Bayesian Dynamical Modeling of Count Data

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of the Ohio State University

By
Lili Zhuang, M.S.
Graduate Program in Statistics

The Ohio State University
2011

Dissertation Committee:
Professor Noel Cressie, Advisor
Professor Laura Kubatko
Professor Desheng Liu
ABSTRACT

Problems involving discrete data, especially count data, are increasingly common in many important fields, such as cancer mapping and influenza epidemiology. Compared to a large amount of highly developed attractive models for spatio-temporal continuous data (e.g., Cressie and Wikle, 2011), modeling the underlying dynamical process for count data is less well advanced.

Thus, the primary goal that ties together the two main chapters of this dissertation, is to develop dynamical approaches for better capturing the true process that underlies count data. Typically, the statistical dependence in the underlying process can be defined through a mathematical graph consisting of nodes (vertices) and edges. Nodes represent individuals or objects, while edges represent the (dependence) relationships between them. Mathematical graphs can be further divided into different classes based on the properties of their edges and the paths formed by edges (e.g., Lauritzen, 1996). In this dissertation, we use graphs that define spatio-temporal dependence (Chapter 2) and temporal dependence (Chapter 3).

Specifically, we start with spatio-temporal count data in the field of non-contagious disease mapping, namely, yearly sudden infant death syndrome (SIDS) information, from 1979 to 1984, for the counties of North Carolina. These data have been analyzed before as temporally aggregated spatial data (Cressie and Chan, 1989). We incorporate the new temporal aspect by presenting a spatio-temporal model from which
optimal smoothing of SIDS rates can be derived. Specifically, we use a Bayesian hierarchical statistical model (BHM) with a hidden dynamical Markov random field and extra-Poisson variability. The graph arises by evolving the Markov random field via an autoregressive matrix. Potential confounding of sources of variability is avoided by calibrating the extra-Poisson variability with the microscale variation in an approximate Gaussian model.

We also consider temporal (but non-spatial) count data resulting from a pandemic influenza outbreak (e.g., H1N1). The classic SIR (susceptible-infectious-recovered) model, which has been used extensively to study the dynamical process of an infectious disease in a large population, assumes observed counts are “mass balanced.” Here, mass balance means that total count equals the sum of counts of the individual components of the model. However, since the observed counts have errors, we assign the mass balance to the hidden process of a (Bayesian) hierarchical SIR (HSIR) model. Another challenge is to capture the stochastic or random nature of the epidemic process, yet still retain the mass-balance constraint. The HSIR model accomplishes this through modeling the dynamics on a transformed scale. To capture bias due to the transformation, we extend the HSIR model to include an extra level of randomness in the hidden process. Through simulation, we compare the HSIR model and its extension to the classic SIR model (where the mass balance appears on the observed counts). In these processes, the graph reduces to an acyclic directed graph.
DEDICATION

This dissertation is dedicated to my parents, Youming Zhuang, and Qiongfang Luo,
and my husband, Yunwei Qi.
ACKNOWLEDGMENTS

My deep and sincere gratitude first and foremost goes to my advisor, Dr. Noel Cressie, for his inspirational and patient guidance, general support, and unselfish help. Without his encouragement and constructive comments, I could never finished my dissertation successfully.

I want to thank members of my committee, Dr. Laura Kubatko, Dr. Desheng Liu, and Dr. Xinyi Xu, for their suggestions, especially for to look for a temperature effect in the research of Chapter 2. I also want to thank Dr. Laura Pomeroy for providing useful prior information and references in the research of Chapter 3.

My gratitude also goes to Dr. Chris Holloman and Jeni Squiric, for their guidance and help during my two years as a Statistical Consulting Service RA; Dr. Elizabeth Stasny and Dr. Doug Wolfe, for their help throughout my PhD study; Dr. William Venables, Dr. Yougan Wang, and Jay Harrison, for their kind support and valuable advice that made my two internships productive and exciting.

Special thanks go to my friends Ruoxi Xu, Emily L. Kang, and Jingyuan Yang, who gave me so much kindly help during my time at OSU.

Last but not least, I would like to thank my parents and my husband, for their love and encouragement. Thanks for always being there for me and supporting me in all my pursuits.
VITA

2006 ................................. B.S., Mathematical Statistics, Shanghai University of Finance and Economics, Shanghai, China

2008 ................................. M.S., Statistics, The Ohio State University, Columbus, Ohio

2007-2010 ............................ Graduate Teaching Assistant, The Ohio State University, Columbus, Ohio

2007-2010 ............................ Consultant, Statistical Consulting Service, The Ohio State University, Columbus, Ohio

2010-2011 ............................ Graduate Research Assistant, The Ohio State University, Columbus, Ohio

PUBLICATIONS


FIELDS OF STUDY

Major Field: Statistics
# TABLE OF CONTENTS

Abstract ................................................................. ii
Dedication ................................................................. iii
Acknowledgments ......................................................... v
Vita .......................................................................... vi
List of Tables ............................................................ ix
List of Figures ........................................................... x

Chapters:
1 Introduction ......................................................... 1
2 Spatio-Temporal Modeling of Sudden Infant Death Syndrome Data .. 5
   2.1 Introduction ................................................... 5
   2.2 Exploratory Data Analysis of Yearly SIDS Data ............... 7
      2.2.1 Data ...................................................... 7
      2.2.2 The Temporal Variability .............................. 10
      2.2.3 The Large-Scale Variation ............................... 12
      2.2.4 The Small-Scale Variation ............................... 14
   2.3 Model .......................................................... 17
      2.3.1 Data Model .............................................. 17
      2.3.2 Process Model .......................................... 18
      2.3.3 Parameter Model ....................................... 22
      2.3.4 Prior Distribution on Extra-Poisson-Variation Component 23
   2.4 Results ....................................................... 27
      2.4.1 Fitting the Hierarchical Spatio-Temporal CAR Model .... 27
      2.4.2 Posterior Analysis ..................................... 33
   2.5 Discussion and Conclusions .................................. 40
3 Bayesian Hierarchical Statistical SIR Models ....................... 42
   3.1 Introduction .................................................. 42
### Chapter 3

#### 3.2 Bayesian Hierarchical Statistical SIR (HSIR) Models
- 3.2.1 The HSIR Model
- 3.2.2 An Extension of the HSIR Model

#### 3.3 W-Scale Approximations for HSIR and Extended HSIR Models
- 3.3.1 A Linear Approximation to the HSIR Model
- 3.3.2 A Linear Approximation to the Extended HSIR Model

#### 3.4 Results of Analysis on Simulated Epidemics
- 3.4.1 Simulated Data
- 3.4.2 Fitting the ASIR and Extended ASIR Models
- 3.4.3 Posterior Analysis and Forecasting

#### 3.5 Discussion and Conclusions

---

### Appendices

#### A Details of Deriving the Prior Distribution on the Extra-Poisson-Variation Component of Chapter 2
- A.1
- A.2
- A.3
- A.4
- A.5

#### B Details of Deriving the Nonlinear W-Scale Processes and the Linear Approximations of HSIR and Extended HSIR Models of Chapter 3
- B.1
- B.2
- B.3
- B.4

### Bibliography
<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>30</td>
</tr>
<tr>
<td>3.1</td>
<td>58</td>
</tr>
<tr>
<td>3.2</td>
<td>76</td>
</tr>
<tr>
<td>3.3</td>
<td>77</td>
</tr>
<tr>
<td>3.4</td>
<td>77</td>
</tr>
<tr>
<td>3.5</td>
<td>77</td>
</tr>
<tr>
<td>3.6</td>
<td>78</td>
</tr>
<tr>
<td>3.7</td>
<td>78</td>
</tr>
<tr>
<td>3.8</td>
<td>79</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>An example of the type of chain graph used in Chapter 2 with spatio-temporal dependencies. Nodes within the same time may have undirected edges among them if they are neighbours. Nodes at adjacent time points may have directed edges among them if they are geographically close.</td>
</tr>
<tr>
<td>1.2</td>
<td>An example of the type of ADG used in Chapter 3, where there is only temporal dependence.</td>
</tr>
<tr>
<td>2.1</td>
<td>A map with the 100 counties of North Carolina numbered in alphabetical order. The number of each county appears at the location of its county seat. (The locations of county seats are based on longitudes and latitudes instead of the Cartesian coordinates provided in Cressie, 1993). The four shaded counties contain a city whose NREL monthly temperature data is used to construct the covariate, “minimum winter temperature.”</td>
</tr>
<tr>
<td>2.2</td>
<td>Yearly choropleth maps for minimum winter temperature. The scale is in degrees centigrade, and the ranges in the legend correspond to seven percentile ranges: “&lt; 5%,” “5%-20%,” “20%-35%,” “35%-65%,” “65%-80%,” “80%-95%,” and “&gt; 95%,” respectively.</td>
</tr>
<tr>
<td>2.3</td>
<td>Diagnostic plots for the Poisson GLM defined in equations (2.2.3) and (2.2.4).</td>
</tr>
<tr>
<td>2.4</td>
<td>Robust empirical semivariogram based on the weighted residuals of a weighted linear regression model that regresses FTSMR on FTNR and lag-1-year MWT. Units on the horizontal axis are in miles.</td>
</tr>
<tr>
<td>2.5</td>
<td>Autocorrelation plot based on the weighted residuals of a weighted linear regression model that regresses FTSMR on FTNR and lag-1-year MWT. Dashed horizontal lines on the plot represent the 95% confidence limits for the sample autocorrelations.</td>
</tr>
</tbody>
</table>
2.6 A straight line fitted by regressing the robust empirical semivariograms on the spatial lags near the origin. The intercept of this line with the vertical axis is an estimate of \((\sigma_\xi^2)^2\). Units on the horizontal axis are in miles.

2.7 Histograms of posterior samples of \(\beta_0\), \(\beta_1\), and \(\beta_2\).

2.8 The goodness-of-fit measurements \(\{w_t(s_i)\}\) versus \(\{1/\sqrt{E_t(s_i)}\}\), for \(i = 1, ..., N\), and \(t = 2\), namely 1980-1981. A linear relationship is expected.

2.9 Posterior mean and 95% HPD credible intervals of yearly weighted average of posterior samples of \(\{\bar{\lambda}_t(s_i)\}\) defined in (2.4.2), plotted against \(t = 1, ..., 5\).

2.10 Boxplots of yearly weighted average of posterior samples of \(\{\bar{\mu}_t(s_i)\}\) defined in (2.4.3), plotted against \(t = 1, ..., 5\).

2.11 Boxplots of yearly weighted average of posterior samples of \(\{\bar{\eta}_t(s_i)\}\) defined in (2.4.4), plotted against \(t = 1, ..., 5\).

2.12 (a) Choropleth maps of raw SMRs for five years; (b) Choropleth maps of posterior median of relative risk \(\{\lambda_t(s_i)\}\) for five years.

2.13 Choropleth maps of posterior median of \(\{\mu_t(s_i)\}\) for five years. The ranges in the legend correspond to seven percentile ranges: “< 5%,” “5%-20%,” “20%-35%,” “35%-65%,” “65%-80%,” “80%-95%,” and “> 95%,” respectively.

2.14 Choropleth maps of posterior median of \(\{\eta_t(s_i)\}\) for five years. The ranges in the legend correspond to seven percentile ranges: “< 5%,” “5%-20%,” “20%-35%,” “35%-65%,” “65%-80%,” “80%-95%,” and “> 95%,” respectively.

3.1 Plots of counts \(\{Z(t)\}\) for the HSIR, ESIR, and NSIR datasets as a function of time. The vertical pink line in each plot emphasizes that the data are only available on days 1-35, and they are missing on days 36-45.

3.2 Trace plots for the case AH: (a) trace plot for \(\beta\); (b) trace plot for \(\gamma\); (c) trace plot for \(\sigma_{\xi S}^2\); (d) trace plot for \(\sigma_{\xi I}^2\); (e) trace plot for \(\phi\). Red lines in plots (a)-(e) indicate the true values of the parameters used for simulating the HSIR dataset, namely, \(\beta^*\), \(\gamma^*\), \(\sigma_{\xi S}^2\), \(\sigma_{\xi I}^2\), and \(\phi^*\), respectively.
3.3 Prior (red line) and Posterior (histogram) distributions of $\beta$ for (a) case AH; (b) case EH; (c) case AE; (d) case EE; (e) case AN; (f) case EN. In each plot, blue dashed lines indicate posterior 95% CIs; the green line indicates the true value. 76

3.4 Scatter plot of $\psi \left( Z_{rep}^{(k)}, \lambda^{(k)} \right)$ against $\psi \left( Z; \lambda^{(k)} \right)$ for the case AH 82

3.5 Estimated infectious counts obtained from fitting the ASIR model, the extended ASIR model, and the classic SIR model to the HSIR, ESIR, and NSIR datasets, as a function of time. The vertical pink line in each plot emphasizes that data are only available on days 1-35, and that they are missing on days 36-45. The blue and green lines are posterior medians of $\{\lambda_I(t)\}$ obtained from the ASIR model and the extended ASIR model, respectively. The black line is the estimate of the infectious counts from the classic SIR model. The pink stars give the true infectious counts. 85

3.6 Posterior quantities of $\{\lambda_I(t)\}$ obtained from the ASIR model and the extended ASIR model during the forecasting period from day 36 to day 45 (based on data from day 1 to day 35, from the respective datasets). Lines indicate the 0.025, 0.25, 0.5, 0.75 and 0.975 quantiles, while pink stars give the true values of $\{\lambda_I(t)\}$. 86

3.7 Estimated susceptible counts obtained from fitting the ASIR model, the extended ASIR model, and the classic SIR model to the HSIR, ESIR, and NSIR datasets, as a function of time. The pink vertical line in each plot emphasizes that data are only available on days 1-35, and that they are missing on days 36-45. The blue and green lines are posterior medians of $\{\lambda_S(t)\}$ obtained from the ASIR model and the extended ASIR model, respectively. The black line is the estimate of the susceptible counts from the classic SIR model. The pink stars give the true susceptible counts. 87

3.8 Posterior quantities of $\{\lambda_S(t)\}$ obtained from the ASIR model and the extended ASIR model during the forecast period from day 36 to day 45 (based on data from day 1 to day 35 of the respective datasets). Lines indicate the 0.025, 0.25, 0.5, 0.75 and 0.975 quantiles, while pink stars give the true values of $\{\lambda_S(t)\}$. 88

3.9 Estimates of $R_0$ obtained from the ASIR model, the extended ASIR model, and the classic SIR model fitted to each of the three datasets. Boxplots indicate posteriors of $R_0$ obtained from the ASIR and extended ASIR models. The black line gives the estimate of $R_0$ from the classic SIR model, the red lines give the posterior medians, and the green line gives the true value of $R_0$. 90
INTRODUCTION

The problem of capturing the underlying true process behind observed phenomena has arisen in many areas of science and has been addressed in the statistical literature through (Bayesian) hierarchical modeling. Such models involve multivariate statistical-dependence structures that can be represented through mathematical graphs.

Typically, a mathematical graph is defined as $G = (V, E)$, which consists of a finite set of nodes (or vertices) $V$, where nodes represent individuals or objects, and edges $E$, where edges specify their (dependence) relationships. Graphs can be further divided into different classes, according to the nature of their edges as well as the paths formed by edges; see Lauritzen (1996) for further details. In this dissertation, we focus on modeling the dependence through a type of graph, called a chain graph. A chain graph is a combination of both undirected graphs and acyclic directed graphs (ADGs, or sometimes abbreviated as DAGs), with the overall structure being guided by an ADG.

Specifically, an undirected graph is a graph that has only undirected edges, which can be used to define a symmetric dependence relationships between nodes. For example, a Markov random field (e.g., Besag, 1974), which is extensively used for specifying small-scale spatial dependence, derives its neighborhood structure from an undirected graph. On the contrary, an ADG is a graph that has only directed edges.
with no cycles, and it is widely used for capturing asymmetric dependence relationships, such as causal effects (e.g., Ellis and Wong, 2008), temporal dependence (e.g., Dechter et al., 1991), and so forth. Thus, by combining these two types of graphs, a chain graph has greater flexibility to define relationships that may consist of both symmetric and asymmetric dependence relationships, such as for spatio-temporal dependencies (e.g., Cressie and Wikle, 2011) and social networks (e.g., Kolaczyk, 2009).

Bayesian hierarchical statistical models (BHMs) that capture dependencies are featured in this dissertation. A hierarchical model typically consists of three components, namely, the conditional probability distribution of data given the process and parameters (data model), the conditional probability distribution of the process given parameters (process model), and the probability distribution of the parameters (parameter model, sometimes called the prior distribution). An important example of a BHM involves both a Gaussian data model and a Gaussian process model, for the purpose of modeling continuous data.

Models for discrete data, especially count data, are needed increasingly in important disciplines, such as non-contagious disease mapping, influenza epidemiology, traffic control, etc. In this dissertation, we develop BHM approaches for modeling count data with dependencies in the process model expressed through chain graphs and ADGs.

The chain graph is used in Chapter 2 to define the spatio-temporal dependencies in the hidden process for sudden infant death syndrome (SIDS) data. We assume that spatial dependence can only happen within “neighborhoods” (a neighborhood here contains nodes that are within a certain geographical distance from each other); and temporal dependence is assumed to be Markov. Therefore, the chain graph here consists of both undirected (spatial dependence) and directed edges (temporal dependence); an example is given in Figure 1.1.
Figure 1.1: An example of the type of chain graph used in Chapter 2 with spatio-temporal dependencies. Nodes within the same time may have undirected edges among them if they are neighbours. Nodes at adjacent time points may have directed edges among them if they are geographically close.

In Chapter 3, the ADG defines the underlying process in a BHM for modeling influenza epidemic data; here the dependence is purely temporal resulting in a graph with only directed edges. We assume that the dynamical flow in the process model is based on the classic SIR (susceptible-infectious-recovered) model; an example is given in Figure 1.2.
In this dissertation, we investigate the discrete multivariate distributions defined by graphs under the two different types of settings discussed above. We develop BHMs that model temporal or spatio-temporal count data directly, rather than using approximate Gaussian models on their transformed scales.

Dr. Noel Cressie proposed the problems solved in each chapter. Chapter 2 has been accepted for publication as: Zhuang, L. and Cressie, N. (2011). Spatio-Temporal Modeling of Sudden Infant Death Syndrome Data. Statistical Methodology, doi:10.1016/j.stamet.2011.01.006. Chapter 3 will form the basis of articles by L. Zhuang and N. Cressie that will be submitted for publication.
Chapters: 2

SPATIO-TEMPORAL MODELING OF SUDDEN INFANT DEATH SYNDROME DATA

2.1 Introduction

Sudden infant death syndrome (SIDS) is a classification of unexpected, sudden death in apparently healthy infants under one year old. Before the 1990s, SIDS resulted in about 7,000 infant deaths per year (i.e., 200 SIDS deaths per 100,000 live births) in the United States. Although the current rate of SIDS is declining dramatically, at present about 60 SIDS deaths per 100,000 live births per year occur in the United States (e.g., Best, 2009). Much of the early research on SIDS was physiological and focused on the relationship between SIDS and certain risk factors, such as apnea (Steinschneider, 1972), cold temperature (Goldberg and Stein, 1978), and poor nutrition (Fogerty et al., 1984). Even to the present day, the etiology of SIDS is largely a mystery.

Atkinson (1978) presented a data set that contained the number of SIDS and the number of live births from 1974-1978 in the 100 counties of North Carolina. Symons et al. (1983) augmented the data set and, based on independent Poisson assumptions, found possible geographical clustering that seemed to correlate with inhomogeneously distributed racial groups across North Carolina. Cressie and Read (1985) performed a spatial exploratory data analysis on the same data set and found both geographical
clustering and spatial dependence. Cressie and Chan (1989) used a Markov-random-field model of spatial dependence to analyze this data set, along with a similar one for the period 1979-1984. Although they confirmed a dependence of SIDS rate on the proportion of non-white live births in the data aggregated over 1974-1978, they did not find dependence for data aggregated over 1979-1984.

Cressie and Chan (1989) accounted for spatial dependence and thus improved the overall fit, however, they did not consider the temporal aspect because the dataset was aggregated over time. In this chapter, the availability of yearly SIDS data for each of the 100 counties of North Carolina from July 1979 to June 1984, has allowed us to introduce a temporal aspect to the existing spatial analyses. For example, we can investigate an important possible (lag) effect caused by low winter temperatures. Hence, our spatio-temporal statistical analysis may suggest further physiological research and thus may be helpful from a public-health perspective.

Since we are able to study both the spatial and temporal variability of SIDS in North Carolina, we propose a hierarchical spatio-temporal CAR model for the yearly data from 1979-1984. In Section 2.2, we perform exploratory data analysis on the yearly SIDS data set for the 100 counties of North Carolina. We further construct and compare Poisson generalized linear models (GLMs) for the yearly data and the aggregated five-yearly data. Our exploratory analysis does not include spatial dependence initially. Important covariates in the Poisson GLM were found to be the one-year-lag minimum winter temperature and the non-white live birth rate. In Section 2.3, we propose a Bayesian hierarchical spatio-temporal CAR model for the yearly SIDS dataset, which includes a hidden dynamical Markov random field. Section 2.3.4 gives a calibration method based on a Gaussian approximation, to resolve the Poisson and extra-Poisson variability. The posterior analysis is performed in Section 2.4, followed by discussions and conclusions in Section 2.5.
2.2 Exploratory Data Analysis of Yearly SIDS Data

2.2.1 Data

The SIDS data set we analyze in this chapter contains yearly SIDS counts from 1979 to 1984 for the 100 counties of North Carolina. The data set also contains information for subgroups of the population of live births, stratified by race and gender, namely, “white male,” “white female,” “non-white male,” and “non-white female.” For each subgroup, we know the number of live births and the number of SIDS in the 100 counties of North Carolina, yearly from 1979 to 1984. Using the same labels as Cressie and Read (1985) and Cressie and Chan (1989), we number the counties from 1 to 100, as shown in Figure 2.1. The longitude and latitude information for each of the 100 county seats is also available (e.g., Bivand, 2010). Furthermore, for the four cities, Cape Hatteras, Charlotte, Greensboro, and Raleigh, whose counties are shaded in Figure 2.1, we have the monthly temperature data from 1961 to 1990 based on “The National Solar Radiation Data Base (1961-1990),” provided by the National Renewable Energy Laboratory (NREL, 1990). Notice that in our data set, year $t$ runs from July 1, 1978+$t$ through June 30, 1978+$t$+1, for $t=1,\ldots,5$. Hence, we define the average of the monthly minimum temperature of December of year $t$, January of year $t + 1$, and February of year $t + 1$ as “minimum winter temperature” for year $t$. We then used a Multilevel B-spline Approximation (MBA) algorithm (e.g., Lee et al., 1997, Finley and Banerjee, 2010) to estimate the yearly minimum winter temperature for each of the 100 counties during year $t$; $t = 0, 1, \ldots, 5$. (For this purpose, we consider a county centered at its county seat.)
Figure 2.1: A map with the 100 counties of North Carolina numbered in alphabetical order. The number of each county appears at the location of its county seat. (The locations of county seats are based on longitudes and latitudes instead of the Cartesian coordinates provided in Cressie, 1993). The four shaded counties contain a city whose NREL monthly temperature data is used to construct the covariate, “minimum winter temperature.”

We first introduce the following important notation:

\( t \) denotes a time point for \( t = 1, \ldots, T \). In our case, \( T = 5 \), and \( t=1, \ldots,5 \) represents years 1979-1980,...,1983-1984, respectively.

\( s_i \) denotes the location of a county, for \( i = 1, \ldots, N \). In our case, the location is the (latitude, longitude) coordinates of the county seats, and \( N = 100 \).

\( Z_t(s_i) \) denotes the observed number of SIDS in county \( i \) at time \( t \).

\( Z_{kt}(s_i) \) denotes the observed number of SIDS for subgroup \( k \) in county \( i \) at time \( t \), for \( k = 1, \ldots, K \). In our case \( K = 4 \), and \( k = 1, \ldots, 4 \) represents the subgroups, “white male,” “white female,” “non-white male,” and “non-white female,” respectively.

\( B_t(s_i) \) denotes the total number of live births in county \( i \) at time \( t \).
\( B_{kt}(s_i) \) denotes the number of live births for subgroup \( k \) in county \( i \) at time \( t \), for \( k = 1, \ldots, K \). Recall that in our case, \( K = 4 \), and the subgroups are the same as given just above.

\( E_t(s_i) \) denotes the number of SIDS expected to occur in county \( i \) at time \( t \), which we define using internal standardization:

\[
E_t(s_i) = \sum_{k=1}^{4} B_{kt}(s_i) \left( \sum_{j=1}^{N} Z_{kt}(s_j) / \sum_{j=1}^{N} B_{kt}(s_j) \right); \quad i = 1, \ldots, N, \; t = 1, \ldots, T. \tag{2.2.1}
\]

\( \text{SMR}_t(s_i) \) denotes the standardized mortality ratio (SMR) in county \( i \) at time \( t \); that is,

\[
\text{SMR}_t(s_i) \equiv Z_t(s_i) / E_t(s_i); \quad i = 1, \ldots, N, \; t = 1, \ldots, T.
\]

\( \text{FTSMR}_t(s_i) \) denotes the Freeman-Tukey transformed SMR in county \( i \) at time \( t \), defined as

\[
\text{FTSMR}_t(s_i) = \sqrt{1000 Z_t(s_i) / E_t(s_i)} + \sqrt{1000 (Z_t(s_i) + 1) / E_t(s_i)},
\]

\[
\quad \text{for } i = 1, \ldots, N, \; t = 1, \ldots, T. \tag{2.2.2}
\]

\( \text{FTNR}_t(s_i) \) denotes the Freeman-Tukey transformed non-white live birth rate in county \( i \) at time \( t \), defined as

\[
\text{FTNR}_t(s_i) = \sqrt{1000 \left[ B_{3t}(s_i) + B_{4t}(s_i) \right] / B_t(s_i)}
\]

\[
\quad + \sqrt{1000 \left[ B_{3t}(s_i) + B_{4t}(s_i) + 1 \right] / B_t(s_i)},
\]

\[
\quad \text{for } i = 1, \ldots, N, \; t = 1, \ldots, T. \quad \text{Recall that } B_{3t}(s_i) \text{ and } B_{4t}(s_i) \text{ are the numbers of non-white-male live births and non-white-female live births, respectively, in county } i \text{ at time } t.
\]
MWT$_t(s_i)$ denotes the minimum winter temperature in county $i$ at time $t$, for $i = 1,\ldots,N$, $t = 0, 1,\ldots,T$. Recall that these are estimated by a spline smoothing of average minimum temperatures in December, January, and February for the four cities in North Carolina (NREL, 1990).

