Cai, J. Jiao, Z. Gui, Y. Liu, and W. Wang, “Environmental variability in a stochastic epidemic model,” *Applied Mathematics and Computation*, vol. 329, pp. 210–226, 2018.
- [103] D. J. Higham, “An algorithmic introduction to numerical simulation of stochastic differential equations,” *SIAM Rev.*, vol. 43, pp. 525–546, 2001.
- [104] D. J. Higham and P. E. Kloeden, *An Introduction to the Numerical Simulation of Stochastic Differential Equations*. SIAM, 2021.
- [105] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition. Springer, 2009.
- [106] B. Efron and T. Hastie, *Computer Age Statistical Inference*. Cambridge University Press, 2016.

- [107] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [108] M. Kim, Y. Kim, and K. Nah, “Predicting seasonal influenza outbreaks with regime shift-informed dynamics for improved public health preparedness,” *Scientific Reports*, vol. 14, p. 12 698, 2024. DOI: [10.1038/s41598-024-63573-z](https://doi.org/10.1038/s41598-024-63573-z). URL: <https://doi.org/10.1038/s41598-024-63573-z>.
- [109] R. A. Morbey, D. Todkill, C. Watson, and A. J. Elliot, “Machine learning forecasts for seasonal epidemic peaks: Lessons learnt from an atypical respiratory syncytial virus season,” *PLOS ONE*, vol. 18, no. 9, e0291932, 2023. DOI: [10.1371/journal.pone.0291932](https://doi.org/10.1371/journal.pone.0291932). URL: <https://doi.org/10.1371/journal.pone.0291932>.
- [110] Y. Zhang, L. Yakob, M. B. Bonsall, and W. Hu, “Predicting seasonal influenza epidemics using cross-hemisphere influenza surveillance data and local internet query data,” *Scientific Reports*, vol. 9, p. 3262, 2019. DOI: [10.1038/s41598-019-39871-2](https://doi.org/10.1038/s41598-019-39871-2). URL: <https://doi.org/10.1038/s41598-019-39871-2>.
- [111] World Health Organization, *FluNet: Data portal, global influenza programme*, <https://www.who.int/tools/flunet>, 2024.
- [112] PyTorch Contributors, *PyTorch: An open source machine learning framework*, <https://pytorch.org/>, Accessed on September 1, 2024.
- [113] TensorFlow Developers, *TensorFlow: Large-scale machine learning on heterogeneous systems*, <https://www.tensorflow.org/>, Accessed on September 1, 2024.
- [114] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010. DOI: [10.1109/TKDE.2009.191](https://doi.org/10.1109/TKDE.2009.191).
- [115] G. Hinton, O. Vinyals, and J. Dean, *Distilling the knowledge in a neural network*, 2015. arXiv: [1503.02531](https://arxiv.org/abs/1503.02531).
- [116] G. A. Pavliotis, *Stochastic Processes and Applications: Diffusion Processes, the Fokker–Planck and Langevin Equations*. Springer, 2014.
- [117] F. C. Klebaner, *Introduction to Stochastic Calculus with Applications*, 3rd. World Scientific Publishing, 2012.