As in Cressie and Read (1985), we are interested in the Freeman-Tukey (square-root) transformation of SMRs and non-white live birth rates. Not only does the Freeman-Tukey transformation solve a problem of very skewed data, it is also a variance-controlling transformation as shown in (A.1.6) in Appendix A.1. That is, using the delta method,

$$\text{var}(\text{FTSMR}_t(s_i)) \propto \frac{1}{E_t(s_i)}.$$  

However, unlike Cressie and Read (1985) and Cressie and Chan (1989), the model we propose in Section 3.2 is applied directly to the SIDS counts, although there are occasions when we need to use the Freeman-Tukey transformation.

### 2.2.2 The Temporal Variability

We first construct simple Poisson GLMs without assuming any spatio-temporal statistical dependence. This is done for the five-year-aggregated SIDS data set and for the yearly SIDS data sets. For each model, we assume that the mean of the Poisson GLM is equal to a product of the (known) expected number of SIDS $\{E_t(s_i)\}$ and the unknown relative risk parameter, the logarithm of which is a linear function of covariates. Our initial choice of covariates was motivated by the two most important SIDS risk factors proposed by previous research (e.g., Willinger, 1995; Byrd et al., 2005; Gilbert et al., 1992; etc.), namely, a baby’s sleeping position and the manner in which a baby is wrapped. Because we do not have any direct information on either factors, some proxies were chosen. Byrd et al. (2005) mentioned that non-white people tend to put their babies on their stomach to sleep, while white people
tend to put their babies on their back to sleep. Furthermore, it is well known that winter temperature could affect the manner in which people wrap babies, more so for lower winter temperatures. Therefore, we used (Freeman-Tukey transformed) non-white live birth rates (FTNR) and the minimum winter temperature (MWT) as two covariates. We also included year as a factor in the model to capture any obvious temporal trend. When fitting the five-yearly-aggregated SIDS data, the covariates were also aggregated.

In this paragraph, we summarize the results of our Poisson GLM model fitting. Two models were fitted, one where all the yearly count data are incorporated into one Poisson GLM and the variability is captured by a loglinear mean; and the other where the aggregated five-yearly count data are incorporated into a Poisson GLM and again the variability is captured by a log linear (aggregated) mean. For both models, the effects of FTNR and (non-lagged) MWT were not significant (the 95% confidence interval includes zero), however, the model based on yearly data indicates a very different behavior of SIDS in the third year. Choropleth maps of MWT for the five years, shown in Figure 2.2, indicate an unusually cold winter in the second year (1980-1981), which suggests a possible lag effect of MWT on SIDS. As we already mentioned in Section 2.1, some previous research did consider the relationship between SIDS and winter temperature in the current year (e.g., Goldberg and Stein, 1978); the hypothesis was based on “overbundling” or “overheating” of infants, or possibly the higher risk of infection that infants face during the cold winter months. However, the lag effect of MWT suggests an interesting hypothesis that pregnant women face a higher risk of infection during colder winter months, which may increase the risk of SIDS for babies born the following year.
2.2.3 The Large-Scale Variation

Notice that a spatio-temporal process usually consists of both large-scale and small-scale variation. We think of the large-scale variation as deterministic trend and, here, for count data, we use a Poisson GLM. An alternative would be to transform
Figure 2.3: Diagnostic plots for the Poisson GLM defined in equations (2.2.3) and (2.2.4).

the SIDS rates and fit a Gaussian model (e.g., Cressie and Chan, 1989). Based on the analysis in Section 2.2.2, we included FTNR and one-year-lagged (i.e., lag-1) MWT as covariates; we also tried including the covariate defined by the difference between lag-1 and current MWT, but it did not improve the model fit. Hence, we fitted the following Poisson GLM:

\[
Z_t(s_i) | \lambda_t(s_i) \sim \text{ind. Poisson}(E_t(s_i) \lambda_t(s_i)), \tag{2.2.3}
\]

where

\[
\log(\lambda_t(s_i)) = \beta_0 + \beta_1 \text{FTNR}_t(s_i) + \beta_2 \text{MWT}_{t-1}(s_i), \tag{2.2.4}
\]
for $i = 1, \ldots, N$, $t = 1, \ldots, T$. In (2.2.3) and (2.2.4), the quantities $Z_t(s_i)$, $E_t(s_i)$, 
$\text{FTNR}_t(s_i)$, and $\text{MWT}_{t-1}(s_i)$ are defined in Section 2.2.1; and $\log(E_t(s_i))$ is sometimes 
termed an offset.

Define $\lambda_t(s_i)$ to be the relative risk for county $i$ at time $t$, which accounts for a 
departure from the expected number of SIDS, and $\log(\lambda_t(s_i))$ is the log relative risk.
The diagnostic plots in Figure 2.3 indicate that the model does capture the large-scale 
variation.

### 2.2.4 The Small-Scale Variation

To perform exploratory data analysis on the small-scale variation, we expect the 
approximation derived in Appendix A.1, that considers a linear-model analysis on 
$\text{FTSMR}$, to capture approximately the same variation as a Poisson GLM analysis on 
SMR. We use data from all 100 counties for every year of the five available years to fit 
a simple linear model without using any spatio-temporal statistical dependence when 
carrying out the fit. In order to take into account the unequal variation resulting 
from unequal numbers of expected SIDS ($E_t(s_i)$), the linear model is weighted by 
$\sqrt{E_t(s_i)}$. That is,

$$
\text{FTSMR}_t^w(s_i) = (X_t^w(s_i))' \beta^w + \delta_t^w(s_i); \quad i = 1, \ldots, N, \quad t = 1, \ldots, T,
$$

(2.2.5)

where

$$
\text{FTSMR}_t^w(s_i) \equiv \sqrt{E_t(s_i)} \text{FTSMR}_t(s_i),
$$

(2.2.6)

$$
X_t^w(s_i) \equiv \sqrt{E_t(s_i)} X_t(s_i).
$$

(2.2.7)

In (2.2.7), $X_t(s_i) = (1, \text{FTNR}_t(s_i), \text{MWT}_{t-1}(s_i))'$; recall that the quantities $\text{FTNR}_t(s_i)$ 
and $\text{MWT}_{t-1}(s_i)$ are defined in Section 2.2.1. In (2.2.5), $\beta^w = (\beta_0^w, \beta_1^w, \beta_2^w)'$ is the 
vector of fixed effects, and $\delta_t^w(s_i)$ is the error term. Figure 2.4 shows a robust
Figure 2.4: Robust empirical semivariogram based on the weighted residuals of a weighted linear regression model that regresses FTSMR on FTNR and lag-1-year MWT. Units on the horizontal axis are in miles.

semivariogram estimator (Cressie, 1993, Section 2.2.2) based on the weighted residuals:

\[ R_t^w(s_i) \equiv \text{FTSMR}_t^w(s_i) - (X_t^w(s_i))' \hat{\beta}^w; \quad i = 1, ..., N, \ t = 1, ..., T, \quad (2.2.8) \]

where \( \hat{\beta}^w \) is the ordinary least squares estimator of \( \beta^w \) obtained from the \( NT \)-dimensional data vector whose elements are \( \{\text{FTSMR}_t^w(s_i) : i = 1, ..., N, \ t = \)
Specifically, we compute the estimator proposed by Cressie and Hawkins (1980):

\[
2\gamma(h) \equiv \frac{1}{0.457 + 0.494/(|N(h)|T)} \left\{ \sum_t \sum_{N(h)} |R_t^w(s_i) - R_t^w(s_j)|^{1/2} / |N(h)|T \right\}^4,
\]

where \(N(h) \equiv \{(i, j) : |s_i - s_j| = h\}\), \(h\) is the spatial lag, and \(|N(h)|\) is the number

![Small-Scale Temporal Correlation](image)

**Figure 2.5**: Autocorrelation plot based on the weighted residuals of a weighted linear regression model that regresses FTSMR on FTNR and lag-1-year MWT. Dashed horizontal lines on the plot represent the 95% confidence limits for the sample autocorrelations.
of distinct elements of $N(h)$. Figure 2.4 is a plot of the semivariogram $2\bar{\gamma}(h)$ versus $h$; we can see strong evidence of spatial dependence as the semivariogram increases from smaller spatial lags to level out at a “sill” value at larger spatial lags. Furthermore, we also adapted the formula in (2.2.9) so that we can compute yearly variograms, $2\bar{\gamma}_t(h)$, for $t = 1, ..., 5$; in all of the five semivariogram plots (not shown), structure similar to that shown in Figure 2.4 was seen. We conclude that the spatial dependence is consistent through time.

Another way to look for small-scale variation is through temporal dependence. Figure 2.5 plots autocorrelations (based on weighted residuals) versus temporal lag, from which it appears that the temporal dependence is not very strong.

### 2.3 Model

Assume that there is a true unobserved spatio-temporal process hidden behind the yearly North Carolina SIDS counts, which we incorporate into the framework of a Bayesian hierarchical statistical model. The spatial domain is discrete and consists of the 100 counties in North Carolina. Consequently, we propose to model the data on a spatial lattice of county centroids whose size is fixed at 100. Such models typically consist of three components, namely, the conditional probability distribution of data given processes and parameters (data model), the conditional probability distribution of processes given parameters (process model), and the probability distribution of parameters (parameter model).

#### 2.3.1 Data Model

Rather than transforming the data and fitting Gaussian models (Cressie and Chan, 1989), we model the counts directly. Specifically, the data model is assumed to
be a product of independent Poisson distributions (e.g., Mugglin et al., 2002). That is, for \( i = 1, \ldots, N, t = 1, \ldots, T, \)

\[
Z_t(s_i) | \lambda_t(s_i) \sim \text{ind. Poisson}(E_t(s_i)\lambda_t(s_i)),
\]

(2.3.1)

where recall that \( Z_t(s_i) \) and \( E_t(s_i) \) are the observed number of SIDS and the expected number of SIDS in the county located at \( s_i \) at time \( t \), respectively. Then \( \lambda_t(s_i) \) is the underlying relative risk parameter that accounts for a departure from the expected number of SIDS, \( E_t(s_i) \), at county \( i \) and time \( t \), and hence \( \lambda_t(s_i) \) is the hidden process of interest. The conditional distribution of \( \lambda_t(s_i) \) given parameters, namely the process model, is where the spatio-temporal dependence can be specified. Define

\[
\lambda_t(S) \equiv (\lambda_t(s_1), \ldots, \lambda_t(s_N))',
\]

which is an \( N \)-dimensional vector, and define \( \log \lambda_t(S) \) to be the elementwise logarithms.

### 2.3.2 Process Model

Our process model is:

\[
\log(\lambda_t(S)) = \mu_t(S) + \eta_t(S) + \xi_t(S),
\]

(2.3.2)

where we now discuss each of the components of (2.3.2), in turn. The vector \( \mu_t(S) \equiv (\mu_t(s_1), \ldots, \mu_t(s_N))' \) captures the large-scale variation, or spatial trend. In our case, we put

\[
\mu_t(S) = X_t(S)\beta,
\]

(2.3.3)

where \( X_t(S) \) is an \( N \times p \) matrix of fixed covariates that are specified as a result of our exploratory data analysis in Section 2.2.3. That is, \( X_t(S) \) is the \( N \times 3 \) matrix, \( X_t(S) = (1, \text{FTNR}_t(S), \text{MWT}_{t-1}(S)) \), where \( 1 \) is an \( N \)-dimensional column vector of ones, \( \text{FTNR}_t(S) \equiv (\text{FTNR}_t(s_1), \ldots, \text{FTNR}_t(s_N))' \), and \( \text{MWT}_{t-1}(S) \equiv \ldots \)
We also define $\beta = (\beta_0, \beta_1, \beta_2)'$ to be the vector of corresponding fixed effects. The $N$-dimensional vector, $\eta_t(S) \equiv (\eta_t(s_1), ..., \eta_t(s_N))'$, captures the small-scale variation which, from Section 2.2.4, captures any spatio-temporal statistical dependence. The marginal spatial dependence is modeled according to an inhomogeneous CAR model (e.g., Cressie and Kapat, 2008), however, the temporal dependence in $\eta_t(S)$ requires more structure. Motivated by Mugglin et al. (2002), we define $\eta_t(S)$ to be a multivariate Gaussian AR(1) process:

$$\eta_{t+1}(S) = H_{t+1}(S)\eta_t(S) + \zeta_{t+1}(S); \quad t = 1, ..., T - 1,$$

(2.3.4)

where the instantaneous spatial correlation, $\zeta_{t+1}(S) \equiv (\zeta_{t+1}(s_1), ..., \zeta_{t+1}(s_N))'$, is independent from $\eta_t(S)$ and is distributed according to the Gaussian Markov random field,

$$\zeta_{t+1}(S) \sim MVN(0, U_{t+1}(S)).$$

(2.3.5)

In (2.3.5), $0$ is an $N$-dimensional vector of zeros, and for $t = 1, ..., T$,

$$U_t(S) = \tau_t^2 [I - \phi_t C_t(S)]^{-1} M_t(S),$$

(2.3.6)

where $I$ is the $N \times N$ identity matrix, and

$$M_t(S) \equiv \text{diag} \left( E_t(s_1)^{-1}, ..., E_t(s_N)^{-1} \right).$$

(2.3.7)

This model for $U_t(S)$ is justified in Cressie et al. (2005). The $N \times N$ matrix, $C_t(S) = (c_{tij})$, captures the spatial dependence; we assume it has the following form:

$$c_{tij} = \begin{cases} C(k_t) d_{ij}^{-k_t} \left( \frac{E_t(s_j)}{E_t(s_i)} \right)^{\frac{1}{2}}, & \text{if } s_j \in N_t(s_i) \\ 0, & \text{otherwise}, \end{cases}$$

(2.3.8)

where $N_t(s_i)$ is a set of neighborhood counties of county $i$ at time $t$. In our case, the same 100 counties of North Carolina are involved at each time point and, based on the yearly semivariogram plots referred to earlier, we assume that the neighborhood
structure does not change over time; that is, \( N_t(s_i) = N(s_i) \), for all \( t = 1, ..., T \). We define \( N(s_i) \) in the same manner as Cressie and Chan (1989) did:

\[
    j \in N(s_i) \quad \text{if} \quad d_{ij} \leq 30 \text{ miles}, \tag{2.3.9}
\]

where \( d_{ij} \) is the great-circle distance between the \( i \)-th and \( j \)-th county seats; see also Figure 2.4, where the spatial correlation becomes approximately zero at distances beyond 30 miles. Define \( C(k_t) = C(k) = (\min\{d_{ij} : s_j \in N(s_i), \ i = 1, ..., N\})^k \), which allows us to compare the interpretation of \( \phi_t \) for different values of \( k \) at each time period. We can specify \( k_t \) according to how fast \( c_{tij} \) will decrease as a function of the great-circle distance \( d_{ij} \) between locations \( s_i \) and \( s_j \). A choice of \( k_t \equiv 0 \) amounts to equal weights for all neighbors, which was preferred by Cressie and Chan (1989). In this case, \( C(k_t) = 1; t = 1, ..., T \). According to Cressie (1993, p.437), the partial correlations of the CAR model defined above can be written as

\[
    \text{corr}^2 (\zeta_t(s_i), \zeta_t(s_j))|\{\zeta_t(s_k) : k = 1, ..., N \text{ and } k \neq i, j\}) = \phi_t^2 c_{tij} c_{tji} = \phi_t^2 \left( \frac{C(k_t)}{d_{ij}^{k_t}} \right)^2, \tag{2.3.10}
\]

for \( i, j = 1, ..., N, \) and \( i \neq j, \ t = 1, ..., T \). Thus,

\[
    0 \leq \phi_t^2 \left( \frac{C(k_t)}{d_{ij}^{k_t}} \right)^2 \leq 1. \tag{2.3.11}
\]

Now, when \( k_t = 0 \), (2.3.11) results in \( \phi_t^2 \in [0, 1] \), and therefore \( \phi_t \) can be interpreted as a unitless correlation taking values from \([-1, 1]\). Furthermore, according to Besag (1974), only when \( M_t(S)^{-1} [I - \phi_t C_t(S)] \) is symmetric and positive-definite, the CAR model discussed above can imply a valid joint distribution defined in (2.3.5). From (2.3.7) and (2.3.8), we can see that \( C_t(S)M_t(S) \) is symmetric, because we have

\[
    \frac{c_{tij}}{E_t(s_j)} = \frac{C(k_t) d_{ij}^{-k_t}}{[E_t(s_i) E_t(s_j)]^{1/2}} I(s_j \in N(s_i)) = \frac{c_{tji}}{E_t(s_i)}, \quad i, j = 1, ..., N, \tag{2.3.12}
\]

where \( I(\cdot) \) is an indicator function. Hence, \( M_t(S)^{-1/2} C_t(S) M_t(S)^{1/2} \) is symmetric, which implies that \( M_t(S)^{-1} [I - \phi_t C_t(S)] \) is symmetric. Cressie (1993, p.559) pointed
out that the matrix $M_t(S)^{-1}[I - \phi_t C_t(S)]$ is positive-definite if and only if the matrix $(I - \phi_t M_t(S)^{-1/2}C_t(S)M_t(S)^{1/2})$ is positive-definite. Consequently, the lower ($\phi_{\text{min},t}$) and upper ($\phi_{\text{max},t}$) bounds on $\phi_t$ are $1/\psi_{\text{min},t}$ and $1/\psi_{\text{max},t}$, where $\psi_{\text{min},t} < 0 < \psi_{\text{max},t}$ are the smallest and largest eigenvalues of the matrix $M_t(S)^{-1/2}C_t(S)M_t(S)^{1/2}$. Notice from (2.3.6) that the structure of $U_t(S)$ is motivated by Cressie and Chan (1989) but, in our work, we model the count data directly, and we have the opportunity to incorporate the temporal aspect into the CAR process. To capture that spatio-temporal variability, we use (2.3.4) and define the $N \times N$ matrix $H_t(S) \equiv (h_{tij})$ as follows. For $t = 2, ..., T$, $i, j = 1, ..., N$:

$$h_{tij} \equiv \begin{cases} \varphi_{t0}, & \text{if } j = i; \\ \varphi_{t1}, & \text{if } s_j \in N(s_i); \\ \varphi_{t2}, & \text{if } s_j \in N^{(2)}(s_i); \\ 0, & \text{otherwise}, \end{cases}$$

where $\varphi_{t0}$, $\varphi_{t1}$, and $\varphi_{t2}$ all belong to (-1,1) and represent the global measures of (i) the impact of the site itself, (ii) the impact of neighbors (or first-order neighbors), and (iii) the impact of second-order neighbors, respectively, on the process at the previous time point. Recall that $N(s_i)$ is the neighborhood set (or first-order neighborhood) of $s_i$ defined in (2.3.9). We define $N^{(2)}(s_i)$ as the second-order neighborhood of $s_i$, namely, if site $s_j \notin N(s_i)$, but there exists a site $s_k$ such that $s_k \in N(s_i)$ and $s_j \in N(s_k)$, then $s_j \in N^{(2)}(s_i)$ (e.g., Mugglin et al., 2002).

The final component of (2.3.2), $\xi_t(S) \equiv (\xi_t(s_1), ..., \xi_t(s_N))'$, is an $N$-dimensional vector that captures the extra-Poisson variation, which we assume is distributed according to

$$\xi_t(S) \sim \text{MVN}(0, \sigma^2 \mathbf{D}); \quad t = 1, ..., T.$$ 

(2.3.14)
In (2.3.14), $\sigma^2_\xi$ is a nonnegative variance-component parameter that does not depend on time or space, and $D$ is a given diagonal matrix. In our case, we capture the heterogeneity of counties by assuming $D \equiv \text{diag}(E_t(s_1)^{-1}, \ldots, E_t(s_N)^{-1})$.

Compared to other spatio-temporal CAR models for counts (e.g., Cressie and Mugglin, 2000, Mugglin et al., 2002, Kottas et al., 2008, and Martínez-Beneito et al., 2008, etc.), the spatio-temporal CAR model discussed above has several advantages. Rather than simply using a single parameter to specify the temporal dependence between current and previous processes over all locations (e.g., Kottas et al., 2008 and Martínez-Beneito et al., 2008), we incorporate the spatial information in the autoregressive matrix $H_t(S)$ through (2.3.13), which allows the temporal dependence to vary according to the distance between the locations of current and previous processes. Although Mugglin et al. (2002) and Cressie and Mugglin (2000) use a similar autoregressive structure, they do not incorporate the temporal aspect in the covariance of the CAR model as our approach does in (2.3.4)-(2.3.6). Furthermore, unlike some recent spatio-temporal CAR models for counts (e.g., Mugglin et al., 2002, Kottas et al., 2008, etc.), our approach includes the extra-Poisson variation, $\{\xi_t(s_i)\}$, in the model (as does that of Besag et al., 1991).

### 2.3.3 Parameter Model

Now we specify the joint prior distribution for the parameters to complete the Bayesian hierarchical model. Recall that the parameters include the fixed-effect coefficient $\beta = (\beta_0, \beta_1, \beta_2)'$, the variance and correlation parameters $\{\tau^2_t\}$ and $\{\phi_t\}$, the variance parameter of the extra-Poisson variation $\sigma^2_\xi$, and the autoregressive parameters $\{\varphi_{tl}\}$. Notice that a transformation

$$\theta_{tl} = \log \left[ \frac{(1 + \varphi_{tl})}{(1 - \varphi_{tl})} \right]; \quad l = 0, \ldots, 2, \quad t = 2, \ldots, T, \quad (2.3.15)$$
changes the support of $\varphi_t$ to $(-\infty, \infty)$ and allows a Gaussian prior to be used. Therefore, instead of assigning prior distributions directly to $\{\varphi_t\}$, we put prior distributions on $\{\theta_t\}$. Hence, assuming independence and using $[Y]$ as generic notation for the density of $Y$, we have

$$
[\beta, \tau_1^2, \ldots, \tau_T^2, \phi_1, \ldots, \phi_T, \theta_{20}, \ldots, \theta_{T2}, \sigma_\xi^2] = [\beta][\tau_1^2] \cdots [\tau_T^2][\phi_1] \cdots [\phi_T][\theta_{20}] \cdots [\theta_{T2}][\sigma_\xi^2],
$$

(2.3.16)

where the prior distributions of individual parameters are specified as follows:

$$
\beta_p \sim \text{Gaussian}(\mu_{\beta_p}, \sigma_{\beta_p}^2); \quad p = 0, 1, 2,
$$

$$
\tau_t^2 \sim \text{Inverse gamma}(a_t, b_t); \quad t = 1, \ldots, T,
$$

$$
\phi_t \sim \text{Uniform}(\phi_{\min,t}, \phi_{\max,t}); \quad t = 1, \ldots, T,
$$

$$
\theta_{tl} \sim \text{Gaussian}(\mu_{\theta_{tl}}, \sigma_{\theta_{tl}}^2); \quad t = 2, \ldots, T, \quad l = 0, 1, 2.
$$

Recall that the hyperparameters $\phi_{\min,t}$ and $\phi_{\max,t}$ should be determined such that the matrix $(I - \phi_tM_t(S)^{-1/2}C_t(S)M_t(S)^{1/2})$ is positive-definite, as discussed in Section 2.3.2. The other hyperparameters, namely, $\{\mu_{\beta_p}\}$, $\{\sigma_{\beta_p}^2\}$, $\{a_t\}$, $\{b_t\}$, $\{\mu_{\theta_{tl}}\}$, and $\{\sigma_{\theta_{tl}}^2\}$, are specified in Section 2.4.1.

### 2.3.4 Prior Distribution on Extra-Poisson-Variation Component

Putting a prior distribution on the extra-Poisson component of variation, $\sigma_\xi^2$, is problematic because there is confounding of $\sigma_\xi^2$ and $\{\tau_t^2\}$; in (2.3.2), $\text{var}(\log(\lambda_t(S)))$ has diagonal elements close to $(\tau_t^2/E_t(s_i)) + (\sigma_\xi^2/E_t(s_i))$, for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, and hence in the MCMC, sampling from the full conditionals does not mix. In a Gaussian hierarchical model for regional data, where the setting was purely spatial,
Kang et al. (2009) resolved a similar problem of confounding between the small-scale variation and the measurement-error variance; they used properties of spatial dependence in small regions to tie down the value of the measurement-error variance.

However, in our case, we do not have additive Gaussian models for the data and the process. Hence, we take an indirect route through calibrating the Poisson spatio-temporal hierarchical model to an approximately additive Gaussian model based on the Freeman-Tukey transformation of the counts. Our approach is to derive a model based on FTSMR-values to approximate the Poisson hierarchical model. Then we apply the methodology of Kang et al. (2009) on the approximate Gaussian model and, through the approximation, we are able to tie down the value of $\sigma^2_\xi$ in the Poisson hierarchical model. Finally, we put a Dirac delta prior on the $\sigma^2_\xi$ we obtain.

We now give the details of this strategy: From the Poisson hierarchical model defined in (2.3.1) and (2.3.2) in Section 2.3, an approximate Gaussian model based on FTSMR is derived, as shown in (A.2.3) in Appendix A.2. Then, from (A.1.6) in Appendix A.1, we see that the weighted FTSMRs, $\{\sqrt{E_t(s_i)}\text{FTSMR}_t(s_i)\}$, have approximately equal variability for all locations at time $t$, $t = 1, ..., T$. Therefore, in order to take into account the unequal variation resulting from unequal numbers of expected SIDS ($E_t(s_i)$), we weight (A.2.3) in Appendix A.2 by $\{\sqrt{E_t(s_i)}\}$ to obtain

$$\text{FTSMR}^w_t(s_i) = \mu^w_t(s_i) + \eta^w_t(s_i) + \varepsilon^w_t(s_i); \quad (2.3.17)$$

from (A.2.3) - (A.2.7) in Appendix A.2, we have

$$\text{FTSMR}^w_t(s_i) \equiv \sqrt{E_t(s_i)}\text{FTSMR}_t(s_i), \quad (2.3.18)$$

$$\mu^w_t(s_i) \equiv \sqrt{E_t(s_i)}\mu^FT_t(s_i) = \sqrt{E_t(s_i)}(X_t(s_i))'\beta^FT = \sqrt{E_t(s_i)}\left( J + \frac{1}{2}(X_t(s_i))'\beta \right), \quad (2.3.19)$$

$$\eta^w_t(s_i) \equiv \sqrt{E_t(s_i)}\eta^FT_t(s_i) = \frac{J}{2} \frac{\sqrt{E_t(s_i)}}{\eta_t(s_i)} \eta_t(s_i), \quad (2.3.20)$$
Figure 2.6: A straight line fitted by regressing the robust empirical semivariograms on the spatial lags near the origin. The intercept of this line with the vertical axis is an estimate of \((\sigma^w)^2\). Units on the horizontal axis are in miles.

\[
\varepsilon_w^u(s_i) \equiv \sqrt{E_t(s_i)\varepsilon^{FT}_t(s_i)} = \sqrt{E_t(s_i)} \left( \frac{J}{2} \xi_t(s_i) + \nu_t(s_i) \right),
\]

(2.3.21)

for \(J = 2\sqrt{1000}\) and \(i = 1, \ldots, N, t = 1, \ldots, T\); recall that \(\mu_t(s_i), \eta_t(s_i), \) and \(\xi_t(s_i)\) are the trend, small-scale variability, and extra-Poisson variability, respectively, of the Poisson hierarchical model as defined in Section 2.3.2; and \(\{\nu_t(s_i)\}\) represents the
error after accounting for the epidemiological variation in the (conditional) mean, as
given in Appendix A.2.

Recall that the small-scale variation terms \( \{ \eta_t(s_i) \} \) defined in (2.3.4) in Section
2.3.2 has a spatio-temporal CAR structure. Therefore, from (2.3.19)-(2.3.21), we see
that model (2.3.17) is a Gaussian hierarchical spatio-temporal CAR model, and the
term \( \varepsilon^w_t(s_i) \) can be treated as a measurement error for the approximate Gaussian
model. Here, we use \((\sigma^w_\varepsilon)^2\) to represent the variance of \( \varepsilon^w_t(s_i) \), for \( i = 1, \ldots, N \) and \( t = 1, \ldots, T \). Following the approach of Kang et al. (2009), we obtain the robust
variogram estimator \( 2\hat{\gamma}(h) \) (Cressie and Hawkins, 1980) through (2.2.9) in Section
2.2.4. This estimator is based on the weighted residuals \( \{ R^w_t(s_i) \} \) defined by (2.2.8),
which are obtained from a weighted linear regression of FTSMR on FTNR and lag-1
MWT, weighted by \( \sqrt{E_t(s_i)} \). Here, we use the same covariates as in the Poisson
model because of the relation shown in (2.3.19). From the robust semivariogram
estimator at spatial lags close to zero, we fitted a linear semivariogram using weighted
least squares (Cressie, 1985); see Figure (2.6). Using the same reasoning as Kang et al.
(2009), \((\sigma^w_\varepsilon)^2\) can be estimated in an unbiased manner from the intercept of the fitted
line, namely,

\[
(\hat{\sigma}^w_\varepsilon)^2 = \hat{\gamma}(0+). \tag{2.3.22}
\]

Finally, in Appendix A.3-A.5, by using model (2.3.17) to approximate the Poisson
hierarchical model defined in (2.3.1) and (2.3.2), we show that the extra-Poisson
variation \( \sigma^2_\xi \) can be written approximately as a function of \((\sigma^w_\varepsilon)^2\). In Appendix A.5,
we derive equation (A.5.8), which is repeated here for completeness:

\[
\sigma^2_\xi \approx Q_0 + Q_1 \cdot (\sigma^w_\varepsilon)^2 + Q_2 \cdot (\sigma^w_\varepsilon)^4,
\]

where

\[
Q_0 = - \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{(1 + \hat{\mu}_t(s_i))}{E_t(s_i)},
\]
\[ Q_1 = \frac{4}{J^4} \sum_{i=1}^{N} \sum_{t=1}^{T} \left( \frac{\hat{\mu}_t^w(s_i)^2}{E_t(s_i)^2} \right), \]

\[ Q_2 = \frac{2}{J^4} \sum_{i=1}^{N} \sum_{t=1}^{T} \left( \frac{1}{E_t(s_i)^2} \right), \]

and recall that \( J = 2\sqrt{1000}. \) Therefore, by substituting \((\hat{\sigma}_w^2)^2\) obtained by (2.3.22) into the right-hand side of (A.5.8), we obtain \( \hat{\sigma}_w^2. \) In our case, \((\hat{\sigma}_w^2)^2 = 701.67, \) and hence we obtain \( \hat{\sigma}_w^2 = 0.0246. \) Thus, our prior on \( \sigma_w^2 \) is a Dirac delta function at \( \hat{\sigma}_w^2. \)

### 2.4 Results

#### 2.4.1 Fitting the Hierarchical Spatio-Temporal CAR Model

From Section 2.3, we obtain the joint posterior distribution of all parameters, which is proportional to a product of the data model, the process model, and the parameter model, as follows:

\[
\begin{align*}
&\left[ \beta, \{ \tau_t^2 \}, \{ \phi_t \}, \{ \theta_u \}, \{ \eta_t(s_i) \}, \{ \xi_t(s_i) \}, \sigma_\xi^2 | \{ Z_t(s_i) \} \right] \\
&\propto \left( \prod_{t=1}^{T} \prod_{i=1}^{N} [Z_t(s_i)|\beta, \eta_t(s_i), \xi_t(s_i)] \right) [\eta_1(S)|\tau_1^2, \phi_1] \\
&\quad \cdot [\eta_2(S)|H_2(S)\eta_1(S), \tau_2^2, \phi_2] \cdots [\eta_T(S)|H_T(S)\eta_{T-1}(S), \tau_T^2, \phi_T][\xi_t(S)|\sigma_\xi^2] \\
&\quad \cdot [\beta][\tau_1^2] \cdots [\tau_T^2][\phi_1] \cdots [\phi_T][\theta_20] \cdots [\theta_{T2}][\sigma_\xi^2],
\end{align*}
\]

(2.4.1)

where we specify the starting value of \( \{ \eta_t(S) \} \) as \( \eta_1(S) \sim \text{MVN}(0, \kappa^2U_1), \) and \( \kappa^2 > 1 \) is used to capture the additional uncertainty on \( \eta_1(S). \) In our application, we chose \( \kappa^2 = 1.5. \) Recall that we assume the parameter model consists of independent prior distributions on each parameter, as in (2.3.16) in Section 2.3.3; and parameter \( \{ \theta_u \} \) is a transformation of \( \{ \phi_u \}, \) which is the parameter of the autoregressive matrix \( H_t(S); \) see (2.3.15) in Section 2.3.3. Furthermore, in Section 2.3.4, \( \sigma_\xi^2 \) was estimated with \( \hat{\sigma}_\xi^2. \)
and “plugged into” the Bayesian analysis; this is equivalent to putting a degenerate prior on $\hat{\sigma}_2^2$. The prior distributions of the other parameters are discussed in Section 2.3.3 and further specified in Table 2.1, based on some exploratory data analysis and our preference for vague priors.

Notice that it is difficult to obtain a closed-form expression for the joint posterior distribution in (2.4.1) as well as the marginal posterior distributions of individual parameters; instead, we obtain the posterior through a Markov chain Monte Carlo (MCMC) simulation with a Gibbs sampler incorporating Metropolis-Hastings steps where necessary (e.g., Waller et al., 1997), based on the full conditional distributions listed as follows (we use “rest” to represent all other parameters as well as the data \{Z_t(s_i)\}):

- \{\beta_0, \beta_1, \beta_2\}

\[
[\beta_0, \beta_1, \beta_2|\text{rest}] \propto \prod_{t=1}^{T} \prod_{i=1}^{N} Z_t(s_i)|\beta_0, \beta_1, \beta_2, \eta_t(s_i), \xi_t(s_i)| \cdot [\beta_0] \cdot [\beta_1] \cdot [\beta_2].
\]

- \{\tau^2_t\}: For $t = 1, ..., T$,

\[
[\tau^2_t|\text{rest}] \propto \begin{cases} 
[\eta_t(S)|\tau^2_t, \phi_t] \cdot [\tau^2_t], & \text{if } t = 1; \\
[\eta_t(S)|H_t(S)\eta_{t-1}(S), \tau^2_t, \phi_t] \cdot [\tau^2_t], & \text{if } t = 2, ..., T.
\end{cases}
\]

- \{\phi_t\}: For $t = 1, ..., T$,

\[
[\phi_t|\text{rest}] \propto \begin{cases} 
[\eta_t(S)|\tau^2_t, \phi_t] \cdot [\phi_t], & \text{if } t = 1; \\
[\eta_t(S)|H_t(S)\eta_{t-1}(S), \tau^2_t, \phi_t] \cdot [\phi_t], & \text{if } t = 2, ..., T.
\end{cases}
\]

- \{\theta_{l,t}\}: For $l = 0, ..., 2$, $t = 2, ..., T$,

\[
[\theta_{l,t}|\text{rest}] \propto [\eta_t(S)|H_t(S)\eta_{t-1}(S), \tau^2_t, \phi_t] \cdot [\theta_{l,t}];
\]

recall that \{\theta_{l,t}\} are transformations of \{\varphi_{l,t}\}, which are the parameters of $H_t(S)$; see (2.3.13) in Section 2.3.2 and (2.3.15) in Section 2.3.3.
• \{\eta_t(s_i)\}: For \(i = 1, ..., N, t = 1, ..., T,\)

\[
[\eta_t(s_i)|\text{rest}] \propto \begin{cases} 
[Z_t(s_i)|\beta_0, \beta_1, \beta_2, \eta_t(s_i), \xi_t(s_i)] \cdot [\eta_t(S)|\tau^2_t, \phi_t] \\
\cdot [\eta_{t+1}(S)|H_{t+1}(S)\eta_t(S), \tau^2_t, \phi_t], & \text{if } t = 1; \\
[Z_t(s_i)|\beta_0, \beta_1, \beta_2, \eta_t(s_i), \xi_t(s_i)] \cdot [\eta_t(S)|H_t(S)\eta_{t-1}(S), \tau^2_t, \phi_t] \\
\cdot [\eta_{t+1}(S)|H_{t+1}(S)\eta_t(S), \tau^2_t, \phi_t], & \text{if } t = 2, ..., T - 1; \\
[Z_t(s_i)|\beta_0, \beta_1, \beta_2, \eta_t(s_i), \xi_t(s_i)] \\
\cdot [\eta_t(S)|H_t(S)\eta_{t-1}(S), \tau^2_t, \phi_t], & \text{if } t = T.
\end{cases}
\]

• \{\xi_t(s_i)\}: For \(i = 1, ..., N, t = 1, ..., T,\)

\[
[\xi_t(s_i)|\text{rest}] \propto [Z_t(s_i)|\beta_0, ..., \beta_2, \eta_t(s_i), \xi_t(s_i)][\xi_t(s_i)].
\]

In the simulation procedure, we first specify the initial values for all parameters. Then we draw an observation cyclically from each full conditional distribution conditioning on the most recent values of the other parameters. This procedure allows us to obtain a Markov chain whose stationary distribution is the joint posterior distribution of all parameters given the data. Hence, after “burn-in,” we obtain a random sample from the posterior distribution. Notice that, in our case, except for \([\tau^2_t|\text{rest}],\) which follows the Inverse Gamma distribution for \(t = 1, ..., T\) and can be sampled through Gibbs updating (e.g., Casella and George, 1992 and Cressie and Mugglin, 2000), the full conditional distributions of all the other parameters discussed above are not recognizable forms; therefore, Metropolis-Hastings updates are applied (e.g., Hastings, 1970 and Cressie and Mugglin, 2000).

We ran five parallel independent MCMC chains, each of which contained 10,000 iterations. After a burn-in of 2,000 iterations for each chain, we obtained a total of 40,000 realizations from the posterior distribution. Table 2.1 summarizes the posterior
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior distribution</th>
<th>Prior quantiles</th>
<th>Posterior quantiles</th>
<th>Posterior</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>Gaussian(0, 4)</td>
<td>-3.29 0 3.29</td>
<td>-0.196 -0.0112 -0.0103 -0.1024 0.0022</td>
<td></td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>Gaussian(0, 4)</td>
<td>-3.29 0 3.29</td>
<td>-0.1600 -0.0425 0.0790 -0.0424 0.0037</td>
<td></td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>Gaussian(0, 4)</td>
<td>-3.29 0 3.29</td>
<td>-0.1591 -0.0535 0.0501 -0.0424 0.0029</td>
<td></td>
</tr>
<tr>
<td>$\tau_2^1$</td>
<td>Inverse Gamma(0.25, 0.4)</td>
<td>1.457 57.2 $9.48 \times 10^6$</td>
<td>0.2411 0.4118 0.7100 0.4279 0.01453</td>
<td></td>
</tr>
<tr>
<td>$\tau_2^2$</td>
<td>Inverse Gamma(0.25, 0.4)</td>
<td>1.457 57.2 $9.48 \times 10^6$</td>
<td>0.1813 0.2978 0.4960 0.3085 0.0067</td>
<td></td>
</tr>
<tr>
<td>$\tau_2^3$</td>
<td>Inverse Gamma(0.25, 0.4)</td>
<td>1.457 57.2 $9.48 \times 10^6$</td>
<td>0.1848 0.3018 0.5097 0.0070 0.3138</td>
<td></td>
</tr>
<tr>
<td>$\tau_2^4$</td>
<td>Inverse Gamma(0.25, 0.4)</td>
<td>1.457 57.2 $9.48 \times 10^6$</td>
<td>0.1812 0.2958 0.4976 0.0066 0.3070</td>
<td></td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>Uniform(-0.331, 0.199)</td>
<td>-0.3178 -0.066 0.1858 -0.1391 -0.00011 0.1185 -0.0026 0.0045</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>Uniform(-0.331, 0.199)</td>
<td>-0.3178 -0.066 0.1858 -0.1415 -0.00393 0.1165 -0.0061 0.0044</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\phi_3$</td>
<td>Uniform(-0.331, 0.199)</td>
<td>-0.3178 -0.066 0.1858 -0.1395 -0.00326 0.1161 -0.0059 0.0043</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\phi_4$</td>
<td>Uniform(-0.331, 0.199)</td>
<td>-0.3178 -0.066 0.1858 -0.1427 -0.00447 0.1167 -0.0072 0.0045</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\phi_5$</td>
<td>Uniform(-0.331, 0.199)</td>
<td>-0.3178 -0.066 0.1858 -0.1418 -0.00225 0.1170 -0.0054 0.0045</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\theta_{20}$</td>
<td>Gaussian(0, 4)</td>
<td>-3.29 0 3.29</td>
<td>-0.7715 0.0867 0.9829 0.0882 0.1950</td>
<td></td>
</tr>
<tr>
<td>$\theta_{21}$</td>
<td>Gaussian(0, 4)</td>
<td>-3.29 0 3.29</td>
<td>-0.3094 0.0208 0.3484 0.0210 0.0274</td>
<td></td>
</tr>
<tr>
<td>$\theta_{22}$</td>
<td>Gaussian(0, 4)</td>
<td>-3.29 0 3.29</td>
<td>-0.2795 -0.01701 0.2458 -0.0174 0.0175</td>
<td></td>
</tr>
<tr>
<td>$\theta_{30}$</td>
<td>Gaussian(0, 4)</td>
<td>-3.29 0 3.29</td>
<td>-0.7024 0.0878 0.9206 0.0921 0.1739</td>
<td></td>
</tr>
<tr>
<td>$\theta_{31}$</td>
<td>Gaussian(0, 4)</td>
<td>-3.29 0 3.29</td>
<td>-0.3237 -0.004703 0.3143 -0.0063 0.0295</td>
<td></td>
</tr>
<tr>
<td>$\theta_{32}$</td>
<td>Gaussian(0, 4)</td>
<td>-3.29 0 3.29</td>
<td>-0.2388 0.0123 0.2611 0.01138 0.0157</td>
<td></td>
</tr>
<tr>
<td>$\theta_{40}$</td>
<td>Gaussian(0, 4)</td>
<td>-3.29 0 3.29</td>
<td>-0.6755 0.0738 0.9058 0.0844 0.1631</td>
<td></td>
</tr>
<tr>
<td>$\theta_{41}$</td>
<td>Gaussian(0, 4)</td>
<td>-3.29 0 3.29</td>
<td>-0.3112 0.0114 0.3406 0.0129 0.0262</td>
<td></td>
</tr>
<tr>
<td>$\theta_{42}$</td>
<td>Gaussian(0, 4)</td>
<td>-3.29 0 3.29</td>
<td>-0.2353 0.00447 0.2516 0.0059 0.0153</td>
<td></td>
</tr>
<tr>
<td>$\theta_{50}$</td>
<td>Gaussian(0, 4)</td>
<td>-3.29 0 3.29</td>
<td>-0.7961 0.04878 0.9070 0.0508 0.1850</td>
<td></td>
</tr>
<tr>
<td>$\theta_{51}$</td>
<td>Gaussian(0, 4)</td>
<td>-3.29 0 3.29</td>
<td>-0.3143 0.006294 0.3332 0.0072 0.0272</td>
<td></td>
</tr>
<tr>
<td>$\theta_{52}$</td>
<td>Gaussian(0, 4)</td>
<td>-3.29 0 3.29</td>
<td>-0.2624 -0.0051 0.2537 -0.0040 0.0165</td>
<td></td>
</tr>
</tbody>
</table>

The posteriors of all parameters are clearly much tighter than the priors, which indicates that there is substantial learning about these parameters, in spite of having some zero counts in each year. We notice that all the regression parameters, except the intercept $\beta_0$, have 95% credible intervals that contain zero. However, both Table 2.1 and the histograms in Figure 2.7 show a smaller variation for the posterior distribution of $\beta_2$ (the coefficient of the lag-1 minimum winter temperature) than those of the other regression parameters.
To provide information on how far the posterior median is from zero, we compute

\[ A_j \equiv \frac{\text{posterior median of } \beta_j}{\text{posterior IQR of } \beta_j}, \quad j = 0, 1, 2, \]

where IQR represents the interquartile range. In our case, \( A_0 = -1.58 \), \( A_1 = -0.52 \) and \( A_2 = -0.73 \). This suggests a negative value for \( \beta_2 \); moreover, the histogram of \( \beta_2 \) in Figure 2.7 and the fact that 84\% of the posterior samples of \( \beta_2 \) are less than zero, support our conclusion that \( \beta_2 < 0 \). Recall the hypothesis that unusually cold
weather in winter may affect the health of pregnant women and thus may increase the risk of SIDS for the babies born in the following year.

Table 2.1 suggests that all of the transformed autoregression parameters $\theta_{tl}, t = 2, ..., 5, l = 0, 1, 2$, have 95% credible intervals that include zero. It would appear that the lag temperature in the trend has explained most of the temporal variability, leaving only a small amount of temporal dependence in the spatial process $\{\eta_t(S)\}$. This result also agrees with the exploratory data analysis in Section 2.2.4; see Figure 2.5.

Figure 2.8: The goodness-of-fit measurements $\{w_t(s_i)\}$ versus $\{1/\sqrt{E_t(s_i)}\}$, for $i = 1, ..., N$, and $t = 2$, namely 1980-1981. A linear relationship is expected.
Cressie and Mugglin (2000) proposed a goodness-of-fit measure to assess the performance of a hierarchical spatio-temporal model. Consider the measure

\[ W(s_i) \equiv \frac{\sum_t w_t(s_i)}{\sum_t 1}, \]

where \( w_t(s_i) \) is the width of the 95\% Bayesian credible interval from the posterior distribution of \( \eta_t(s_i); i = 1, ..., N. \) They plotted \( \{W(s_i)\} \) against the inverse of the square root of the number of expected cases, \( \{1/\sqrt{E(s_i)}\} \), to see whether there was any departure from linearity. They pointed out that if a model fits well, an inverse proportionality between \( W(s_i) \) and \( \sqrt{E(s_i)} \) can be expected because the variances of posterior distributions of parameters tend to be small in those places with many expected SIDS counts. In our case, we introduce time by plotting \( \{w_t(s_i)\} \), the length of the 95\% credible interval at time \( t \) versus \( \{1/\sqrt{E_t(s_i)}\}; t = 1, ..., T. \) All plots show a strong linear relationship between \( w_t(s_i) \) and \( 1/\sqrt{E_t(s_i)} \), indicating a good overall fit of our model. In Figure 2.8, we show the plot for \( t = 2 \) because there is an apparent outlier in year 1980-1981, where Alleghany county (county 3) looks unusual. In fact, it had only two SIDS cases in that year, but that was an unusually high number for a county that had only 117 live births that year. (The overall SIDS rate during the study period was about 0.002, and 2/117 is approximately ten times higher.)

### 2.4.2 Posterior Analysis

Based on the posterior samples of the parameters listed in Table 2.1, the posterior distributions of various quantities of interest can be obtained. In order to see clearly the temporal variability of relative risks at each time point, for each year, we take a weighted average of them over all counties; that is,

\[
\bar{\lambda}_t^{(j)} \equiv \frac{\sum_{i=1}^N \sqrt{E_t(s_i)} \lambda_t^{(j)}(s_i)}{\sum_{i=1}^N \sqrt{E_t(s_i)}}; \quad j = 1, ..., m, \quad t = 1, ..., T, \quad (2.4.2)
\]
where $j$ represents the corresponding sample from the $j$-th iteration and $m$ is the number of total posterior samples. As mentioned in Section 2.4.1, $m = 40,000$.

Figure 2.9: Posterior mean and 95% HPD credible intervals of yearly weighted average of posterior samples of $\{\bar{\lambda}^{(j)}_t\}$ defined in (2.4.2), plotted against $t = 1, \ldots, 5$.

Figure 2.9 shows a plot of the posterior means and 95% credible intervals of the weighted average relative risks, plotted against time. Notice that the means of overall relative risks are around 1 for all five years, indicating (as expected, due to internal standardization given by (2.2.1)) that the observed number of SIDS is almost equal to the expected number of SIDS for each year. However, Figure 2.9 also suggests an increase of the average relative risk in the third year (1981-1982), as well as a decrease in the fifth year (1983-1984). Recall that there was an unusually cold winter in the
second year and a warm winter in the fourth year. Thus, the temporal variability shown in Figure 2.9 further supports a possible lag effect due to winter temperatures.

Figure 2.10: Boxplots of yearly weighted average of posterior samples of \( \{ \bar{\mu}_t^{(j)} \} \) defined in (2.4.3), plotted against \( t = 1, \ldots, 5 \).

From (2.3.2), the hidden spatio-temporal process of log relative risks, can be further decomposed into three components, namely, the large-scale variation (spatial trend) \( \mu_t(S) \), the small-scale variation \( \eta_t(S) \), and the extra-Poisson variation \( \xi_t(S) \). To be clearer about the temporal change of the relative risks, we can average each component of (2.3.2) over the 100 counties and look at their respective posterior distributions; that is, 

\[
\bar{\mu}_t^{(j)} = \frac{\sum_{i=1}^{N} \sqrt{E_t(s_i)} \mu_t^{(j)}(s_i)}{\sum_{i=1}^{N} \sqrt{E_t(s_i)}}; \quad j = 1, \ldots, m, \ t = 1, \ldots, T \tag{2.4.3}
\]
Figure 2.11: Boxplots of yearly weighted average of posterior samples of \( \bar{\eta}_t^{(j)} \) defined in (2.4.4), plotted against \( t = 1, \ldots, 5 \).

\[
\bar{\eta}_t^{(j)} \equiv \frac{\sum_{i=1}^{N} \sqrt{E_t(s_i)}\eta_t^{(j)}(s_i)}{\sum_{i=1}^{N} \sqrt{E_t(s_i)}}; \quad j = 1, \ldots, m, \quad t = 1, \ldots, T \tag{2.4.4}
\]

\[
\bar{\xi}_t^{(j)} \equiv \frac{\sum_{i=1}^{N} \sqrt{E_t(s_i)}\xi_t^{(j)}(s_i)}{\sum_{i=1}^{N} \sqrt{E_t(s_i)}}; \quad j = 1, \ldots, m, \quad t = 1, \ldots, T; \tag{2.4.5}
\]

where we recall that \( j \) represents the corresponding sample from the \( j \)-th iteration and \( m = 40,000 \) is the number of total posterior samples. Boxplots are obtained from these; for example, Figures 2.10 and 2.11 show the posterior distribution of \( \{\bar{\eta}_t^{(j)}\} \) and \( \{\bar{\eta}_t^{(j)}\} \), respectively. From Figure 2.10, we can see clearly that the trend in the large-scale variation increases in the third year and decreases in the fifth year. For the small-scale variation, there is no obvious temporal pattern shown in Figure 2.11, and the same is true (not shown) for the posterior distributions of the extra-Poisson
variation, \( \{ \tilde{\xi}^{(j)} \} \). Thus, these figures further support the existence of a lag-winter-temperature effect.

Now, let us turn our attention to the spatio-temporal variability. Figure 2.12a shows yearly choropleth maps of raw SMRs and Figure 2.12b shows the yearly choropleth maps of the posterior medians of relative risk. As expected, the latter are smoother and hence they capture the general spatial pattern and its changes over the five years during the study period. For example, in the first year, the counties with high values seem to cluster in a small area of the north east and in a horizontal band across the middle part of the state. However, in the following years, areas with high values appear more often in the west and south. Furthermore, the posterior median relative risks seem to increase in most counties in the third year and decrease in the fifth year, which agrees with the earlier discussion of the average relative risks shown in Figure 2.9.

We further obtain choropleth maps of the posterior median for each of the three components for each of the five years; for example, Figures 2.13 and 2.14 show the posterior median for \( \mu_t(S) \) and \( \eta_t(S) \), respectively. As expected, the large-scale variation (deterministic trend) term is much smoother than the small-scale-variation term and the extra-Poisson-variation term. Figure 2.13 clearly indicates the increase in the third year, and the decrease in the fifth year. The yearly maps in Figure 2.13 also indicate higher-risk areas clustered on the western border and a band from north-west to south-east for most of these five years. For the small-scale variation, Figure 2.14 indicates that there seem to be higher-risk counties in a middle horizontal band across the state, however, we are unable to see any clear temporal patterns. For the extra-Poisson variation, the choropleth maps (not shown) do not suggest any obvious spatial or temporal pattern, as expected. These figures provide further evidence, now in the posterior analysis, of a lag-winter-temperature effect; the clustering in the
small-scale variation shows a spatial-dependence effect in neighboring counties; and there is no evidence of any remaining dynamical temporal structure.

Figure 2.12: (a) Choropleth maps of raw SMRs for five years; (b) Choropleth maps of posterior median of relative risk \( \lambda_t(s_i) \) for five years.
Figure 2.13: Choropleth maps of posterior median of $\{\mu(t)\}$ for five years. The ranges in the legend correspond to seven percentile ranges: “< 5%,” “5%-20%,” “20%-35%,” “35%-65%,” “65%-80%,” “80%-95%,” and “> 95%,” respectively.
Figure 2.14: Choropleth maps of posterior median of \( \{ \eta_t(s_i) \} \) for five years. The ranges in the legend correspond to seven percentile ranges: “< 5%,” “5%-20%,” “20%-35%,” “35%-65%,” “65%-80%,” “80%-95%,” and “> 95%,” respectively.

2.5 Discussion and Conclusions

In this chapter, we study sudden infant death syndrome (SIDS) in North Carolina, from not only a spatial perspective (by county), but also include a temporal perspective (by year). The latter allows epidemiological questions to be addressed that have heretofore been unavailable. That is, in the data we analyzed, we saw a
negative relationship between lag winter temperature and the relative risks of SIDS; one might hypothesize that unusually cold weather may affect the health of pregnant women and, hence, may increase the SIDS risk of babies born the following year.

In order to capture the spatio-temporal process hidden behind the noisy data, we propose a hierarchical spatio-temporal CAR model for the yearly SIDS dataset. When using a traditional MCMC approach to fit the hierarchical model, we encountered a difficult problem of confounding between sources of variability. To deal with this, we adapted the approach of Kang et al. (2009) to our setting where we have spatio-temporal count data and a Poisson data model. In effect, we use variance-stabilizing approximations and calibration ideas from geostatistics to infer the extra-Poisson variance term, instead of putting a vague prior on it.

From the MCMC results in Section 2.4, we conclude that our Bayesian analysis can capture the spatio-temporal process hidden behind the noisy data. All our results from the posterior analysis suggest a negative relationship between lag winter temperature and the relative risks of SIDS. Furthermore, we detect quite strong residual spatial dependence between nearby neighbors but no residual temporal dependence. Our spatio-temporal statistical model indicates that Bayesian procedures could be used to forecast behaviors of SIDS in North Carolina.
3.1 Introduction

A pandemic (e.g., caused by viruses such as H1N1, H5N1) is an epidemic of an infectious disease spreading through human populations across a large region (e.g., Potter, 2001). Recently, the risk of pandemic influenza has been a significant public health concern, and much attention has been paid to achieve more precise and timely estimates and predictions of influenza activity.

Compartment epidemic models, such as the SIR (susceptible-infectious-recovered) model, have been widely used to study the dynamical process of an infectious disease in a large population. The SIR model was first proposed by Kermack and McKendrick (1927) to explain plague and cholera epidemics. Assume that at any given time $t$, a fixed population can be split into three compartments (susceptible, infectious, and recovered). Then the dynamical process is captured through the following set of nonlinear ordinary differential equations (ODEs),

\[
\frac{dS}{dt} = -\beta SI + \phi R, \quad (3.1.1)
\]

\[
\frac{dI}{dt} = \beta SI - \gamma I, \quad (3.1.2)
\]

\[
\frac{dR}{dt} = \gamma I - \phi R, \quad (3.1.3)
\]
where $\beta$ is the transmission rate (also referred as the contact or infection rate) per unit time, which can be expressed as that fraction of contacts between susceptible individuals and infectious individuals that result in an infection; $\gamma$ is the rate of recovery per unit time, which is the rate at which infectious individuals stop being infectious due to recovery (or death); $\phi$ is the rate of loss of immunity of recovered individuals per unit time, which is the average rate at which recovered individuals become susceptible again. In (3.1.1)-(3.1.3), $S(t)$, $I(t)$, and $R(t)$ are functions of time $t$ and represent the numbers in the population that are susceptible, infectious, and recovered, respectively. Notice that no births are assumed, nor are deaths from causes other than the disease itself. Thus, the total number of people in all three compartments together is constant through time; specifically,

$$S(t) + I(t) + R(t) = N,$$

where $N$ is the size of the population; notice that (3.1.4) implies that

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0.$$

The assumption (3.1.4) is commonly referred to as mass balance (e.g., Reluga, 2004). Since the SIR model in (3.1.1)-(3.1.3) assumes members of the recovered class can rejoin the susceptible class, it is also referred to as a SIRS model in some articles (e.g., Dushoff et al., 2004); the traditional SIR model is obtained when $\phi = 0$.

In past decades, this deterministic continuous-time SIR model, or classic SIR model, and its various extensions (e.g., including birth and death rates, migration, etc.), have been used extensively for infectious-disease estimation and prediction in large and well mixed populations (e.g., Schenzle, 1984; Anderson and May, 1991; Keeling and Rohani, 2007). The classic SIR model is appealing, mainly because the model of the dynamical epidemic process is straightforward and its parameters have a simple interpretation. However, there are various sources of uncertainty in the
model: First, there is uncertainty in the counts \(\{S(t), I(t), R(t)\}\) themselves; that is, the counts in the compartments are observed with error. Second, the rather simple model (3.1.1)-(3.1.3) may not capture the dynamics of the epidemic exactly, and third, the values of the parameters \(\beta, \gamma, \) and \(\phi\) are uncertain.

A variety of stochastic models have been developed recently, through a probabilistic mechanism that involves a Markov chain of SIR states (e.g., Allen, 2003; Xu et al., 2007). Some more recent stochastic models involve complex networks to avoid the assumption of homogenous mixing (e.g., Halloran et al., 2002; Zhou et al., 2006; Volz, 2008). However, these stochastic models ignore the noisy nature of data, and they improperly apply mass balance to the observed counts. Furthermore, these models typically rely on many carefully chosen parameters, such as transmission rates, recovery rates, and so forth in heterogeneous populations; uncertainty in where the parameter vector is located in the parameter space is not accounted for. Our strategy is to deal with each source of uncertainty using a Bayesian hierarchical statistical model, and we make the dynamical structure embodied in (3.1.1)-(3.1.3) stochastic.

Bayesian hierarchical models have proved popular for mapping non-infectious diseases; while they aim to capture the true spatial process hidden behind noisy data (e.g., Besag et al., 1991; Carlin and Banerjee, 2002), their process models and parameter models are not appropriate for epidemics. Those that do have a dynamical spatial statistical component have not generally been parameterized in terms of the interpretable components of the epidemic (e.g., Mugglin et al., 2002).

Recently, Đukić et al. (2010) proposed a (non-spatial) dynamical Bayesian hierarchical model that is a state-space extension of a classic SEIR model (SEIR has a compartment of exposed (E) individuals, beyond the classic SIR model). In their model, they pay attention to the underlying true process hidden behind the noisy data, and they incorporate a source of variation that captures randomness in the
(hidden) epidemic process. However, they ignore the mass-balance property when incorporating the extra source of variation, which may introduce biased results. Their model is based on rates rather than counts; and they develop a particle learning algorithm for obtaining the posterior distribution and updating it as more data arrive.

In this chapter, we return to the classic SIR model for motivation, and we propose a mass-balanced (discrete-time) Bayesian hierarchical SIR, or HSIR, model, which is based directly on counts and correctly imposes mass balance on the underlying true counts, rather than on the observed counts. Our model captures the randomness in the epidemic process by assuming that the dynamical process occurs on a log-odds-ratio scale, transformed from the mass-balanced true counts. We believe this form of an infectious-disease model is new, and it shows how more general hierarchical statistical models for infectious diseases can be built for heterogeneous populations.

In Section 3.2, we propose our HSIR model (and an extension) for infectious-disease data. The actual computations associated with the posterior analysis involve local linearization of difference equations; see Section 3.3. In Section 3.4, we simulate data sets from an HSIR model (and its extension), as well as from a Neo-classic SIR (NSIR) model that incorporates observation error, and then we infer all unknowns of the model through Markov chain Monte Carlo (MCMC) analysis. Comparisons are also given to classic-SIR-model based inference. Discussion and conclusions are given in Section 3.5.

### 3.2 Bayesian Hierarchical Statistical SIR (HSIR) Models

Assume that there is a true unobserved process hidden behind the observed epidemic counts, which we incorporate into the framework of a Bayesian hierarchical statistical model. A Bayesian hierarchical model typically consists of three components: the data model (i.e., the conditional distribution of the data given hidden
processes and parameters); the process model (i.e., the conditional distribution of
the hidden processes given parameters); and the parameter model (i.e., the prior
distribution of the parameters).

3.2.1 The HSIR Model

3.2.1.1 Data Model

Observations in influenza epidemiology usually appear as counts. Rather than
using Gaussian distributions to (approximately) model the rates derived from the
counts (e.g., Dukić et al., 2010), our approach is to model the raw counts directly. To
ensure the discrete nature of the process of influenza, the data model is assumed to
consist of (conditionally) independent Poisson distributions that are discrete in time.
That is, for time points \( t = 1, 2, ..., T \), in units of \( \Delta \) days, the data model is

\[
Z_S(t) | P_S(t) \sim \text{ind. Poisson}(\lambda_N P_S(t)),
\]

\[
Z_I(t) | P_I(t) \sim \text{ind. Poisson}(\lambda_N P_I(t)),
\]

where \( Z_S(t) \) and \( Z_I(t) \) are the observed number of susceptible and infectious indi-
viduals at time \( t \), respectively; \( \lambda_N \) denotes the true total population count and \( P_S(t) \)
and \( P_I(t) \) are the underlying true rates of susceptible and infectious individuals at
time \( t \), respectively. Since \( \lambda_N \) is known from demography, the observed number of
recovered individuals, \( Z_R(t) \), is easily obtained by subtraction. Thus, the data are
\( \{(Z_S(t), Z_I(t)) : t = 1, 2, ..., T\} \).

3.2.1.2 Process Model

Recall the classic SIR model defined by (3.1.1)-(3.1.3), where it is assumed that
mass balance happens on the observed population. Because of the noisy nature of
observations, this is not true. The appropriate place to impose mass balance is on
the true (hidden) process. That is, for \( t = 1, 2, \ldots \), we have

\[
\lambda_S(t) + \lambda_I(t) + \lambda_R(t) = \lambda_N, \quad (3.2.3)
\]

where \( \lambda_S(t) \), \( \lambda_I(t) \), and \( \lambda_R(t) \) are the underlying true counts of susceptible, infectious, and recovered individuals at time \( t \), respectively. Now define the true (hidden) rates, \( P_S(t) \), \( P_I(t) \), and \( P_R(t) \), via

\[
\lambda_S(t) \equiv \lambda_N P_S(t); \quad (3.2.4)
\]

\[
\lambda_I(t) \equiv \lambda_N P_I(t); \quad (3.2.5)
\]

\[
\lambda_R(t) \equiv \lambda_N P_R(t), \quad (3.2.6)
\]

where \( P_R(t) \) denotes the underlying true rate of recovered individuals at time \( t \). Then by substituting (3.2.4)-(3.2.6) into (3.2.3), it is straightforward to see that the mass balance in (3.2.3) can be rewritten as,

\[
P_S(t) + P_I(t) + P_R(t) = 1. \quad (3.2.7)
\]

Notice that under the mass-balance assumption in (3.2.7), the value of \( P_R(t) \) is determined from \( P_S(t) \) and \( P_I(t) \) by subtraction:

\[
P_R(t) = 1 - P_S(t) - P_I(t); \quad t = 1, 2, \ldots. \quad (3.2.8)
\]

Recall that classic SIR ODEs defined in (3.1.1)-(3.1.3) give easily interpretable
dynamics, in which individuals move from the susceptible state, to the infectious
state, to the recovered state (some individuals go back to being susceptible again).
We refer to this as SIRS dynamics, and we recognize that \( t \) is discrete (in units of \( \Delta \) days) by deriving a set of deterministic difference equations on the hidden process, \( \lambda_S(t) \), \( \lambda_I(t) \), and \( \lambda_R(t) \). That is, for \( t = 1, 2, \ldots \), the process model becomes,

\[
\lambda_S(t + 1) = \lambda_S(t) - \beta \Delta \lambda_S(t) \lambda_I(t) + \phi \Delta \lambda_R(t), \quad (3.2.9)
\]

47
\[ \lambda_I(t+1) = \lambda_I(t) + \beta \Delta \lambda_S(t) \lambda_I(t) - \gamma \Delta \lambda_I(t), \]  
\[ \lambda_R(t+1) = \lambda_R(t) + \gamma \Delta \lambda_I(t) - \phi \Delta \lambda_R(t), \]  

where the SIRS dynamics has been preserved, and the rate parameters \( \beta, \phi, \) and \( \gamma \) are in units of per day \( (d^{-1}) \).

According to the definition of \( \lambda_S(t), \lambda_I(t), \) and \( \lambda_R(t) \) in (3.2.4)-(3.2.6), equations (3.2.9)-(3.2.11) can be rewritten based on the true proportions, \( P_S(t), P_I(t), \) and \( P_R(t) \),

\[ P_S(t+1) = P_S(t) - \beta \Delta \lambda_N P_S(t) P_I(t) + \phi \Delta P_R(t), \]  
\[ P_I(t+1) = P_I(t) + \beta \Delta \lambda_N P_S(t) P_I(t) - \gamma \Delta P_I(t), \]  
\[ P_R(t+1) = P_R(t) + \gamma \Delta P_I(t) - \phi \Delta P_R(t). \]  

The deterministic difference equations in (3.2.12)-(3.2.14) are unable to capture all the uncertainties in the hidden epidemic process. Possible reasons for this are heterogeneous populations and the existence of other categories such as exposed individuals. To handle this complexity, we use stochastic components to capture the uncertainties in the true process, but we are careful to preserve the mass-balance constraint. Consider the logit transformations on \( \{P_S(t)\} \) and \( \{P_I(t)\} \):

\[ W_S(t) \equiv \log \left( \frac{P_S(t)}{P_R(t)} \right), \]  
\[ W_I(t) \equiv \log \left( \frac{P_I(t)}{P_R(t)} \right), \]

where \( W_S(t) \) and \( W_I(t) \) are the log odds ratios of susceptible over recovered population, and infectious over recovered population, respectively, at time \( t \). The logit transformations change the support of \( \{P_S(t)\} \) and \( \{P_I(t)\} \) from \([0,1]\) to \((-,\infty)\) for \( W_S(t) \) and \( W_I(t) \). Therefore, by modeling on the odds-ratio scale (W-scale),
uncertainties in the epidemic process can be modeled flexibly while preserving the associated mass-balance constraint.

For discrete time $t = 1, 2, \ldots$, in units of $\Delta$ days, our process model is

$$W(t + 1) = \mu^W(t) + \xi(t + 1),$$

(3.2.17)

where $W(t)$ denotes the hidden log-odds-ratio vector given by $W(t) \equiv (W_S(t), W_I(t))'$. We now discuss each of the components of (3.2.17), in turn. The vector $\mu^W(t) \equiv (\mu_S^W(t), \mu_I^W(t))'$ is the dynamical process that captures the temporal dependence. In Appendix B.1, from the epidemic process defined in (3.2.12)-(3.2.14) and the relationship between $W(t)$ and $P(t)$ defined in (3.2.15) and (3.2.16), we are able to derive the nonlinear dynamical structure of $\mu^W(t)$ as below. This derivation retains the SIRS dynamics on the hidden process; that is, for discrete time $t = 1, 2, \ldots$, in units of $\Delta$ days,

$$\mu_S^W(t) = W_S(t) + \log \left[ 1 + \frac{\phi \Delta}{\exp(W_S(t))} - \frac{\beta \Delta \lambda_N \exp(W_I(t))}{1 + \exp(W_S(t)) + \exp(W_I(t))} \right],$$

(3.2.18)

and

$$\mu_I^W(t) = W_I(t) + \log \left[ 1 - \gamma \Delta + \frac{\beta \Delta \lambda_N \exp(W_S(t))}{1 + \exp(W_S(t)) + \exp(W_I(t))} \right],$$

(3.2.19)

where recall that $\beta$, $\gamma$, and $\phi$, are the transmission rate, recovery rate, and loss-of-immunity rate per day, respectively.
The vector, $\xi(t) \equiv (\xi_S(t), \xi_I(t))'$, is the small-scale variation that captures the uncertainties in the hidden epidemic process. For $t = 1, 2, ...$, we define

$$\xi(t) \sim \text{MVN}(0, \Sigma_{\xi}(t)),$$

a multivariate normal distribution with mean 0 and diagonal covariance matrix $\Sigma_{\xi}(t) \equiv \text{diag}(\sigma^2_{\xi_S}(t), \sigma^2_{\xi_I}(t))$, with nonnegative variance-components, $\sigma^2_{\xi_S}(t)$ and $\sigma^2_{\xi_I}(t)$. For the sake of simplicity, in this chapter, we assume that $\sigma^2_{\xi_S}(t) = \sigma^2_{\xi_I}(t)$, for all $t = 1, 2, ...$

Notice that the strategy of transforming from the hidden proportion scale ($P$-scale) to the hidden log-odds-ratio scale ($W$-scale) and adding the small-scale variation on the $W$-scale rather than on the $P$-scale, is key to retaining the mass-balance constraint while allowing flexible SIRS dynamics to be handled. To our knowledge, this is a new approach to infectious-disease modeling; other Bayesian approaches (e.g., Dukić et al., 2010) are unable to preserve mass balance after building uncertainties into the process model.

### 3.2.1.3 Parameter Model

To complete the HSIR model, we now specify the joint prior distribution for the parameters, which includes the transmission rate per day, $\beta$; the rate of recovery per day, $\gamma$; the loss-of-immunity rate per day, $\phi$; and the variance components of the small-scale variations, $\{\sigma^2_{\xi_S}\}$ and $\{\sigma^2_{\xi_I}\}$. Notice that the ODEs in (3.1.1)-(3.1.3) impose a natural constraint on $\beta$; that is, for any time $t$,

$$\beta S(t)I(t) < N - (1 - \phi)R(t),$$

because the number of individuals that become infectious at a certain time $t$ should be less than the total number that could be infected at that time. Hence,

$$0 < \beta < \frac{N - (1 - \phi)R(t)}{S(t)I(t)}.$$
In the context of the HSIR model, this amounts to ensuring that $\beta$ is bounded above by a hyperparameter, $\beta_{\text{max}}$. Furthermore, it is straightforward to see that $\gamma \in [0, 1]$ and $\phi \in [0, 1]$, due to their definition. Typically, information about the recovery rate $\gamma$ is easier to obtain than the other rate parameters; we apply a logit transformation to $\gamma$,

$$
\theta_{\gamma} = \log\left(\frac{\gamma}{1 - \gamma}\right),
$$

(3.2.23)

and assign a Gaussian prior to $\theta_{\gamma}$.

Assuming statistical independence of parameters and using $[Y]$ as generic notation for the probability distribution of $Y$, we assume the parameter model can be written as

$$
[\beta, \theta_{\gamma}, \phi, \sigma_{\xi_S}^2, \sigma_{\xi_I}^2] = [\beta][\theta_{\gamma}][\phi][\sigma_{\xi_S}^2][\sigma_{\xi_I}^2],
$$

(3.2.24)

where we specify the prior distributions of individual parameters as follows:

$$
\beta \sim \text{Uniform}[0, \beta_{\text{max}}],
$$

$$
\theta_{\gamma} \sim \text{Normal}(\mu_{\theta_{\gamma}}, \sigma_{\theta_{\gamma}}^2),
$$

$$
\phi \sim \text{Uniform}[0, 1],
$$

$$
\sigma_{\xi_S}^2 \sim \text{Inverse Gamma} (a_{\xi_S}, b_{\xi_S}),
$$

$$
\sigma_{\xi_I}^2 \sim \text{Inverse Gamma} (a_{\xi_I}, b_{\xi_I}).
$$

Notice that the uniform distributions on $\beta$ and $\phi$ could be easily replaced, respectively, by a Generalized Beta distribution on $[0, \beta_{\text{max}}]$ and a Beta distribution on $[0, 1]$.

In practice, the hyperparameter $\beta_{\text{max}}$ could be data-driven using (3.2.22):

$$
\beta_{\text{max}} \equiv \min_t \left(\frac{\lambda_N - Z_R(t)}{Z_S(t)Z_I(t)}\right),
$$

(3.2.25)
where

$$Z_R(t) = \max\{1, \lambda_N - Z_S(t) - Z_I(t)\}. \quad (3.2.26)$$

The hyperparameter $\mu_\theta$, could be data-driven through the inverse relation between recovery rate and the mean duration of the infectious period for individuals (e.g., Lloyd, 2001), which is about 3 days for common influenzas (Centers for Disease Control; http://www.cdc.h1n1flu/recommendations.htm). In our case, we use $\mu_\theta = \log(0.33/(1 - 0.33)) = -0.708$, with $\sigma^2_\theta = 0.01$. The other hyperparameters, namely, $a_{\xi_S}, b_{\xi_S}, a_{\xi_I},$ and $b_{\xi_I}$, can be further specified, for example, the choice, $a_{\xi_S} = a_{\xi_I} = 0.25$ and $b_{\xi_S} = b_{\xi_I} = 0.4$, can result in a fairly vague prior for the variance components.

### 3.2.2 An Extension of the HSIR Model

#### 3.2.2.1 Data Model

In the HSIR model presented in Section 3.2.1, we incorporate the uncertainties after transforming from the $P$-scale to the $W$-scale. However, the deterministic transformation defined in (3.2.15) and (3.2.16) may result in biases when inverting the transformation. Therefore, we extend the process model in order to capture uncertainty in the transformation, while leaving the data model alone. We call this generalization the extended HSIR model.

#### 3.2.2.2 Process Model

The process model of the extended HSIR model involves an extra level of conditional probability modeling. Instead of applying the deterministic transformation as defined in (3.2.15) and (3.2.16) to the true proportion vector, $P(t)$, we now assume that, conditional on $W(t)$, $P(t)$ is from a joint Logistic Normal distribution (e.g., Aitchison and Shen, 1980). That is, for $t = 1, 2, ..., \text{in units of } \Delta \text{ days},$

$$P(t)|W(t), \Sigma_p(t) \sim \text{ind. Logistic Normal}(W(t), \Sigma_p(t)), \quad (3.2.27)$$
where $\mathbf{\Sigma}_P(t) \equiv \text{diag}(\sigma^2_{Ps}(t), \sigma^2_{P_I}(t))$ is a $2 \times 2$ diagonal covariance matrix with non-negative variance components, $\sigma^2_{Ps}(t)$ and $\sigma^2_{P_I}(t)$, on the diagonal. So,

$$
[\mathbf{P}(t)|\mathbf{W}(t), \mathbf{\Sigma}_P(t)] = \frac{1}{2\pi} |\mathbf{\Sigma}_P(t)|^{-1/2} (P_S(t)P_I(t)P_R(t))^{-1} \cdot \exp \left\{-\frac{1}{2} [\mathbf{Y}_P(t) - \mathbf{W}(t)]' \mathbf{\Sigma}_P(t)^{-1} [\mathbf{Y}_P(t) - \mathbf{W}(t)] \right\},
$$

(3.2.28)

where

$$
\mathbf{Y}_P(t) \equiv \left( \log \left( \frac{P_S(t)}{P_R(t)} \right), \log \left( \frac{P_I(t)}{P_R(t)} \right) \right)'.
$$

We use $E[Y_1|Y_2]$ as generic notation for the mean of $Y_1$ conditional on $Y_2$, and we define

$$
E_{\mathbf{P}_{\mathbf{\Sigma}}(t)}|\mathbf{W}(t) \equiv E \left( \frac{P_S(t)}{P_R(t)} \mathbf{W}(t), \mathbf{\Sigma}_P(t) \right) \quad (3.2.29)
$$

$$
E_{\mathbf{P}_{\mathbf{\Sigma}}(t)}|\mathbf{W}(t) \equiv E \left( \frac{P_I(t)}{P_R(t)} \mathbf{W}(t), \mathbf{\Sigma}_P(t) \right). \quad (3.2.30)
$$

According to Aitchison and Shen (1980),

$$
E_{\mathbf{P}_{\mathbf{\Sigma}}(t)}|\mathbf{W}(t) = \exp \left( W_S(t) + \frac{1}{2} \sigma^2_{Ps}(t) \right) \quad (3.2.31)
$$

$$
E_{\mathbf{P}_{\mathbf{\Sigma}}(t)}|\mathbf{W}(t) = \exp \left( W_I(t) + \frac{1}{2} \sigma^2_{P_I}(t) \right), \quad (3.2.32)
$$

for $t = 1, 2, \ldots$. Uncertainty in the transformation is captured through the variance components, $\{\sigma^2_{Ps}(t)\}$ and $\{\sigma^2_{P_I}(t)\}$. Furthermore, for $t = 1, 2, \ldots$, we define

$$
D_S(t) \equiv W_S(t) + \frac{1}{2} \sigma^2_{Ps}(t), \quad (3.2.33)
$$

$$
D_I(t) \equiv W_I(t) + \frac{1}{2} \sigma^2_{P_I}(t), \quad (3.2.34)
$$

to be the \emph{adjusted log odds ratios} of susceptible over recovered, and infectious over recovered populations at time $t$, respectively. Notice that when $\sigma^2_{Ps}(t) \equiv 0 \equiv \sigma^2_{P_I}(t)$, we obtain the deterministic transformations (3.2.15) and (3.2.16).
We are now able to incorporate the small-scale variation on the $W$-scale, as in (3.2.17), to handle the uncertainties in the underlying epidemic process. For discrete time $t = 1, 2, ..., $ in units of $\Delta$ days, the second level of the process model of the extended HSIR model is:

$$W(t + 1) = \mu^D(t) - \frac{1}{2} V(t + 1) + \xi(t + 1),$$

(3.2.35)

where recall that $W(t) \equiv (W_S(t), W_I(t))'$ denotes the hidden odds-ratio vector. We now discuss each of the components of (3.2.35), in turn. The vector $\mu^D(t) \equiv (\mu^D_S(t), \mu^D_I(t))'$ is the dynamical process that captures the temporal dependence. In Appendix B.3, from the epidemic process defined in (3.2.12)-(3.2.14) and the relationship between $W(t)$ and $P(t)$ defined in (3.2.27), we are able to derive the nonlinear dynamical structure of $\mu^D(t)$ as below, to retain the SIRS dynamics on the hidden process. That is, for discrete time $t = 1, 2, ..., $ in units of $\Delta$ days,

$$\mu^D_S(t) = D_S(t)$$

$$+ \log \left[ 1 + \frac{\phi \Delta}{\exp(D_S(t))} - \frac{\beta \Delta \lambda N \exp(D_I(t))}{1 + \exp(D_S(t)) + \exp(D_I(t))} \right]$$

$$- \log \left[ 1 + \exp(D_I(t)) - \phi \Delta \right],$$

(3.2.36)

and

$$\mu^D_I(t) = D_I(t)$$

$$+ \log \left[ 1 - \gamma \Delta + \frac{\beta \Delta \lambda N \exp(D_S(t))}{1 + \exp(D_S(t)) + \exp(D_I(t))} \right]$$

$$- \log \left[ 1 + \gamma \Delta \exp(D_I(t)) - \phi \Delta \right],$$

(3.2.37)

where recall that $\beta$, $\gamma$, and $\phi$, are the transmission rate, recovery rate, and loss-of-immunity rate per day, respectively. Now $\sigma^2_{P_S}(t)$ and $\sigma^2_{P_I}(t)$ are the variance components that capture the uncertainty in the transformation. For the sake of simplicity, in this chapter we assume that $\sigma^2_{P_S}(t) \equiv \sigma^2_{P_S}$ and $\sigma^2_{P_I}(t) \equiv \sigma^2_{P_I}$, for $t = 1, 2, ....$
In (3.2.35), we define
\[ V(t + 1) \equiv (\sigma_{P_S}^2(t + 1), \sigma_{P_I}^2(t + 1))', \] (3.2.38)
for \( t = 1, 2, \ldots \). Thus, \( \frac{1}{2}V(t + 1) \) can be explained as an adjustment to \( W(t + 1) \) for possible bias due to the transformation.

Similar to the construction of the HSIR model in Section 3.2.1.2, the vector of small-scale variation, \( \xi(t) \equiv (\xi_S(t), \xi_I(t))' \), is modeled via
\[ \xi(t) \sim MVN(0, \Sigma(t)), \] for \( t = 1, 2, \ldots \), where \( \Sigma(t) \equiv \text{diag}(\sigma_{\xi_S}^2(t), \sigma_{\xi_I}^2(t)) \) is a \( 2 \times 2 \) diagonal covariance matrix with nonnegative variance-components, \( \sigma_{\xi_S}^2(t) \) and \( \sigma_{\xi_I}^2(t) \), on the diagonal. Again, we assume that \( \sigma_{\xi_S}^2(t) = \sigma_{\xi_S}^2 \) and \( \sigma_{\xi_I}^2(t) = \sigma_{\xi_I}^2 \), for \( t = 1, 2, \ldots \).

Notice that the extra level of variability does not affect the mass-balance constraint given by (3.2.7). Moreover, (3.2.36) and (3.2.37) reduce to (3.2.18) and (3.2.19), respectively, when the variance components, \( \{\sigma_{P_S}^2(t)\} \) and \( \{\sigma_{P_I}^2(t)\} \) are all identically zero.

### 3.2.2.3 Parameter Model

The parameters now include \( \beta, \gamma, \phi, \sigma_{\xi_S}^2, \) and \( \sigma_{\xi_I}^2 \), as in the HSIR model, as well as the variance components, \( \sigma_{P_S}^2 \) and \( \sigma_{P_I}^2 \), which capture uncertainty due to the transformation from the \( P \)-scale to the \( W \)-scale. Hence, the joint prior distribution for parameters of the extended HSIR model can be written as:
\[ [\beta, \theta, \phi, \sigma_{\xi_S}^2, \sigma_{\xi_I}^2, \sigma_{P_S}^2, \sigma_{P_I}^2] = [\beta][\theta][\phi][\sigma_{\xi_S}^2][\sigma_{\xi_I}^2][\sigma_{P_S}^2][\sigma_{P_I}^2], \] (3.2.39)
where \( \theta, \) is the logit transformation of \( \gamma \) as defined in (3.2.23). In (3.2.39), the prior distributions, \( [\beta], [\theta], [\phi], [\sigma_{\xi_S}^2], \) and \( [\sigma_{\xi_I}^2] \), are the same as those of the HSIR model; see Section 3.2.1.3. For the variance components \( \sigma_{P_S}^2 \) and \( \sigma_{P_I}^2 \), we define
\[ \sigma_{P_S}^2 \sim \text{Inverse Gamma} (a_{P_S}, b_{P_S}), \]
\[ \sigma^2_{pi} \sim \text{Inverse Gamma} \left( a_{pi}, b_{pi} \right). \]

The hyperparameters \( a_{ps}, b_{ps}, a_{pi}, b_{pi} \) need to be specified; for example, the choice \( a_{ps} = a_{pi} = 0.25 \) and \( b_{ps} = b_{pi} = 0.4 \), results in a fairly vague prior for the variance components.

### 3.3 W-Scale Approximations for HSIR and Extended HSIR Models

#### 3.3.1 A Linear Approximation to the HSIR Model

The HSIR process model defined in (3.2.17) is nonlinear and can cause computational difficulties. Here, we approximate it with a well calibrated Gaussian linear process: For \( t = 1, 2, ..., \), replace (3.2.17) in the HSIR model with

\[ W(t + 1) = \mu^{LW}(t) + \zeta(t + 1), \quad (3.3.1) \]

where recall that \( W(t) \equiv (W_S(t), W_I(t))' \) is the true log-odds-ratio vector. In (3.3.1), the vector \( \mu^{LW}(t) \equiv (\mu^{LW}_S(t), \mu^{LW}_I(t))' \) is a linear dynamical process derived through Taylor-series expansions that approximate the nonlinear stochastic process \( \mu^W \) defined in (3.2.18) and (3.2.19). From Appendix B.2, for \( t = 1, 2, ..., \),

\[ \begin{align*}
\mu^{LW}_S(t) &= J_0(t) + J_1(t)W_S(t) + J_2(t)W_I(t), \\
\mu^{LW}_I(t) &= J_3(t) + J_4(t)W_S(t) + J_5(t)W_I(t),
\end{align*} \quad (3.3.2) \]

where

\[ \begin{align*}
J_0(t) &\equiv \log \hat{A}_1(t) + \frac{1}{\hat{A}_1(t)} + \phi \Delta \left[ \frac{\exp(\hat{A}_5(t))(1 - \hat{A}_5(t))}{\hat{A}_1(t)} + \frac{1}{\hat{A}_2(t)} \right] \\
&\quad - \frac{\beta \Delta \lambda_N}{\hat{A}_1(t)} \left( 1 - \frac{1}{1 - \hat{A}_7(t)} + \frac{B_0(t) + \hat{A}_7(t)}{(1 - \hat{A}_7(t))^2} \right) \\
&\quad - \log \hat{A}_2(t) - \frac{1}{\hat{A}_2(t)} - \frac{\gamma \Delta}{\hat{A}_2(t)} e^{\hat{A}_6(t)}(1 - \hat{A}_6(t)).
\end{align*} \quad (3.3.4) \]
where

\( J_1(t) \equiv 1 - \frac{\phi \Delta \exp(\hat{A}_5(t))}{\hat{A}_1(t)} - \frac{\beta \Delta \lambda N B_1(t)}{A_1(t)(1 - \hat{A}_7(t))^2} \); (3.3.5)

\( J_2(t) \equiv -\frac{\beta \Delta \lambda N B_2(t)}{A_1(t)(1 - \hat{A}_7(t))^2} - \frac{\gamma \Delta e^{\hat{A}_6(t)}}{A_2(t)} \); (3.3.6)

\( J_3(t) \equiv \log \hat{A}_3(t) + \frac{1 - \gamma \Delta}{\hat{A}_3(t)} + \frac{\beta \Delta \lambda N}{\hat{A}_3(t)(1 - \hat{A}_10(t))} - \frac{\beta \Delta \lambda N (B_3(t) + \hat{A}_10(t))}{\hat{A}_3(t)(1 - \hat{A}_10(t))^2} \)

\[ -\log \hat{A}_2(t) - \frac{1}{\hat{A}_2(t)} - \frac{\gamma \Delta e^{\hat{A}_6(t)}(1 - \hat{A}_6(t))}{\hat{A}_2(t)} + \frac{\phi \Delta}{\hat{A}_2(t)} \]; (3.3.7)

\( J_4(t) \equiv -\frac{\beta \Delta \lambda N B_4(t)}{A_3(t)(1 - \hat{A}_10(t))^2} \); (3.3.8)

\( J_5(t) \equiv 1 - \frac{\beta \Delta \lambda N B_5(t)}{A_3(t)(1 - \hat{A}_10(t))^2} - \frac{\gamma \Delta e^{\hat{A}_6(t)}}{\hat{A}_2(t)} \). (3.3.9)

Recall that \( \beta, \gamma, \) and \( \phi \) are the transmission rate, recovery rate, and loss-of-immunity rate, respectively, per day, and \( \lambda_N \) is the given total population. Also,

\( B_0(t) \equiv e^{\hat{A}_4(t)}(1 - \hat{A}_4(t))B^*(t) \); (3.3.10)

\( B_1(t) \equiv \frac{e^{\hat{A}_5(t)} + \hat{A}_4(t)(1 - \hat{A}_4(t))}{(1 - \hat{A}_9(t))^2} - e^{\hat{A}_4(t)}B^*(t) \); (3.3.11)

\( B_2(t) \equiv e^{\hat{A}_4(t)}B^*(t) \); (3.3.12)

\( B_3(t) \equiv e^{\hat{A}_4(t)}(1 - \hat{A}_4(t)) + e^{\hat{A}_5(t)}(1 - \hat{A}_5(t)) \); (3.3.13)

\( B_4(t) \equiv -e^{\hat{A}_4(t)} - e^{\hat{A}_5(t)} \); (3.3.14)

\( B_5(t) \equiv e^{\hat{A}_4(t)} \); (3.3.15)

where

\( B^*(t) = \frac{1}{1 - \hat{A}_9(t)} + \frac{e^{\hat{A}_5(t)}(\hat{A}_5(t) - 1) - \hat{A}_9(t)}{(1 - \hat{A}_9(t))^2} \). (3.3.16)
Notice that by combining equations (3.3.1)-(3.3.16), we are able to write \( W(t+1) \) in a multivariate autoregressive form:

\[
W(t+1) = C(t) + H(t)W(t) + \zeta(t+1), \tag{3.3.17}
\]

where \( C(t) = (J_0(t), J_3(t))' \) and

\[
H = \begin{pmatrix}
J_1(t) & J_2(t) \\
J_4(t) & J_5(t)
\end{pmatrix}.
\]

### Table 3.1: Table of the initializations \( \{\hat{A}_i(t) : i = 1, ..., 10\} \) and the quantities \( \{A_i(t) : i = 1, ..., 10\} \) that they approximate.

<table>
<thead>
<tr>
<th>Value</th>
<th>Initializations (( \hat{A}(t) ))</th>
<th>Nonlinear Components in HSIR</th>
<th>Nonlinear Components in Extended HSIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A_1(t))</td>
<td>(1 + \frac{\Delta \phi S(t)}{Z_S(t)} - \Delta \beta Z(t))</td>
<td>(1 + \frac{\Delta \phi}{e^{W_S(t)}} - \frac{\Delta \beta N S(t) I(t)}{1 + e^{W_S(t)} + e^{W_I(t)}})</td>
<td>(1 + \frac{\Delta \phi}{e^{W_S(t)}} - \frac{\Delta \beta N S(t) I(t)}{1 + e^{W_S(t)} + e^{W_I(t)}})</td>
</tr>
<tr>
<td>(A_2(t))</td>
<td>(1 + \Delta \gamma \frac{Z(t)}{Z(t)})</td>
<td>(1 + \Delta \gamma - \Delta \phi)</td>
<td>(1 + \Delta \gamma - \Delta \phi)</td>
</tr>
<tr>
<td>(A_3(t))</td>
<td>((1 - \Delta \gamma N(t) + \Delta \beta Z_S(t)))</td>
<td>((1 - \Delta \gamma) + \frac{\Delta \beta N S(t) I(t)}{1 + e^{W_S(t)} + e^{W_I(t)}})</td>
<td>((1 - \Delta \gamma) + \frac{\Delta \beta N S(t) I(t)}{1 + e^{W_S(t)} + e^{W_I(t)}})</td>
</tr>
<tr>
<td>(A_4(t))</td>
<td>(\log \frac{Z(t)}{Z(t)})</td>
<td>((W_I(t) - W_S(t)))</td>
<td>((D_I(t) - D_S(t)))</td>
</tr>
<tr>
<td>(A_5(t))</td>
<td>(-\log \frac{Z(t)}{Z(t)})</td>
<td>(-W_S(t))</td>
<td>(-D_S(t))</td>
</tr>
<tr>
<td>(A_6(t))</td>
<td>(\log \frac{Z(t)}{Z(t)})</td>
<td>(W_I(t))</td>
<td>(D_I(t))</td>
</tr>
<tr>
<td>(A_7(t))</td>
<td>(\frac{Z_N - Z(t)}{Z(t)})</td>
<td>(-\frac{e^{W_S(t)}}{1 + e^{W_S(t)}})</td>
<td>(-\frac{e^{W_S(t)}}{1 + e^{W_S(t)}})</td>
</tr>
<tr>
<td>(A_8(t))</td>
<td>(\frac{Z(t)}{Z(t)})</td>
<td>(-e^{W_S(t)})</td>
<td>(-e^{W_S(t)})</td>
</tr>
<tr>
<td>(A_9(t))</td>
<td>(\frac{Z(t)}{Z(t)})</td>
<td>(-e^{-W_S(t)})</td>
<td>(-e^{-W_S(t)})</td>
</tr>
<tr>
<td>(A_{10}(t))</td>
<td>(\frac{Z_N - Z(t)}{Z(t)})</td>
<td>(-\frac{1 + e^{W_S(t)}}{e^{W_S(t)}})</td>
<td>(-\frac{1 + e^{W_S(t)}}{e^{W_S(t)}})</td>
</tr>
</tbody>
</table>

The appropriate interpretation of \( \hat{A}_i(t), i = 1, ..., 10 \), is as an initialization of the Taylor-series expansion of the nonlinear process \( \mu^W(t) \) in (3.2.17). Formulas for \( \hat{A}_i(t) \) and the quantity \( A_i(t) \) that it approximates are given in Table 3.1. In practice,
we use empirical values obtained from data \(\{Z_S(t)\} \) and \(\{Z_I(t)\} \) to obtain \(\hat{A}_i(t)\) close to \(A_i(t)\). From the transformation in (3.2.15) and (3.2.16), we can obtain

\[
\lambda_N P_S(t) = \frac{\lambda_N \exp(W_S(t))}{1 + \exp(W_S(t)) + \exp(W_I(t))}
\]

\[
\lambda_N P_I(t) = \frac{\lambda_N \exp(W_I(t))}{1 + \exp(W_S(t)) + \exp(W_I(t))}
\]

\[
\lambda_N P_R(t) = \frac{\lambda_N}{1 + \exp(W_S(t)) + \exp(W_I(t))}.
\]

Now \(\lambda_N P_S(t)\), \(\lambda_N P_I(t)\), and \(\lambda_N P_R(t)\) are the means of \(\{Z_S(t)\}\), \(\{Z_I(t)\}\), and \(\{Z_I(t)\}\), and hence the values in the column “Initializations (\(\hat{A}(t)\))” in Table 3.1 are reasonable choices for \(\{\hat{A}_i(t)\}\). If data are not available at some time point, we obtain \(\{\hat{A}_i(t)\}\) in Table 3.1 by replacing the observed data \(\{Z_S(t)\}\) and \(\{Z_I(t)\}\) with the predicted counts provided by a simple model, such as the classic SIR model. The goal is to obtain \(\{\hat{A}_i(t)\}\) as close as possible to \(\{A_i(t)\}\).

We further denote \(\beta_0\), \(\gamma_0\), and \(\phi_0\) in Table 3.1 to be initial estimates of \(\beta\), \(\gamma\), and \(\phi\), respectively, aiming for better performance in the Taylor-series expansions. We use estimates obtained from the classic SIR model (e.g., Anderson and May, 1991; Wearing et al., 2005; Burr and Chowell, 2006) as the values of \(\beta_0\), \(\gamma_0\), and \(\phi_0\). Further details of implementation can be found in Section 3.4.3. We performed sensitivity studies and noticed that the MCMC based on this linear approximation is not sensitive to the initializations (even in forecasting), because the small-scale variation terms in (3.3.1) can absorb the higher-order terms in the Taylor-series expansions (e.g., Cressie and Wikle, 2011, Section 7.3.3).

Now we discuss the small-scale variation vector \(\zeta(t) \equiv (\zeta_S(t), \zeta_I(t))'\) in (3.3.1) that captures the uncertainties in the epidemic process as well as the higher-order terms in the Taylor-series expansions. For \(t = 1, 2, \ldots\), we assume that

\[
\zeta(t) \sim \text{MVN}(0, \Sigma_\zeta(t)),
\]
where $\Sigma_\zeta(t) \equiv \operatorname{diag}(\sigma^2_{\zeta_S}(t), \sigma^2_{\zeta_I}(t))$, is the covariance matrix of $\zeta(t)$, and $\sigma^2_{\zeta_S}(t)$ and $\sigma^2_{\zeta_I}(t)$ are nonnegative variance-component parameters. Typically, the components of $\Sigma_\zeta(t)$ are larger than respective components of $\Sigma_\xi(t)$, because $\{\zeta(t)\}$ also captures the higher-order terms in the linear approximation. For the sake of simplicity, in this chapter we further assume that $\sigma^2_{\zeta_S}(t) = \sigma^2_{\zeta_S}$ and $\sigma^2_{\zeta_I}(t) = \sigma^2_{\zeta_I}$, for $t = 1, 2, \ldots$.

Like the data model, the parameter model is unchanged, except that $\xi$ is replaced with $\zeta$, see Section 3.2.1.3 for details. We refer to the hierarchical model that consists of the data model defined in (3.2.1) and (3.2.2), the linear dynamical process model for $\{W(t)\}$ defined in (3.3.1), and the parameter model defined in (3.2.24) with $\xi$ replaced by $\zeta$, as the approximate HSIR model, which we abbreviate to ASIR model.

### 3.3.2 A Linear Approximation to the Extended HSIR Model

The extended HSIR model defined in (3.2.35) is nonlinear and can be approximated by a well calibrated Gaussian linear process, using an approach similar to Section 3.3.1: For $t = 1, 2, \ldots$, replace (3.2.35) in the extended HSIR model with

$$W(t + 1) = \mu^{LD}(t) - \frac{1}{2} V(t + 1) + \zeta(t + 1),$$

(3.3.19)

where recall that $W(t) \equiv (W_S(t), W_I(t))'$ is the hidden log-odds-ratio vector, and the vector $\frac{1}{2} V(t + 1)$ is the adjustment on $W(t + 1)$ defined in (3.2.38). In (3.3.19), the vector $\mu^{LD}(t) \equiv (\mu^{LD}_S(t), \mu^{LD}_I(t))'$ is a linear dynamical process derived through Taylor-series expansions that approximate the nonlinear stochastic process $\mu^D$ defined in (3.2.36) and (3.2.37). From Appendix B.4, for $t = 1, 2, \ldots$,

$$\mu^{LD}_S(t) = J_0(t) + J_1(t)D_S(t) + J_2(t)D_I(t),$$

(3.3.20)

$$\mu^{LD}_I(t) = J_3(t) + J_4(t)D_S(t) + J_5(t)D_I(t),$$

(3.3.21)

where recall that $D_S(t)$ and $D_I(t)$ are adjusted log odds ratios defined in (3.2.33) and (3.2.34), respectively. The coefficients, $\{J_i(t) : i = 1, \ldots, 5\}$, are the same as that of
the ASIR model defined in (3.3.4)-(3.3.9) in Section 3.3.1, but \( \{A_i(t) : i = 1, \ldots, 10\} \) are modified and shown in the column “Nonlinear Components in Extended HSIR” in Table 3.1. Nevertheless, the initializations \( \{\hat{A}_i(t)\} \) obtained from the HSIR model are still appropriate.

Again, combining equations (3.3.19)-(3.3.21), we are able to write \( W(t+1) \) in a multivariate autoregressive form:

\[
W(t+1) = C(t) + H(t)W(t) + \frac{1}{2}H(t)V(t) - \frac{1}{2}V(t+1) + \zeta(t+1),
\]

where \( C(t) \) and \( H(t) \) are the same quantities that appear in equation (3.3.17).

The vector \( \zeta(t) \equiv (\zeta_s(t), \zeta_I(t))^T \) in equation (3.3.19) is the small-scale variation defined as in (3.3.18); \( \zeta(t) \) captures both the uncertainties in the epidemic process as well as the higher-order terms in the Taylor-series expansions. Like the data model, the parameter model is unchanged, except that \( \xi \) is replaced by \( \zeta \); see Section 3.2.2.3 for details.

We refer to the hierarchical model that consists of the data model defined in (3.2.1) and (3.2.2), the first level of the process model for \( \{P(t)\} \) defined in (3.2.27), the linear dynamical process model for \( \{W(t)\} \) defined in (3.3.19), and the parameter model defined in (3.2.39), with \( \xi \) replaced by \( \zeta \), as the extended ASIR model. This extends the ASIR model defined in Section 3.3.1.

### 3.4 Results of Analysis on Simulated Epidemics

#### 3.4.1 Simulated Data

The simulation of the epidemics are from three processes meant to mimic an H1N1 outbreak in Franklin County, Ohio. The three datasets are from an HSIR model (defined in Section 3.2.1), an extended HSIR model (defined in Section 3.2.2),
and a Neo-classic SIR model (defined below), which we refer to as the HSIR dataset, the ESIR dataset, and the NSIR dataset, respectively.

### 3.4.1.1 The HSIR dataset

For simulating the HSIR dataset, we select \( \gamma^* = 0.33 \) per day as the true value of \( \gamma \), since there is an inverse relation between \( \gamma \) and mean duration of the infectious period (e.g., Lloyd, 2001); this is about 3 days for common influenza (Centers for Disease Control; http://www.cdc.h1n1flu/recommendations.htm). Furthermore, in epidemiology, the basic reproduction number, \( R_0 \), is defined as the number of secondary infections caused by a single infective, introduced into a population made up entirely of susceptible individuals, over this individual’s course of the infection. Therefore, \( R_0 \) can be obtained by

\[
R_0 = \frac{\beta \lambda_N}{\gamma^*}.
\]  

Typically, \( R_0 \) has a value between 1 and 2 for new strains of influenza A in human communities (e.g., Anderson, 2006). Therefore, since \( \lambda_N = 1,068,978 \) in the 2000 Census in Franklin County, Ohio, and \( \gamma^* = 0.33 \), if we select \( \beta^* = 5.1 \times 10^{-7} \) as the true value of \( \beta \), then we obtain \( R_0 \approx 1.65 \) to mimic a pandemic flu. We select \( \phi^* = 0.05, \sigma^2_{\xi_S} = \sigma^2_{\xi_I} = 0.1 \), as the true values of \( \phi, \sigma^2_{\xi_S}, \) and \( \sigma^2_{\xi_I} \), respectively, and \( \Sigma^*_\xi \equiv \text{diag}(\sigma^2_{\xi_S}, \sigma^2_{\xi_I}) \). Furthermore, we assume that \( P_{S^*}(1) = 0.99 \) and \( P_{I^*}(1) = 0.001 \), which results in

\[
\mu^*_W(1) = \left( \log \left( \frac{0.99}{0.009} \right), \log \left( \frac{0.001}{0.009} \right) \right)' = (4.7, -2.197)',
\]

the initial mean of the hidden log-odds-ratio vector, \( W(1) \).

We simulate daily data for 45 days. Specifically, for \( t = 1 \), we simulate

\[
W(1) \sim \text{MVN} \left( \mu^*_W(1), \Sigma^*_\xi \right);
\]
for \( t = 2, \ldots, 45 \), we simulate \( \{W(t)\} \) using (3.2.17) and then obtain \( \{P(t)\} \) using transformations defined in (B.1.5) and (B.1.6). Finally, we generate observed counts of susceptible and infectious individuals, \( \{Z_S(t)\} \) and \( \{Z_I(t)\} \), from the Poisson distribution defined in (3.2.1) and (3.2.2) conditional on \( \{P(t)\} \). These counts, \( \{Z_S(t)\} \) and \( \{Z_I(t)\} \), represent the HSIR dataset.

### 3.4.1.2 The ESIR dataset

For simulating the ESIR dataset, we also simulate daily data for 45 days. For the parameters, \( \beta, \gamma, \phi, \sigma^2_{\xi_S}, \) and \( \sigma^2_{\xi_I} \), the mean of the initial vector \( W(1) \), and the total population \( \lambda_N \), we select the same values as in Section 3.4.1.1 for the HSIR dataset. Further, for comparison purpose, we select \( \sigma^2_{P_S} = \sigma^2_{P_I} = 1 \) as the true values for extra variance components, \( \sigma^2_{P_S} \) and \( \sigma^2_{P_I} \), to incorporate sufficient variations into the ESIR dataset, and \( \Sigma^*_P \equiv \text{diag}(\sigma^2_{P_S}, \sigma^2_{P_I}) \). For \( t = 1 \), we simulate

\[
W(1) \sim \text{MVN} \left( \mu^*_W(1), \Sigma^*_\xi \right);
\]

for \( t = 2, \ldots, 45 \), we simulate \( \{W(t)\} \) using equation (3.2.35) and then, conditional on \( \{W(t)\} \), we use the logistic normal distribution defined in (3.2.27) to generate \( \{P(t)\} \). Finally, we generate observed counts of susceptible and infectious individuals, \( \{Z_S(t)\} \) and \( \{Z_I(t)\} \), from the Poisson distributions defined in (3.2.1) and (3.2.2) conditional on \( \{P(t)\} \). These counts, \( \{Z_S(t)\} \) and \( \{Z_I(t)\} \), represent the ESIR dataset.

### 3.4.1.3 The NSIR dataset

Before we illustrate the procedure for generating the NSIR dataset, we first define the Neo-classic SIR (NSIR) model. Recall that the classic SIR model is deterministic; but the NSIR model is a hierarchical statistical model that consists of a data model defined by (3.2.1) and (3.2.2), a \textit{deterministic} process model defined by the classic
SIR model in (3.2.12)-(3.2.14), and a parameter model. That is, for \( t = 1, 2, \ldots \), in units of \( \Delta \) days, an NSIR model can be written as:

**Data model:**

\[
Z_S(t) | \lambda_S(t) \sim \text{ind. Poisson}(\lambda_N P_S(t)),
\]

\[
Z_I(t) | \lambda_I(t) \sim \text{ind. Poisson}(\lambda_N P_I(t)).
\]

**Process model:**

\[
P_S(t+1) = P_S(t) - \beta \Delta \lambda_N P_S(t) P_I(t) + \phi \Delta P_R(t),
\]

\[
P_I(t+1) = P_I(t) - \beta \Delta \lambda_N P_S(t) P_I(t) - \gamma \Delta P_I(t),
\]

\[
P_R(t+1) = P_R(t) + \gamma \Delta P_I(t) - \phi \Delta P_R(t).
\]

**Parameter model:**

\[
[\beta, \gamma, \phi] = [\beta][\gamma][\phi].
\]

In this hierarchical statistical model, recall that \( Z_S(t) \) and \( Z_I(t) \) are the observed susceptible and infectious counts, respectively; \( P_S(t), P_I(t), \) and \( P_R(t) \) are the hidden true proportions; \( \beta, \gamma, \) and \( \phi \) are transmission rate, recovery rate, and loss-of-immunity rate, respectively, per day; and \( \Delta \) is the time step. In our case, \( \Delta = 1 \).

Notice that in the NSIR model, the process model does not capture any uncertainty in the hidden epidemic process.

For simulating the NSIR dataset, we also simulate daily data for 45 days. For the unknown parameters, \( \beta, \gamma, \) and \( \phi \), we select the same values in Section 3.4.1.1, and we select the starting proportions, \( P_S(1) = 0.09 \) and \( P_I(1) = 0.001 \), to be consistent with the other two datasets. Then for \( t = 2, \ldots, 45 \), we simulate \( \{ P(t) \} \) from the deterministic process defined in (3.2.12)-(3.2.14), and we finally generate \( \{ Z_S(t) \} \).
and \( \{Z_I(t)\} \) using the Poisson distribution defined in (3.2.1) and (3.2.2) based on \( \{P(t)\} \). These counts, \( \{Z_S(t)\} \) and \( \{Z_I(t)\} \), represent the \textit{NSIR dataset}.

Notice that the HSIR dataset and ESIR dataset both mimic noisy observations from an epidemic with uncertainties in the underlying epidemic process, while the NSIR dataset mimics noisy observations from a deterministic epidemic process. Although for each dataset, we simulate data for 45 days, when we fit models in Section 3.4.2, we assume that data are only available on the first \( T = 35 \) days and are missing in the last \( F = 10 \) days; that is, \( T + F = 45 \). Thus, we can assess the performance of these models on forecasting by comparing their forecasts to the true values.

Figure 3.1a-3.1f show the daily observed susceptible counts \( \{Z_S(t)\} \) and infectious counts \( \{Z_I(t)\} \), for each of the three datasets, as a function of time (for all 45 days). Notice that we use the pink vertical line on each plot to emphasize that we assume data are only available for the first \( T = 35 \) days. Comparing these plots, we can clearly see that the epidemic patterns in the NSIR dataset (Figure 3.1a-3.1f) are much smoother than those in the other two datasets, as expected. Comparing the epidemic patterns of the HSIR dataset (Figure 3.1a-3.1b), we see that they are smoother than those of the extended HSIR dataset (Figure 3.1c-3.1d), again as expected. In general, the three datasets suggest similar epidemic patterns in the 45 days. That is, all of them indicate clearly an epidemic between day 25 to day 45, and the infectious population reaches its peak in the period between day 33 and day 38.
3.4.2 Fitting the ASIR and Extended ASIR Models

As explained in Section 3.3, to reduce computational complexity, we perform posterior analyses and forecasting based on the ASIR model and the extended ASIR model, which have been well calibrated to the HSIR model and the extended HSIR model, respectively. Recall that within the 45-day study period, data are only available on the first $T = 35$ days and are missing for the last $F = 10$ days.

Notice that our method is not “real-time” Bayesian statistical methodology that updates current and past posteriors as more data arrive, such as the particle filtering
approach in Dukić et al. (2010). If we have new data, we have to re-run the MCMC from the beginning.

3.4.2.1 Fitting the ASIR Model

Based on the ASIR model specified in Section 3.3.1, the joint posterior distribution of all “unknowns” is proportional to a product of the data model, the process model, and the parameter model. Notice that \( \{W_S(t)\} \) and \( \{W_I(t)\} \) are fixed transformations of \( \{P_S(t)\} \) and \( \{P_I(t)\} \) in the ASIR model; thus, combining equations (B.1.5) and (B.1.6) in Appendix B.1, the ASIR’s data model given by (3.2.1) and (3.2.2) can be rewritten: For \( t = 1, ..., T \),

\[
Z_S(t)|P_S(t) = Z_S(t)|W(t) \sim \text{ind. Poisson} \left( \frac{\lambda_N \exp(W_S(t))}{1 + \exp(W_I(t)) + \exp(W_S(t))} \right)
\]

\[
Z_I(t)|P_I(t) = Z_I(t)|W(t) \sim \text{ind. Poisson} \left( \frac{\lambda_N \exp(W_I(t))}{1 + \exp(W_I(t)) + \exp(W_S(t))} \right).
\]

Therefore, the joint posterior distribution of all unknowns can be obtained as follows:

\[
[\beta, \theta_\gamma, \phi, \sigma_{\zeta_S}^2, \sigma_{\zeta_I}^2, \{W_S(t)\}, \{W_I(t)\}|Z_S(1), ..., Z_S(T), Z_I(1), ..., Z_I(T)]
\]

\( \propto \prod_{t=1}^{T} [Z_S(t)|W_S(t), W_I(t)] \cdot \prod_{t=1}^{T} [Z_I(t)|W_S(t), W_I(t)] \cdot [W_S(1), W_I(1)|\sigma_{\zeta_S}^2, \sigma_{\zeta_I}^2]
\]

\[
\cdot \left( \prod_{t=2}^{T+F} [W_S(t), W_I(t)|\beta, \theta_\gamma, \phi, W_S(t-1), W_I(t-1), \sigma_{\zeta_S}^2, \sigma_{\zeta_I}^2] \cdot [\beta|\theta_\gamma][\phi][\sigma_{\zeta_S}^2][\sigma_{\zeta_I}^2] \right).
\]

(3.4.2)

We specify the hidden log-odds-ratio vector, \( W(1) \equiv (W_S(1), W_I(1))' \), as

\[
W(1)|\sigma_{\zeta_S}^2, \sigma_{\zeta_I}^2 \sim \text{MVN} (\mu_W(1), \Sigma_\zeta^2),
\]

where the hyperparameter \( \mu_W(1) \) is specified as

\[
\mu_W(1) = \mu_W^*(1) = \left( \log \left( \frac{0.99}{0.009} \right), \log \left( \frac{0.001}{0.009} \right) \right)' = (4.7, -2.197)'.
\]
(Recall from Section 3.4.1 that $\mu_{W}^*(1)$ is the initial mean of the log-odds-ratio vector used for simulating the data.) The parameter model defined in (3.2.1.3) consists of independent prior distributions on each parameter. Regarding prior-parameter specification, we specify fairly vague priors for $\sigma_{\zeta S}^2$ and $\sigma_{\zeta I}^2$ by choosing independent Inverse Gamma distributions with hyperparameters, $a_{\zeta S} = a_{\zeta I} = 0.25$, $b_{\zeta S} = b_{\zeta I} = 0.4$. Recall from Section 3.2.1.3 that we also specify $\beta \sim \text{Uniform}[0, \beta_{\text{max}}]$, $\theta, \gamma \sim \text{Normal}(0.33, 0.01)$, $\phi \sim \text{Uniform}[0, 1]$.

Notice that it is difficult to obtain a closed-form expression for the joint posterior distribution in (3.4.2), as well as for the marginal posterior distributions of individual parameters; instead, we obtain the posterior through a Markov chain Monte Carlo (MCMC) simulation with a Gibbs sampler that incorporates Metropolis-Hastings steps where necessary (e.g., Waller et al., 1997), based on the full conditional distributions listed below. (We use “rest” to represent all other unknowns as well as the data $\{Z_S(1), ..., Z_S(T)\}$ and $\{Z_I(1), ..., Z_I(T)\}$.)

- $\beta$

$$[\beta|\text{rest}] \propto \prod_{t=2}^{T+F} [W_S(t), W_I(t)|W_S(t-1), W_I(t-1), \beta, \theta, \gamma, \phi, \sigma_{\zeta S}^2, \sigma_{\zeta I}^2][\beta].$$

- $\theta, \gamma$

$$[\theta, \gamma|\text{rest}] \propto \prod_{t=2}^{T+F} [W_S(t), W_I(t)|W_S(t-1), W_I(t-1), \beta, \theta, \gamma, \phi, \sigma_{\zeta S}^2, \sigma_{\zeta I}^2][\theta, \gamma].$$

- $\phi$

$$[\phi|\text{rest}] \propto \prod_{t=2}^{T+F} [W_S(t), W_I(t)|W_S(t-1), W_I(t-1), \beta, \theta, \gamma, \phi, \sigma_{\zeta S}^2, \sigma_{\zeta I}^2][\phi].$$

- $\sigma_{\zeta S}^2$

$$[\sigma_{\zeta S}^2|\text{rest}] \propto [W_S(1), W_I(1)|\sigma_{\zeta S}^2, \sigma_{\zeta I}^2] \\ \cdot \prod_{t=2}^{T+F} [W_S(t), W_I(t)|W_S(t-1), W_I(t-1), \beta, \theta, \gamma, \phi, \sigma_{\zeta S}^2, \sigma_{\zeta I}^2][\sigma_{\zeta S}^2].$$
• $\sigma^2_{\zeta_l}$

$$[\sigma^2_{\zeta_l}|\text{rest}] \propto \left[ W_S(1), W_I(1)|\sigma^2_{\zeta_S}, \sigma^2_{\zeta_l} \right]$$

$$\times \prod_{t=2}^{T+F} [W_S(t), W_I(t)|W_S(t-1), W_I(t-1), \beta, \theta, \phi, \sigma^2_{\zeta_S}, \sigma^2_{\zeta_l}]$$

$$\times [\sigma^2_{\zeta_I}|\text{rest}]$$

• $\{W_S(t), W_I(t)\}$: for $t = 1, ..., T, T+1, ..., T+F$,

$$[W_S(t), W_I(t)|\text{rest}] \propto \begin{cases} 
[Z_S(t)|W_S(t), W_I(t)] [Z_I(t)|W_S(t), W_I(t)] \\
[W_S(t+1), W_I(t+1)|W_S(t), W_I(t), \beta, \theta, \phi, \sigma^2_{\zeta_S}, \sigma^2_{\zeta_I}] \\
[W_S(t), W_I(t)|\sigma^2_{\zeta_S}, \sigma^2_{\zeta_I}]
\end{cases}$$

if $t = 1$;

$$[W_S(t+1), W_I(t+1)|W_S(t), W_I(t), \beta, \theta, \phi, \sigma^2_{\zeta_S}, \sigma^2_{\zeta_I}]$$

$$[W_S(t), W_I(t)|W_S(t-1), W_I(t-1), \beta, \theta, \phi, \sigma^2_{\zeta_S}, \sigma^2_{\zeta_I}]$$

if $2 \leq t \leq T$;

$$[W_S(t+1), W_I(t+1)|W_S(t), W_I(t), \beta, \theta, \phi, \sigma^2_{\zeta_S}, \sigma^2_{\zeta_I}]$$

$$[W_S(t), W_I(t)|W_S(t-1), W_I(t-1), \beta, \theta, \phi, \sigma^2_{\zeta_S}, \sigma^2_{\zeta_I}]$$

if $T < t < T+F$;

$$[W_S(t), W_I(t)|W_S(t-1), W_I(t-1), \beta, \theta, \phi, \sigma^2_{\zeta_S}, \sigma^2_{\zeta_I}]$$

if $t = T+F$.

In the MCMC simulation, we first specify initial values for all unknowns. Then we draw an observation cyclically from each full conditional distribution, conditioning on the most recent values of the other parameters. This iterative procedure allows us to obtain a Markov chain whose stationary distribution is the joint distribution of all the unknowns given the data (i.e., the posterior distribution). Hence, after a “burn-in” number of iterations, we obtain realizations from the posterior distribution. Notice that, in our case, except for $[\sigma^2_{\zeta_S}|\text{rest}]$ and $[\sigma^2_{\zeta_I}|\text{rest}]$, which follow the Inverse Gamma
distributions, and \([W_S(t), W_I(t)|\text{rest}]\) for \(t > T\), which follows a multivariate Normal distribution, the full conditional distributions of all the other unknowns cannot be simulated directly, and so Metropolis-Hastings updates are applied (e.g., Robert and Casella, 2004).

### 3.4.2.2 Fitting the Extended ASIR Model

The procedure of fitting the extended ASIR model is similar to that for fitting the ASIR model in Section 3.4.2.1. Specifically, from the extended ASIR model defined in Section 3.3.2, we can obtain the joint posterior distribution of all unknowns as follows:

\[
[\beta, \theta, \phi, \sigma^2_{\xi_S}, \sigma^2_{\zeta_I}, \sigma^2_{PS}, \sigma^2_{PI}, \{W(t)\}, \{P(t)\}|Z_S(1), ..., Z_S(T), Z_I(1), ..., Z_I(T)]
\]

\[
\propto \prod_{t=1}^{T} [Z_S(t)|P_S(t)] \prod_{t=1}^{T} [Z_I(t)|P_I(t)] \cdot \left( \prod_{t=1}^{T+F} [P_S(t), P_I(t)|W_S(t), W_I(t), \sigma^2_{PS}, \sigma^2_{PI}] \right)
\]

\[
\cdot \left( \prod_{t=2}^{T+F} [W_S(t), W_I(t)|W_S(t-1), W_I(t-1), \beta, \theta, \phi, \sigma^2_{\xi_S}, \sigma^2_{\zeta_I}, \sigma^2_{PS}, \sigma^2_{PI}] \right)
\]

\[
\cdot [W_S(1), W_I(1)|\sigma^2_{\xi_S}, \sigma^2_{\zeta_I}][\beta][\theta, \phi][\sigma^2_{PS}][\sigma^2_{PI}][\sigma^2_{\xi_S}][\sigma^2_{\zeta_I}].
\] (3.4.3)

We specify fairly vague priors for the additional parameters, \(\sigma^2_{PS}\) and \(\sigma^2_{PI}\), by choosing independent Inverse Gamma distributions with hyperparameters, \(a_{PS} = a_{PI} = 0.25\), \(b_{PS} = b_{PI} = 0.4\). The prior distributions for all other parameters, as well as the value of \(W(1)\), are the same as for fitting the ASIR model in Section 3.4.2.1.

The posterior of the extended ASIR model is likewise obtained through a Markov chain Monte Carlo (MCMC) simulation based on a Gibbs sampler with Metropolis-Hastings steps where necessary. The full conditional distributions are listed below. (We use “rest” to represent all other unknowns as well as the data \(\{Z_S(1), ..., Z_S(T)\}\) and \(\{Z_I(1), ..., Z_I(T)\}\).)
• $\beta$

$$[\beta|\text{rest}] \propto \prod_{t=2}^{T+F} [W_S(t), W_I(t)|W_S(t-1), W_I(t-1), \beta, \theta, \phi, \sigma_{\xi_S}^2, \sigma_{\xi_I}^2, \sigma_{P_S}^2, \sigma_{P_I}^2][\beta].$$

• $\theta_\gamma$

$$[\theta_\gamma|\text{rest}] \propto \prod_{t=2}^{T+F} [W_S(t), W_I(t)|W_S(t-1), W_I(t-1), \beta, \theta, \phi, \sigma_{\xi_S}^2, \sigma_{\xi_I}^2, \sigma_{P_S}^2, \sigma_{P_I}^2][\theta_\gamma].$$

• $\phi$

$$[\phi|\text{rest}] \propto \prod_{t=2}^{T+F} [W_S(t), W_I(t)|W_S(t-1), W_I(t-1), \beta, \theta, \phi, \sigma_{\xi_S}^2, \sigma_{\xi_I}^2, \sigma_{P_S}^2, \sigma_{P_I}^2][\phi].$$

• $\sigma_{\xi_S}^2$

$$[\sigma_{\xi_S}^2|\text{rest}] \propto \prod_{t=2}^{T+F} [W_S(t), W_I(t)|W_S(t-1), W_I(t-1), \beta, \theta, \phi, \sigma_{\xi_S}^2, \sigma_{\xi_I}^2, \sigma_{P_S}^2, \sigma_{P_I}^2].$$

• $\sigma_{\xi_I}^2$

$$[\sigma_{\xi_I}^2|\text{rest}] \propto \prod_{t=2}^{T+F} [W_S(t), W_I(t)|W_S(t-1), W_I(t-1), \beta, \theta, \phi, \sigma_{\xi_S}^2, \sigma_{\xi_I}^2, \sigma_{P_S}^2, \sigma_{P_I}^2].$$

• $\{W_S(t), W_I(t)\}$: for $t = 1, ..., T, T + 1, ..., T + F$,

$$[W_S(t), W_I(t)|\text{rest}] \propto \begin{cases} [P_S(t), P_I(t)|W_S(t), W_I(t), \sigma_{P_S}^2, \sigma_{P_I}^2][W_S(t), W_I(t)|\sigma_{\xi_S}^2, \sigma_{\xi_I}^2] \\
|W_S(t+1), W_I(t+1)|W_S(t), W_I(t), \beta, \theta, \phi, \sigma_{\xi_S}^2, \sigma_{\xi_I}^2, \sigma_{P_S}^2, \sigma_{P_I}^2], \\
\text{if } t = 1; \\
[P_S(t), P_I(t)|W_S(t), W_I(t), \sigma_{P_S}^2, \sigma_{P_I}^2] \\
|W_S(t), W_I(t)|W_S(t-1), W_I(t-1), \beta, \theta, \phi, \sigma_{\xi_S}^2, \sigma_{\xi_I}^2, \sigma_{P_S}^2, \sigma_{P_I}^2], \\
|W_S(t+1), W_I(t+1)|W_S(t), W_I(t), \beta, \theta, \phi, \sigma_{\xi_S}^2, \sigma_{\xi_I}^2, \sigma_{P_S}^2, \sigma_{P_I}^2], \\
\text{if } 2 \leq t < T + F; \\
[P_S(t), P_I(t)|W_S(t), W_I(t), \sigma_{P_S}^2, \sigma_{P_I}^2] \\
|W_S(t), W_I(t)|W_S(t-1), W_I(t-1), \beta, \theta, \phi, \sigma_{\xi_S}^2, \sigma_{\xi_I}^2, \sigma_{P_S}^2, \sigma_{P_I}^2], \\
\text{if } t = T + F. 
\end{cases}$$
• $\sigma^2_{PS}$

$$\left[ \sigma^2_{PS} | \text{rest} \right] \propto \left[ \sigma^2_{PS} \right] \cdot \prod_{t=1}^{T+F} \left[ P_S(t), P_I(t) | W_S(t), W_I(t), \sigma^2_{PS}, \sigma^2_{PI} \right]$$

$$\quad \cdot \left( \prod_{t=2}^{T+F} \left[ W_S(t), W_I(t) | W_S(t-1), W_I(t-1), \beta, \theta, \phi, \sigma^2_{\zeta_S}, \sigma^2_{\zeta_I}, \sigma^2_{PS}, \sigma^2_{PI} \right] \right).$$

• $\sigma^2_{PI}$

$$\left[ \sigma^2_{PI} | \text{rest} \right] \propto \left[ \sigma^2_{PI} \right] \cdot \prod_{t=1}^{T+F} \left[ P_S(t), P_I(t) | W_S(t), W_I(t), \sigma^2_{PS}, \sigma^2_{PI} \right]$$

$$\quad \cdot \left( \prod_{t=2}^{T+F} \left[ W_S(t), W_I(t) | W_S(t-1), W_I(t-1), \beta, \theta, \phi, \sigma^2_{\zeta_S}, \sigma^2_{\zeta_I}, \sigma^2_{PS}, \sigma^2_{PI} \right] \right).$$

• $\{P_S(t), P_I(t)\}$: for $t = 1, ..., T, T+1, ..., T+F$,

$$\left[ P_S(t), P_I(t) | \text{rest} \right] \propto \begin{cases} 
[ Z_S(t) | P_S(t) ] [ Z_I(t) | P_I(t) ] [ P_S(t), P_I(t) | W_S(t), W_I(t), \sigma^2_{PS}, \sigma^2_{PI} ], & \text{if } t \leq T; \\
[ P_S(t), P_I(t) | W_S(t), W_I(t), \sigma^2_{PS}, \sigma^2_{PI} ], & \text{if } t > T.
\end{cases}$$

Notice that the simulation procedure based on the extended ASIR model is similar to that of the ASIR model presented in Section 3.4.2.1. The full conditional distributions, $[\sigma^2_{\zeta_S} | \text{rest}]$ and $[\sigma^2_{\zeta_I} | \text{rest}]$, follow the Inverse Gamma distribution; the full conditional distributions, $[W_S(t), W_I(t) | \text{rest}]$, $t = 1, ..., T+F$, follow the Multivariate Normal distribution; the full conditional distribution, $[P_S(t), P_I(t) | \text{rest}]$, follows the Logistic Normal distribution when $t > T$. All these can be sampled through Gibbs updating. The full conditional distributions of all other unknowns cannot be simulated directly, and so Metropolis-Hastings updates are applied (e.g., Robert and Casella, 2004).
3.4.3 Posterior Analysis and Forecasting

3.4.3.1 Comparisons between ASIR and Extended ASIR models based on posterior analysis

For the purpose of comparison, we fit both the ASIR model and the extended ASIR model to each of the three datasets, and we refer to them as:

- case AH: fit the ASIR model to the HSIR dataset
- case AE: fit the ASIR model to ESIR dataset
- case AN: fit the ASIR model to the NSIR dataset
- case EH: fit the extended ASIR model to the HSIR dataset
- case EE: fit the extended ASIR model to the ESIR dataset
- case EN: fit the extended ASIR model to the NSIR dataset

For each of these six cases, we ran one MCMC chain of 100,000 iterations. After a burn-in of 3,000 iterations, we obtained a total of 97,000 realizations from the posterior distribution for each case listed above.

Figure 3.2 shows the convergence behavior of $\beta$, $\gamma$, $\sigma^2_{\zeta_S}$, $\sigma^2_{\zeta_I}$, and $\phi$, for the case AH. Similar figures are obtained for the other five cases mentioned above (not shown). The posterior median, mean, variance, and 95% Bayes credible interval (95%CI) for parameters in each of the six cases are shown in Table 3.2-3.7. For all six cases, we can see that, except for $\gamma$ (recall $\theta_\gamma$ in the tables is the logit transformation of $\gamma$), the posteriors for all the other parameters are much tighter than the priors. Hence, despite the missing data during the last 10 days, there is substantial learning about these parameters. Furthermore, because the data are simulated, we have the opportunity to compare the posterior realizations with the true values. As mentioned in Section
3.4.1, for all three datasets, we use $\beta^* = 5.1 \times 10^{-7}$, $\gamma^* = 0.33$, and $\phi^* = 0.05$, as the true values of the transmission rate, the recovery rate, and the loss-of-immunity rate, respectively (in units of per day). Tables 3.2-3.7 show that both ASIR and extended ASIR models perform well on recovering these parameters, and the posterior 95% CIs of these three parameters are much tighter than their priors (except for $\gamma$, which already has a tight prior). Figure 3.2 supports this conclusion further by indicating the true values of the three parameters as red lines in Figures 3.2a, 3.2b, and 3.2e for the case AH. We find the agreements are excellent, especially for $\beta$ and $\gamma$. The same conclusions hold for the other five cases (not shown).

Recall that the HSIR and ESIR datasets have uncertainties associated with the hidden epidemic process, and we selected $\sigma^2_{\xi_S} = \sigma^2_{\xi_I} = 0.1$ as the true values of the small-scale variance components. Tables 3.2, 3.3, 3.5, and 3.6 indicate that when fitting to HSIR and ESIR datasets (case AH, AE, EH, EE, respectively), the posterior realizations of small-scale variances in the ASIR and extended ASIR models (i.e., $\sigma^2_{\xi_S}$ and $\sigma^2_{\xi_I}$) tend to be larger than 0.1. For example, for AH (Table 3.2), the posterior medians of $\sigma^2_{\xi_S}$ and $\sigma^2_{\xi_I}$ are about 0.4. These results agree with our expectation, because, as explained in Section 3.3.1, the small-scale variation terms in the ASIR and extended ASIR models also capture the higher-order terms in the linear approximations.

Recall that the underlying process of the ESIR dataset has extra variation coming from the transformation (from the $\mathbf{P}$-scale to the $\mathbf{W}$-scale), and we selected the extra variance components as $\sigma^2_{P_S} = \sigma^2_{P_I} = 1$. The extended ASIR model has an extra component to capture this; Table 3.6 shows the case EE, where the posteriors of $\sigma^2_{P_S}$ and $\sigma^2_{P_I}$ are much tighter than the priors.
We now compare the performances of fitting the ASIR model and the extended ASIR model to the three datasets. Figure 3.3 shows the prior and posterior distributions of $\beta$ (as a representative parameter) for all six cases. Clearly, fitting the ASIR model yields much tighter posteriors of $\beta$ centered at the true value, for both the HSIR and NSIR datasets. For the ESIR dataset that consists of the most sources of uncertainties, the posteriors of $\beta$ for both the ASIR and extended ASIR models are similar and more skewed. Results in Tables 3.2-3.7 support these conclusions further.

Figure 3.2: Trace plots for the case AH: (a) trace plot for $\beta$; (b) trace plot for $\gamma$; (c) trace plot for $\sigma^2_{\xi_S}$; (d) trace plot for $\sigma^2_{\xi_I}$; (e) trace plot for $\phi$. Red lines in plots (a)-(e) indicate the true values of the parameters used for simulating the HSIR dataset, namely, $\beta^*$, $\gamma^*$, $\sigma^2_{\xi_S}$, $\sigma^2_{\xi_I}$, and $\phi^*$, respectively.
Figure 3.3: Prior (red line) and Posterior (histogram) distributions of $\beta$ for (a) case AH; (b) case EH; (c) case AE; (d) case EE; (e) case AN; (f) case EN. In each plot, blue dashed lines indicate posterior 95% CIs; the green line indicates the true value.

Table 3.2: Table of parameters and summaries of their priors and posteriors for case AH.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior distribution</th>
<th>Prior quantiles</th>
<th>Posterior quantiles</th>
<th>Posterior</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>Uniform(0, $\beta_{\text{max}} = 6.7686e-6$)</td>
<td>$0.025$ 0.5 0.975</td>
<td>$0.025$ 0.5 0.975</td>
<td>mean variance</td>
</tr>
<tr>
<td>$\theta_\gamma$</td>
<td>Gaussian(-0.708, 0.01)</td>
<td>-0.904 -0.708 -0.512</td>
<td>-0.8848 -0.6922 -0.4954</td>
<td>-0.6914 0.0099</td>
</tr>
<tr>
<td>$\phi$</td>
<td>Uniform(0,1)</td>
<td>0.025 0.5 0.975</td>
<td>0.0051 0.0993 0.2990</td>
<td>0.1128 0.0064</td>
</tr>
<tr>
<td>$\sigma_{\xi S}^2$</td>
<td>Inverse Gamma(0.25, 0.4)</td>
<td>1.457 57.2 9.48e6</td>
<td>0.2345 0.4066 0.7829</td>
<td>0.4324 0.0202</td>
</tr>
<tr>
<td>$\sigma_{\xi I}^2$</td>
<td>Inverse Gamma(0.25, 0.4)</td>
<td>1.457 57.2 9.48e6</td>
<td>0.2332 0.4055 0.7774</td>
<td>0.4314 0.0201</td>
</tr>
</tbody>
</table>

76
Table 3.3: Table of parameters and summaries of their priors and posteriors for case AE.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior distribution</th>
<th>Prior quantiles</th>
<th>Posterior quantiles</th>
<th>Posterior mean</th>
<th>variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>Uniform(0, $\beta_{max} = 4.1212e - 6$)</td>
<td>1.030e-7</td>
<td>2.061e-6</td>
<td>4.018e-6</td>
<td>2.4e-7</td>
</tr>
<tr>
<td>$\theta$</td>
<td>Gaussian(-0.708, 0.01)</td>
<td>-0.904</td>
<td>-0.708</td>
<td>-0.512</td>
<td>-0.8840</td>
</tr>
<tr>
<td>$\phi$</td>
<td>Uniform(0,1)</td>
<td>0.025</td>
<td>0.5</td>
<td>0.975</td>
<td>0.0154</td>
</tr>
<tr>
<td>$\sigma^2_{\xi S}$</td>
<td>Inverse Gamma(0.25, 0.4)</td>
<td>1.457</td>
<td>57.2</td>
<td>9.4866</td>
<td>1.7488</td>
</tr>
<tr>
<td>$\sigma^2_{\xi I}$</td>
<td>Inverse Gamma(0.25, 0.4)</td>
<td>1.457</td>
<td>57.2</td>
<td>9.4866</td>
<td>1.4718</td>
</tr>
</tbody>
</table>

Table 3.4: Table of parameters and summaries of their priors and posteriors for case AN.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior distribution</th>
<th>Prior quantiles</th>
<th>Posterior quantiles</th>
<th>Posterior mean</th>
<th>variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>Uniform(0, $\beta_{max} = 9.7523e - 6$)</td>
<td>2.438e-7</td>
<td>4.876e-6</td>
<td>9.508e-6</td>
<td>2.374e-7</td>
</tr>
<tr>
<td>$\theta$</td>
<td>Gaussian(-0.708, 0.01)</td>
<td>-0.904</td>
<td>-0.708</td>
<td>-0.512</td>
<td>-0.8904</td>
</tr>
<tr>
<td>$\phi$</td>
<td>Uniform(0,1)</td>
<td>0.025</td>
<td>0.5</td>
<td>0.975</td>
<td>0.0052</td>
</tr>
<tr>
<td>$\sigma^2_{\xi S}$</td>
<td>Inverse Gamma(0.25, 0.4)</td>
<td>1.457</td>
<td>57.2</td>
<td>9.4866</td>
<td>0.1321</td>
</tr>
<tr>
<td>$\sigma^2_{\xi I}$</td>
<td>Inverse Gamma(0.25, 0.4)</td>
<td>1.457</td>
<td>57.2</td>
<td>9.4866</td>
<td>0.1327</td>
</tr>
</tbody>
</table>

Table 3.5: Table of parameters and summaries of their priors and posteriors for case EH.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior distribution</th>
<th>Prior quantiles</th>
<th>Posterior quantiles</th>
<th>Posterior mean</th>
<th>variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>Uniform(0, $\beta_{max} = 6.7686e - 6$)</td>
<td>1.692e-7</td>
<td>3.384e-6</td>
<td>6.599e-6</td>
<td>6.53e-8</td>
</tr>
<tr>
<td>$\theta$</td>
<td>Gaussian(-0.708, 0.01)</td>
<td>-0.904</td>
<td>-0.708</td>
<td>-0.512</td>
<td>-0.8877</td>
</tr>
<tr>
<td>$\phi$</td>
<td>Uniform(0,1)</td>
<td>0.025</td>
<td>0.5</td>
<td>0.975</td>
<td>0.0117</td>
</tr>
<tr>
<td>$\sigma^2_{\xi S}$</td>
<td>Inverse Gamma(0.25, 0.4)</td>
<td>1.457</td>
<td>57.2</td>
<td>9.4866</td>
<td>0.4548</td>
</tr>
<tr>
<td>$\sigma^2_{\xi I}$</td>
<td>Inverse Gamma(0.25, 0.4)</td>
<td>1.457</td>
<td>57.2</td>
<td>9.4866</td>
<td>0.4359</td>
</tr>
<tr>
<td>$\sigma^2_{PS}$</td>
<td>Inverse Gamma(0.25, 0.4)</td>
<td>1.457</td>
<td>57.2</td>
<td>9.4866</td>
<td>0.3761</td>
</tr>
<tr>
<td>$\sigma^2_{PI}$</td>
<td>Inverse Gamma(0.25, 0.4)</td>
<td>1.457</td>
<td>57.2</td>
<td>9.4866</td>
<td>0.3861</td>
</tr>
</tbody>
</table>
To compare the goodness of fit of the ASIR and extended ASIR models, we can compute the deviance information criterion (DIC) (e.g., Claeskens and Hjort, 2008, Section 3.5) for the six cases.

According to Claeskens and Hjort (2008),

\[
\text{DIC} = \sum_{k=1}^{m} -2\log \left[ \prod_{t=1}^{T} L \left( \mathbf{Z}(t)|\lambda^{(k)}(t) \right) \right] - 2\log \left[ \prod_{t=1}^{T} L \left( \mathbf{Z}(t)|\bar{\lambda}(t) \right) \right] + \sum_{k=1}^{m} -2\log \left[ \prod_{t=1}^{T} L \left( \mathbf{Z}(t)|\lambda^{(k)}(t) \right) \right],
\]

(3.4.4)
where \( k \) denotes the \( k \)th iteration in the MCMC, and \( m \) is the total number of iterations after burn-in; in our case, \( m = 97,000 \). In (3.4.4), recall that \( Z(t) = (Z_S(t), Z_I(t))' \) is the vector of (simulated) observed counts at time \( t \), for \( t = 1, ..., T = 35 \); \( \lambda^{(k)}(t) \equiv \left( \lambda^{(k)}_S(t), \lambda^{(k)}_I(t) \right)' \) denotes the vector of true counts at time \( t \) at the \( k \)th iteration, where

\[
\lambda^{(k)}_S(t) = \lambda_N P^{(k)}_S(t) \tag{3.4.5}
\]

\[
\lambda^{(k)}_I(t) = \lambda_N P^{(k)}_I(t); \tag{3.4.6}
\]

and \( \bar{\lambda}(t) \equiv (\bar{\lambda}_S(t), \bar{\lambda}_I(t))' \) denotes the vector of the posterior means of the true counts at time \( t \). We define \( L(\cdot) \) to be the likelihood function; therefore, in (3.4.4),

\[
\prod_{t=1}^T L\left(Z(t)|\lambda^{(k)}(t)\right) = \prod_{t=1}^T \frac{\lambda^{(k)}_S(t) Z_S(t) e^{-\lambda^{(k)}_S(t)}}{Z_S(t)!} \frac{\lambda^{(k)}_I(t) Z_I(t) e^{-\lambda^{(k)}_I(t)}}{Z_I(t)!}, \quad k = 1, ..., m;
\]

\[
\prod_{t=1}^T L(Z(t)|\bar{\lambda}(t)) = \prod_{t=1}^T \frac{\bar{\lambda}_S(t) Z_S(t) e^{-\bar{\lambda}_S(t)}}{Z_S(t)!} \frac{\bar{\lambda}_I(t) Z_I(t) e^{-\bar{\lambda}_I(t)}}{Z_I(t)!}.
\]

Table 3.8: DICs for fitting the ASIR model and the extended ASIR model to HSIR, ESIR, and NSIR datasets.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>DIC of ASIR</th>
<th>DIC of extended ASIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSIR</td>
<td>1063.0</td>
<td>1065.1</td>
</tr>
<tr>
<td>ESIR</td>
<td>1054.8</td>
<td>1053.4</td>
</tr>
<tr>
<td>NSIR</td>
<td>1091.1</td>
<td>1094.2</td>
</tr>
</tbody>
</table>

Table 3.8 summarizes the DICs for fitting the ASIR model and the extended ASIR model to each of the three datasets. Since small DIC indicates a better fit, Table 3.8
shows that the ASIR model fits better than the extended ASIR model to both the HSIR and NSIR datasets, while the extended ASIR model fits slightly better than the ASIR model to the ESIR dataset, as expected. This result is consistent with the previous results shown in Tables 3.2-3.7 and Figure 3.3. Overall, the simpler ASIR model fits well to all three datasets.

To assess the general performances of the ASIR model and the extended ASIR model to approximating the HSIR model and the extended HSIR model, respectively, we use a discrepancy measure, as described in Cressie and Wikle (2011), Section 2.2.2 and due to Gelman et al. (1996). They point out that, based on a given discrepancy measure, one can obtain a posterior predictive p-value and perform a diagnostic procedure to determine whether the model fits the data.

Specifically, for the $k$th iteration from the MCMC, $k = 1, \ldots, m$, we define a discrepancy measure, $\psi \left( \mathbf{Z}; \lambda^{(k)} \right)$, as follows:

$$
\psi \left( \mathbf{Z}; \lambda^{(k)} \right) \equiv \sum_{t=1}^{T} \left[ \mathbf{Z}(t) - \mathbb{E} \left( \mathbf{Z}(t) | \lambda^{(k)}(t) \right) \right]' \left( \Sigma_Z^{(k)}(t) \right)^{-1} \left[ \mathbf{Z}(t) - \mathbb{E} \left( \mathbf{Z}(t) | \lambda^{(k)}(t) \right) \right],
$$

where recall that we assume data are only available on the first $T = 35$ days for each of the three datasets. Based on the Poisson data model defined in (3.2.1) and (3.2.2) and on the definitions of $\lambda^{(k)}(t)$ in (3.4.5) and (3.4.6), the covariance matrix can be defined as

$$
\left( \Sigma_Z^{(k)}(t) \right) \equiv \text{diag} \left( \lambda_S^{(k)}(t), \lambda_I^{(k)}(t) \right).
$$

Furthermore, for $t = 1, \ldots, T$, if we use $\mathbf{Z}_{rep}(t)$ to denote an independent replicate of the data, then the posterior predictive distribution of $\mathbf{Z}_{rep}(t)$ can be defined as (Gelman et al., 1996),

$$
[\mathbf{Z}_{rep}(t) | \mathbf{Z}(t)] = \int \int [\mathbf{Z}_{rep}(t) | \mathbf{U}(t), \Theta] [\mathbf{U}(t), \Theta | \mathbf{Z}(t)] d\mathbf{U}(t) d\Theta(t),
$$

(3.4.8)
where for ASIR model,
\[ U(t) \equiv \{ W(t) \} \]
\[ \Theta \equiv \{ \beta, \gamma, \phi, \Sigma^2 \} \]

and for extended ASIR model,
\[ U(t) \equiv \{ P(t), W(t) \} \]
\[ \Theta \equiv \{ \beta, \gamma, \phi, \Sigma^2, \Sigma^2_P \} \].

Thus, for the \( k \)th sample, \( Z^{(k)}(t) \), which is drawn from the posterior distribution \( [Z_{rep}(t)|Z(t)], \ t = 1, \ldots, T \), we obtain the discrepancy measure \( \psi(Z^{(k)}_{rep}; \lambda^{(k)}) \) by replacing \( Z \) in (3.4.7) with \( Z^{(k)}_{rep}(t) \) defined in (3.4.8). Therefore, the replicates,
\[ \{ Z^{(1)}_{rep}(t), Z^{(2)}_{rep}(t), \ldots, Z^{(m)}_{rep} \}, \]
should “look like” the data \( Z(t) \) if the model is appropriate. Based on this idea, we can apply posterior predictive diagnostics on the ASIR and extended ASIR models to see if they are “appropriate” for fitting the three datasets. Furthermore, since the HSIR and ESIR datasets are simulated from the HSIR and extended HSIR models, respectively, we can also assess the performances of the ASIR model and the extended ASIR model to approximate the HSIR model and the extended HSIR model, respectively.

The posterior predictive p-value can be obtained as below (e.g., Cressie and Wikle, 2011):
\[
\Pr (\psi(Z_{rep}; \lambda) \geq \psi(Z; \lambda)) = \frac{1}{m} \sum_{k=1}^{m} I \left[ \psi \left( Z^{(k)}_{rep}; \lambda^{(k)} \right) \geq \psi \left( Z; \lambda^{(k)} \right) \right], \quad (3.4.9)
\]
where \( I(\cdot) \) denotes an indicator function.

Based on (3.4.9), we obtain the posterior predictive p-values for the case AH, to be 0.49; hence, there is no evidence for a lack of model fit when fitting the ASIR
model to the HSIR dataset. This conclusion is further supported by Figure 3.4, which plots $\psi\left(Z^{(k)}_{\text{rep}}; \lambda^{(k)}\right)$ against $\psi\left(Z; \lambda^{(k)}\right)$ for the case AH; because all the points on the plots are randomly scattered around the 45-degree line, the replicates “look like” the data. The same is true for the other five cases. Thus, we can conclude that there is no evidence for lack of fit for both the ASIR and the extended ASIR models when fitting to these three datasets; consequently, the ASIR and extended ASIR models appear to approximate well the HSIR and extended HSIR models, respectively.

Figure 3.4: Scatter plot of $\psi\left(Z^{(k)}_{\text{rep}}; \lambda^{(k)}\right)$ against $\psi\left(Z; \lambda^{(k)}\right)$ for the case AH
3.4.3.2 Comparisons of the ASIR, extended ASIR, and classic SIR models based on posterior analysis and forecasting

We now fit the discrete-time classic SIR model to each of the three simulated datasets described in Section 3.4.1. In practice, estimates of $\beta$, $\gamma$, and $\phi$, denoted as $\hat{\beta}$, $\hat{\gamma}$, and $\hat{\phi}$, are needed in order to solve for $\{S(t)\}$ and $\{I(t)\}$ in the discretized version of (3.1.1)-(3.1.3). Here, we minimize the sum of squares between the estimated and observed infectious counts over time (e.g., Anderson and May, 1991; Wearing et al., 2005; Burr and Chowell, 2006). That is,

$$\sum_{t=1}^{T} \left( \hat{I}(t; \hat{\beta}, \hat{\gamma}, \hat{\phi}) - Z_I(t) \right)^2 = \min_{\beta, \gamma, \phi} \left[ \sum_{t=1}^{T} \left( \hat{I}(t; \beta, \gamma, \phi) - Z_I(t) \right)^2 \right],$$

(3.4.10)

where $\hat{I}(t; \beta, \gamma, \phi)$ is the deterministic estimate of infectious counts at time $t$, obtained after substituting $\beta$, $\gamma$, and $\phi$ into equations (3.1.1)-(3.1.3). Recall that $T = 35$. Notice that, in our case, the ordinary least squares estimates, $\hat{\beta}$, $\hat{\gamma}$, and $\hat{\phi}$, are also the values selected in Section 3.3 for the initial MCMC values (denoted $\beta_0$, $\gamma_0$, and $\phi_0$, there.) Moreover, the classic SIR forecasts are, using obvious notation, $\{\hat{S}(t; \hat{\beta}, \hat{\gamma}, \hat{\phi}) : t = T + 1, \ldots, T + F\}$, $\{\hat{I}(t; \hat{\beta}, \hat{\gamma}, \hat{\phi}) : t = T + 1, \ldots, T + F\}$, and $\{\hat{R}(t; \hat{\beta}, \hat{\gamma}, \hat{\phi}) : t = T + 1, \ldots, T + F\}$.

For the ASIR and extended ASIR models, the appropriate predictions are obtained from posterior inference on $\lambda_S(t)$, and $\lambda_I(t)$, for the entire study period $t = 1, \ldots, T, T + 1, \ldots, T + F$. Here, equations (3.2.4)-(3.2.6) are used, along with mass balance given by (3.2.8).

In the infectious-disease setting, the number of infectious individuals is one of the quantities that is of most interest. The estimated trajectories given in Figure 3.5 show infectious counts for the entire 45-day study period, obtained by fitting the ASIR model, the extended ASIR model, and the classic SIR model to each of the three datasets. These are compared to the true (hidden) values of infectious counts. Upon inspection of Figure 3.5c (NSIR dataset), we see that all three models can
capture the overall epidemic pattern very well, not only on the first 35 days when
data are available, but also on the last 10 days when there are no data. However,
when the underlying epidemic process has stochastic components, the disadvantage
of the classic SIR model becomes apparent. Upon inspection of Figure 3.5a (HSIR
dataset), estimates from fitting the classic SIR model are oversmoothed, even on the
days when data are available. On the contrary, the agreement between the true values
and the posterior medians of \{\lambda_I(t)\}, obtained from fitting both the ASIR model and
the extended ASIR model, are excellent when data are available; and in the 10 days
when there are no data, they capture the general pattern reasonably well.

Upon inspection of Figure 3.5b (ESIR dataset), we see again the advantages of
fitting the ASIR model and the extended ASIR model over the classic SIR model. We
can see that estimates from the hierarchical models capture the epidemic pattern very
well on the first 35 days when data are available, however estimates from the classic
SIR are oversmoothed. On the last 10 days when there are no data, the hierarchical
models are able to predict the general downward trend at the end of the epidemic
process; however, the classic SIR model is unable to deal with the uncertainties,
mistakenly forecasting that the epidemic is still going strong in the last 10 days.

Another disadvantage of the classic SIR model is that it is unable to provide
any uncertainty measures to accompany its deterministic modeling strategy. On the
contrary, when fitting hierarchical models, we can obtain uncertainty measures for
any quantity of interest, based on the posterior distribution. For example, Figure 3.6
shows 0.025, 0.25, 0.5, 0.75, 0.975 quantiles of posterior distributions of the hidden
infectious counts, \lambda_I(t), during the forecasting period \{T + 1, ..., T + F\}. We can
see that in all cases, the posterior 50\%CI of \lambda_I(t) obtained from fitting both the
ASIR model and the extended ASIR model, cover the true values most of the time
(except day 38 for the cases AE and EE). Comparing Figures 3.6a-3.6f, we notice that
when fitting to the NSIR and HSIR datasets, the ASIR model can provide slightly better predictions with tighter posterior 50% CIs (green dotted lines) and 95% CIs (blue dashed lines) than can the extended ASIR model. This conclusion is consistent with previous discussions that the advantage of using an extended ASIR model is not obvious when fitting to smoother datasets.

Figure 3.5: Estimated infectious counts obtained from fitting the ASIR model, the extended ASIR model, and the classic SIR model to the HSIR, ESIR, and NSIR datasets, as a function of time. The vertical pink line in each plot emphasizes that data are only available on days 1-35, and that they are missing on days 36-45. The blue and green lines are posterior medians of \( \{\lambda_I(t)\} \) obtained from the ASIR model and the extended ASIR model, respectively. The black line is the estimate of the infectious counts from the classic SIR model. The pink stars give the true infectious counts.
Figure 3.6: Posterior quantities of $\{\lambda_I(t)\}$ obtained from the ASIR model and the extended ASIR model during the forecasting period day from 36 to day 45 (based on data from day 1 to day 35, from the respective datasets). Lines indicate the 0.025, 0.25, 0.5, 0.75 and 0.975 quantiles, while pink stars give the true values of $\{\lambda_I(t)\}$.

Similarly, for susceptible counts, Figure 3.7 and Figure 3.8 clearly indicate that the hierarchical models perform better than the classic SIR model, when fitting to the HSIR and ESIR datasets. Furthermore, the figures show that the extended ASIR model can perform much better than the other two models in forecasting the susceptible counts in the last 10 days (for the HSIR and ESIR datasets).
Figure 3.7: Estimated susceptible counts obtained from fitting the ASIR model, the extended ASIR model, and the classic SIR model to the HSIR, ESIR, and NSIR datasets, as a function of time. The pink vertical line in each plot emphasizes that data are only available on days 1-35, and that they are missing on days 36-45. The blue and green lines are posterior medians of \( \{ \lambda_S(t) \} \) obtained from the ASIR model and the extended ASIR model, respectively. The black line is the estimate of the susceptible counts from the classic SIR model. The pink stars give the true susceptible counts.
Figure 3.8: Posterior quantities of \( \{ \lambda_S(t) \} \) obtained from the ASIR model and the extended ASIR model during the forecast period from day 36 to day 45 (based on data from day 1 to day 35 of the respective datasets). Lines indicate the 0.025, 0.25, 0.5, 0.75 and 0.975 quantiles, while pink stars give the true values of \( \{ \lambda_S(t) \} \).

As well as the infectious counts, another quantity of great interest to epidemiologists is the basic reproduction number, \( R_0 \), as defined in (3.4.1):

\[
R_0 = \frac{\beta \lambda_N}{\gamma}.
\]

Notice that this is a threshold quantifying the transmission potential of a disease, namely, whether an epidemic occurs or the disease simply dies out (e.g., Anderson and May, 1991; Dietz, 1993). If \( R_0 < 1 \), the infection dies out; if \( R_0 > 1 \), there is
a pandemic; in cases where $R_0 = 1$, the disease becomes endemic (i.e. the disease remains in the population at a consistent rate).

For the classic SIR model, an estimate of $R_0$ can be obtained based on $\hat{\beta}$ and $\hat{\gamma}$ obtained from (3.4.10). For example, the estimates of $R_0$ obtained from fitting the classic SIR model to each of the three datasets are indicated by black lines on the plots of Figure 3.9. For the ASIR model and the extended ASIR model, not only can we obtain an estimate of $R_0$ (e.g., posterior median), but we can also obtain its posterior distribution. The boxplots in Figure 3.9 are obtained from the posterior distribution of $R_0$ after fitting the hierarchical models to each of the three datasets. Since the data are simulated, we have the opportunity to compare our estimates of $R_0$ to the true values (indicated by green lines in the plots of Figure 3.9). Comparing Figures 3.9a-3.9c, we can clearly see that the classic SIR model generally gives an estimate of $R_0$ that is biased. Furthermore, Figure 3.9 shows the uncertainty in inference based on the hierarchical models. For example, Figure 3.9 shows that the posterior 50% CIs cover the true value of $R_0$ in all data cases.

Figure 3.9 also suggests that the ASIR model can provide better estimates of $R_0$ with tighter 50% CIs than that of the extended ASIR model for the HSIR and NSIR datasets, while the latter obtains better estimates of $R_0$ than the former based on the ESIR dataset. This result is consistent with previous conclusions that when fitting to a smoother dataset, the simpler hierarchical structure of the ASIR model may be advantageous, but when fitting to a more variable dataset, the extended ASIR model is better equipped to handle the complexity.
3.5 Discussion and Conclusions

In this chapter, we develop a Bayesian hierarchical SIR (HSIR) model (and an extended version) that captures the uncertainties in modeling infectious diseases like seasonal or pandemic influenza. The HSIR model preserves mass balance on the (hidden) true counts rather than on the observed counts, as it should. Furthermore, by modeling the dynamical process on a log-odds-ratios scale, our approach captures the stochastic and discrete nature of the epidemic process, as well as keeping the SIRS
dynamics that underlie the classic SIR model. The extended HSIR model referred to above captures extra uncertainties due to transformation to the log-odds-ratio scale.

We simulated three datasets, HSIR, ESIR, and NSIR datasets, from HSIR, extended HSIR, and Neo-classic SIR models, respectively. For each of the three datasets, we assume data are available on the first 35 days and missing on the following 10 days. Then we use MCMC to fit the HSIR and the extended HSIR models to each of these datasets; for computational efficiency, a well calibrated linear approximation is used.

From the MCMC results in Section 3.4, we can see that the calibrated linear versions of the HSIR and extended HSIR models are very good approximations, and that careful accounting of uncertainty leads to superior performance over the classic deterministic SIR model. We also see that the benefit of extending the HSIR model to account for uncertainty in the log-odds-ratio transformation is not all that great. We conclude from our limited study that the calibrated linear version of the HSIR model offers an accurate and computationally efficient approach to analyzing infectious-disease data.

On the other hand, since model fitting in this Chapter is based on only a single dataset simulated from each of the HSIR, ESIR, and NSIR models, our conclusions are only indicative. In the next step of our research, a carefully designed simulation experiment at various levels of various factors with sufficient replication will be carried out. This will allow us to make definitive comparisons between the HSIR model and the classic SIR model.

In further research, we are also investigating more complicated SIRS dynamics that involve a heterogeneous population. For example, we are looking at incorporating susceptible and infectious populations that are stratified according to individuals’ drug resistance. It is also of interest to incorporate birth, death, and immigration...
processes for appropriate time periods. Finally, incorporation of the spatial aspect into these hierarchical dynamical models, would represent a major advance. For a non-hierarchical approach, see Viboud et al. (2006).
Appendix A

DETAILS OF DERIVING THE PRIOR DISTRIBUTION
ON THE EXTRA-POISSON-VARIATION COMPONENT
OF CHAPTER 2

A.1

Based on the Poisson hierarchical model defined in (2.3.1) and (2.3.2) in Section 2.3, it is straightforward to obtain the mean and variance of the data conditional on the relative risk. For \( i = 1, \ldots, N \) and \( t = 1, \ldots, T \),

\[
E(Z_t(s_i)|\lambda_t(s_i)) = \text{var}(Z_t(s_i)|\lambda_t(s_i)) \equiv E_t(s_i)\lambda_t(s_i), \tag{A.1.1}
\]

where recall that \( E_t(s_i) \) is given by (2.2.1) in Section 2.2.1. Hence, from (A.1.1), we further obtain

\[
E\left(\frac{Z_t(s_i)}{E_t(s_i)} \mid \lambda_t(s_i)\right) = \lambda_t(s_i), \tag{A.1.2}
\]

and

\[
\text{var}\left(\frac{Z_t(s_i)}{E_t(s_i)} \mid \lambda_t(s_i)\right) = \frac{\lambda_t(s_i)}{E_t(s_i)}. \tag{A.1.3}
\]

Recall that the Freeman-Tukey transformation defined in (2.2.2) in Section 2.2.1 can be written as a function of \( \frac{Z_t(s_i)}{E_t(s_i)} \), namely,

\[
\text{FTSMR}_t(s_i) = \sqrt{1000} \frac{Z_t(s_i)}{E_t(s_i)} + \sqrt{1000} \frac{Z_t(s_i)}{E_t(s_i)} + \frac{1}{E_t(s_i)}. \tag{A.1.4}
\]
Thus, for $\frac{1}{E_t(s_i)}$ small, we can obtain the approximate (conditional) mean and variance of the Freeman-Tukey transformation by the delta method:

$$E(\text{FTSMR}_t(s_i)|\lambda_t(s_i)) \approx 2\sqrt{1000}\sqrt{\lambda_t(s_i)};$$  \hspace{1cm} (A.1.5)

and

$$\text{var}(\text{FTSMR}_t(s_i)|\lambda_t(s_i)) \approx \frac{1000}{E_t(s_i)} \propto \frac{1}{E_t(s_i)}.$$  \hspace{1cm} (A.1.6)

From (A.1.6), we can see that the Freeman-Tukey transformation is a variance-stabilizing transformation for the Poisson distribution; see Freeman and Tukey (1950) for more details.

Now, if we expand $\sqrt{\lambda_t(s_i)}$ and $\log(\lambda_t(s_i))$ about 1 using Taylor series, we can obtain

$$\sqrt{\lambda_t(s_i)} \approx 1 + \frac{1}{2}(\lambda_t(s_i) - 1) + O(\lambda_t(s_i)^2);$$  \hspace{1cm} (A.1.7)

$$\log(\lambda_t(s_i)) \approx (\lambda_t(s_i) - 1) + O(\lambda_t(s_i)^2),$$  \hspace{1cm} (A.1.8)

for $i = 1, ..., N$, $t = 1, ..., T$. Therefore, from (A.1.7) and (A.1.8),

$$\log(\lambda_t(s_i)) \approx 2\sqrt{\lambda_t(s_i)} - 2,$$  \hspace{1cm} (A.1.9)

which implies that

$$\sqrt{\lambda_t(s_i)} \approx 1 + \frac{1}{2}\log(\lambda_t(s_i)).$$  \hspace{1cm} (A.1.10)

Hence, we have

$$E(\text{FTSMR}_t(s_i)|\lambda_t(s_i)) \approx J + \frac{J}{2}\log(\lambda_t(s_i)),$$  \hspace{1cm} (A.1.11)

where $J = 2\sqrt{1000}$. 

94
Combining (2.3.2) in Section 2.3.2 and (A.1.11) in Appendix A.1, for $i = 1, ..., N$, $t = 1, ..., T$, we have

$$E(FTSMR_t(s_i)|\lambda_t(s_i)) \approx \left(J + \frac{J}{2} \mu_t(s_i)\right) + \frac{J}{2} \eta_t(s_i) + \frac{J}{2} \xi_t(s_i),$$  \hspace{0.5cm} (A.2.1)

where recall that $J = 2\sqrt{1000}$. Furthermore,

$$FTSMR_t(s_i) = E(FTSMR_t(s_i)|\lambda_t(s_i)) + \nu_t(s_i); \hspace{0.5cm} i = 1, ..., N, \hspace{0.5cm} t = 1, ..., T,$$  \hspace{0.5cm} (A.2.2)

where $\nu_t(s_i)$ represents the error after accounting for the epidemiological variation in the (conditional) mean. Hence, from (A.2.1) and (A.2.2),

$$FTSMR_t(s_i) = \mu^{FT}_t(s_i) + \eta^{FT}_t(s_i) + \varepsilon^{FT}_t(s_i),$$  \hspace{0.5cm} (A.2.3)

where

$$\mu^{FT}_t(s_i) \equiv J + \frac{J}{2} \mu_t(s_i),$$  \hspace{0.5cm} (A.2.4)

$$\eta^{FT}_t(s_i) \equiv \frac{J}{2} \eta_t(s_i),$$  \hspace{0.5cm} (A.2.5)

$$\varepsilon^{FT}_t(s_i) \equiv \frac{J}{2} \xi_t(s_i) + \nu_t(s_i); \hspace{0.5cm} i = 1, ..., N, \hspace{0.5cm} t = 1, ..., T.$$  \hspace{0.5cm} (A.2.6)

Notice that from (2.3.3) in Section 2.3.2, we can rewrite (A.2.4) as

$$\mu^{FT}_t(s_i) = J + \frac{J}{2} (X_t(s_i))' \beta = (X_t(s_i))' \beta^{FT}; \hspace{0.5cm} i = 1, ..., N, \hspace{0.5cm} t = 1, ..., T,$$  \hspace{0.5cm} (A.2.7)

since $(X_t(s_i))'$, the $i$-th row of $X_t(S)$ defined in Section 2.3.2, contains the constant term 1, and $\beta^{FT} = (\beta_0^{FT}, \beta_1^{FT}, \beta_2^{FT})'$ is the vector of corresponding fixed effects for $\mu^{FT}_t(s_i)$.

Recall that the small-scale variation terms $\{\eta_t(s_i)\}$ defined in (2.3.4) have a spatio-temporal CAR structure. Therefore, from (A.2.4)-(A.2.6), we can see that model (A.2.3) based on FTSMR is a Gaussian hierarchical spatio-temporal CAR model that can approximate the Poisson hierarchical model defined in Section 2.3.
A.3

Consider the Poisson hierarchical model given in (2.3.1) and (2.3.2). From (A.1.2) and (A.1.3) in Appendix A.1, we have

\[ E \left( \frac{Z_t(s_i)}{E_t(s_i)} \right) = E(\lambda_t(s_i)), \quad (A.3.1) \]

and

\[ \text{var} \left( \frac{Z_t(s_i)}{E_t(s_i)} \right) = \text{var}(\lambda_t(s_i)) + \frac{E(\lambda_t(s_i))}{E_t(s_i)}; \quad i = 1, \ldots, N, \ t = 1, \ldots, T. \quad (A.3.2) \]

From (2.3.2) in Section 2.3.2,

\[ \lambda_t(s_i) \sim \text{lognormal} \left( \mu_t(s_i), V_t(s_i) \right), \quad (A.3.3) \]

where \( \mu_t(s_i) \) is given by (2.3.3) and

\[ V_t(s_i) = \text{var}(\eta_t(s_i)) + \text{var}(\xi_t(s_i)) = \frac{\tau_t^2 + \sigma_\xi^2}{E_t(s_i)}. \quad (A.3.4) \]

Thus,

\[ E \left( \frac{Z_t(s_i)}{E_t(s_i)} \right) = E(\lambda_t(s_i)) = e^{\mu_t(s_i) + \frac{V_t(s_i)}{2}}, \quad (A.3.5) \]

and

\[ \text{var}(\lambda_t(s_i)) = (e^{V_t(s_i)} - 1)e^{2\mu_t(s_i) + V_t(s_i)}. \quad (A.3.6) \]

Hence, from (A.3.2),

\[ \text{var} \left( \frac{Z_t(s_i)}{E_t(s_i)} \right) = (e^{V_t(s_i)} - 1)e^{2\mu_t(s_i) + V_t(s_i)} + \frac{e^{\mu_t(s_i) + \frac{V_t(s_i)}{2}}}{E_t(s_i)}; \quad (A.3.7) \]

after applying a Taylor expansion for \( \frac{1}{E_t(s_i)} \) small, we obtain

\[ \text{var} \left( \frac{Z_t(s_i)}{E_t(s_i)} \right) \approx \left( \frac{1}{E_t(s_i)} + \frac{1}{2E_t(s_i)^2} \right) \left( \tau_t^2 + \sigma_\xi^2 \right) + \frac{1 + \mu_t(s_i)}{E_t(s_i)}. \quad (A.3.8) \]
Consider the approximate Gaussian model (2.3.17) in Section 2.3.3, based on weighted FTSMRs (derived as an approximation to the Poisson hierarchical model defined in (2.3.1) and (2.3.2)). Specifically, from model (2.3.17), we have, unconditionally,

\[ \text{FTSMR}^w_t(s_i) \sim \text{Normal} \left( \mu^w_t(s_i), V^w_t(s_i) \right); \quad i = 1, \ldots, N, \ t = 1, \ldots, T, \quad (A.4.1) \]

where

\[ V^w_t(s_i) \equiv \text{var} (\eta^w_t(s_i)) + \text{var} (\varepsilon^w_t(s_i)) \approx 1000\tau^2_t + (\sigma^w_\varepsilon)^2. \quad (A.4.2) \]

The approximation in (A.4.2) is because \( \eta^w_t(s_i) \approx \sqrt{1000E_t(s_i)}\eta_t(s_i) \) and \( \text{var}(\eta_t(s_i)) \approx \tau^2_t/E_t(s_i) \), as mentioned in Section 2.3.2. Furthermore, from (2.3.18) in Section 2.3.4 and (A.1.4) in Appendix A.1, for \( 1/E_t(s_i) \) small, we have

\[ \text{FTSMR}^w_t(s_i) = \sqrt{E_t(s_i)}\text{FTSMR}_t(s_i) \approx J \sqrt{E_t(s_i)} \frac{Z_t(s_i)}{E_t(s_i)}, \quad (A.4.3) \]

where recall that \( J = 2\sqrt{1000} \). Now, from (A.4.1) and (A.4.3),

\[ \frac{Z_t(s_i)}{E_t(s_i)} \sim N \left( \frac{\mu^w_t(s_i)}{J \sqrt{E_t(s_i)}}, \frac{V^w_t(s_i)}{(J \sqrt{E_t(s_i)})^2} \right); \quad i = 1, \ldots, N, \ t = 1, \ldots, T. \quad (A.4.4) \]

Define

\[ Z^*_t(s_i) \equiv \sqrt{\frac{Z_t(s_i)}{E_t(s_i)}}/\sqrt{\frac{V^w_t(s_i)}{(J \sqrt{E_t(s_i)})^2}}, \quad i = 1, \ldots, N, \ t = 1, \ldots, T; \quad (A.4.5) \]

thus, approximately,

\[ Z^*_t(s_i) \sim N \left( \frac{\mu^w_t(s_i)}{\sqrt{V^w_t(s_i)}}, 1 \right); \quad i = 1, \ldots, N, \ t = 1, \ldots, T. \quad (A.4.6) \]
Therefore, the distribution of $Z_t^*(s_i)^2$ can be approximated by a non-central chi-square distribution with one degree of freedom. Hence,

$$\text{var}(Z_t^*(s_i)^2) \approx 2 \left( 1 + 2 \frac{(\mu_t^w(s_i))^2}{V_t^w(s_i)} \right); \quad i = 1, ..., N, \ t = 1, ..., T. \quad (A.4.7)$$

Combining (A.4.5) and (A.4.7), we obtain

$$\text{var} \left( \frac{Z_t(s_i)}{E_t(s_i)} \right) \approx \frac{4(\mu_t^w(s_i))^2 V_t^w(s_i)}{J^4 E_t(s_i)^2} + \frac{2(V_t^w(s_i))^2}{J^4 E_t(s_i)^2}. \quad (A.4.8)$$

Then from (A.4.2), (A.4.8) becomes

$$\text{var} \left( \frac{Z_t(s_i)}{E_t(s_i)} \right) \approx \frac{4(\mu_t^w(s_i))^2 (\sigma^w)^2}{J^4 E_t(s_i)^2} + \frac{2(\sigma^w)^4}{J^4 E_t(s_i)^2} + \left[ \frac{(\mu_t^w(s_i))^2 + (\sigma^w)^2}{J^2 E_t(s_i)} + \frac{1}{8} \left( \frac{\tau_t^2}{E_t(s_i)} \right) \right] \left( \frac{\tau_t^2}{E_t(s_i)} \right); \quad (A.4.9)$$

$i = 1, ..., N, \ t = 1, ..., T$. Recall that $J = 2\sqrt{1000}$.

**A.5**

Since (A.4.9) is, to leading order, approximately equal to (A.3.8),

$$\left( \frac{1}{E_t(s_i)} + \frac{1}{2E_t(s_i)^2} \right) \sigma^2 \approx \frac{4(\mu_t^w(s_i))^2 (\sigma^w)^2}{J^4 E_t(s_i)^2} + \frac{2(\sigma^w)^4}{J^4 E_t(s_i)^2} - \frac{(1 + \mu_t)}{E_t(s_i)} + G_t(s_i) \quad (A.5.1)$$

for $i = 1, ..., N, \ t = 1, ..., T$. In (A.5.1),

$$G_t(s_i) \equiv \left[ \frac{(2(\mu_t^w(s_i))^2 - 2J^2 E_t(s_i)) + (2(\sigma^w)^2 - J^2)}{2J^2 E_t(s_i)} \right] \cdot \left( \frac{\tau_t^2}{E_t(s_i)} \right) + \frac{1}{8} \left( \frac{\tau_t^2}{E_t(s_i)} \right)^2. \quad (A.5.2)$$

Recall that in (2.3.19),

$$\mu_t^w(s_i) = \sqrt{E_t(s_i)} \mu_t^{FT}(s_i) \approx \sqrt{E_t(s_i)} \left( J + \frac{J}{2} \mu_t(s_i) \right),$$

and from (2.3.2), if we expect $\lambda_t(s_i)$ to be around 1, then the trend $\mu_t(s_i)$ will be a small number around zero. Thus, in (A.5.2), $2(\mu_t^w(s_i))^2$ is expected to be a value...
closed to $2J^2E_t(s_i)$. Furthermore, from (A.1.6) in Appendix A.1 and (A.2.2) in Appendix A.2, we have

$$\text{var}(\nu_t(s_i)) = \text{var}(\text{FTSMR}_t(s_i)|\lambda_t(s_i)) \approx \frac{1000}{E_t(s_i)}. \quad \text{(A.5.3)}$$

Therefore, from (2.3.14) in Section 2.3.2, (2.3.21) in Section 2.3.4, and (A.5.3), we obtain

$$(\sigma_w^w)^2 \approx \frac{J^2E_t(s_i)}{4} \left( \frac{\sigma_\xi^2}{E_t(s_i)} \right) + \frac{J^2}{4}, \quad \text{(A.5.4)}$$

where recall that $J = 2\sqrt{1000}$. Thus, in (A.5.2),

$$\left[ \frac{(2(\mu_t^w(s_i))^2 - 2J^2E_t(s_i)) + (2(\sigma_w^w(s_i))^2 - J^2)}{2J^2E_t(s_i)} \right] \approx \frac{1}{4} \left( \frac{\sigma_\xi^2}{E_t(s_i)} \right) - \frac{1}{4E_t(s_i)}$$

$$\approx \frac{1}{4} \left( \frac{\sigma_\xi^2}{E_t(s_i)} \right), \quad \text{(A.5.5)}$$

and (A.5.2) can thus be approximated by

$$G_t(s_i) \approx \frac{1}{4} \left( \frac{\sigma_\xi^2}{E_t(s_i)} \right) \left( \frac{\tau_t^2}{E_t(s_i)} \right) + \frac{1}{8} \left( \frac{\tau_t^2}{E_t(s_i)} \right)^2, \quad \text{(A.5.6)}$$

for $i = 1, \ldots, N$, $t = 1, \ldots, T$.

For $\frac{1}{E_t(s_i)}$ small, we expect the small-scale variance, $\frac{\tau_t^2}{E_t(s_i)}$, and the extra-Poisson variance, $\frac{\sigma_\xi^2}{E_t(s_i)}$, to be small. Therefore, (A.5.6) is expected to be small when $\frac{1}{E_t(s_i)}$ is small. Hence, (A.5.1) can be reduced to

$$\left( \frac{1}{E_t(s_i)} + \frac{1}{2E_t(s_i)^2} \right) \sigma_\xi^2 \approx \frac{4(\mu_t^w(s_i))^2(\sigma_w^w)^2}{J^4E_t(s_i)^2} + \frac{2(\sigma_w^w)^4}{J^4E_t(s_i)^2} - \frac{(1 + \mu_t(s_i))}{E_t(s_i)}, \quad \text{(A.5.7)}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$.

In (A.5.7), $\mu_t(s_i)$ can be estimated by $\hat{\mu}_t(s_i)$, which is the fitted value of Poisson GLM defined by (2.2.4) in Section 2.2.3; and $\mu_t^w(s_i)$ can be estimated by $\hat{\mu}_t^w(s_i)$, which is the fitted value of the weighted linear regression model defined by (2.2.5) in Section 2.2.4. If we further aggregate both sides of (A.5.7) over $i = 1, \ldots, N$ and $t = 1, \ldots, T$, $\sigma_\xi^2$ can be written approximately as a function of $(\sigma_w^w)^2$; that is,

$$\sigma_\xi^2 \approx Q_0 + Q_1 \cdot (\sigma_w^w)^2 + Q_2 \cdot (\sigma_w^w)^4, \quad \text{(A.5.8)}$$
where

\[ Q_0 = - \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{(1 + \hat{\mu}_t(s_i))}{E_t(s_i)}, \]  

(A.5.9)

\[ Q_1 = \frac{4}{J^4} \sum_{i=1}^{N} \sum_{t=1}^{T} \left( \frac{\hat{\mu}_t^w(s_i)^2}{E_t(s_i)^2} \right), \]  

(A.5.10)

\[ Q_2 = \frac{2}{J^4} \sum_{i=1}^{N} \sum_{t=1}^{T} \left( \frac{1}{E_t(s_i)^2} \right). \]  

(A.5.11)
Appendix B

DETAILS OF DERIVING THE NONLINEAR W-SCALE PROCESSES AND THE LINEAR APPROXIMATIONS OF HSIR AND EXTENDED HSIR MODELS OF CHAPTER 3

B.1

Here we give the derivation of the nonlinear dynamical structure of $\mu^W(t)$ as in (3.2.18) and (3.2.19) in Section 3.2.1.2.

Assume $P_R(t) > 0$, for $t = 1, 2, \ldots$. From the difference equations (3.2.12)-(3.2.14), we can obtain

$$
P_S(t + 1) = \frac{P_S(t) - \beta \Delta \lambda N P_S(t) P_I(t) + \phi \Delta P_R(t)}{P_R(t) + \gamma \Delta P_I(t) - \phi \Delta P_R(t)}
$$

(B.1.1)

$$
P_I(t + 1) = \frac{P_I(t) + \beta \Delta \lambda N P_S(t) P_I(t) - \gamma \Delta P_I(t)}{P_R(t) + \gamma \Delta P_I(t) - \phi \Delta P_R(t)}.
$$

(B.1.2)

Notice that equations (B.1.1) and (B.1.2) can be rewritten as

$$
P_S(t + 1) = \frac{P_S(t) - \beta \Delta \lambda N P_S(t) P_I(t) + \phi \Delta P_R(t)}{\frac{1}{P_R(t)} + \gamma \Delta \frac{P_I(t)}{P_R(t)} - \frac{\phi \Delta}{P_R(t)}}
$$

(B.1.3)

$$
P_I(t + 1) = \frac{\beta \Delta \frac{P_S(t)}{P_R(t)} P_I(t) + (1 - \gamma \Delta) \frac{P_I(t)}{P_R(t)} - \frac{\phi \Delta}{P_R(t)}}{\frac{1}{P_R(t)} + \gamma \Delta \frac{P_I(t)}{P_R(t)} - \frac{\phi \Delta}{P_R(t)}}.
$$

(B.1.4)

From (3.2.15) and (3.2.16),

$$
P_S(t) = \frac{\exp(W_S(t))}{1 + \exp(W_S(t)) + \exp(W_I(t))}
$$

(B.1.5)
\[ P_I(t) = \frac{\exp(W_I(t))}{1 + \exp(W_S(t)) + \exp(W_I(t))}. \]  
(B.1.6)

Then substituting (B.1.5) and (B.1.6) into (3.2.8), we obtain
\[ P_R(t) = \frac{1}{1 + \exp(W_S(t)) + \exp(W_I(t))}. \]  
(B.1.7)

Hence,
\[ \frac{P_S(t)}{P_R(t)} = \exp(W_S(t)) \]  
(B.1.8)
\[ \frac{P_I(t)}{P_R(t)} = \exp(W_I(t)) \]  
(B.1.9)
\[ \frac{1}{P_R(t)} = \frac{\exp(W_S(t)) + \exp(W_I(t)) + 1}{1}. \]  
(B.1.10)

For \( t = 1, 2, \ldots \), substitute (B.1.8)-(B.1.10) into equations (B.1.3) and (B.1.4) to obtain
\[ \exp(W_S(t+1)) = \exp(W_S(t)) \cdot \left[ 1 + \frac{\phi\Delta}{\exp(W_S(t))} - \frac{(\beta\Delta\lambda_N)\exp(W_I(t))}{1 + \exp(W_S(t)) + \exp(W_I(t))} \right] \cdot \frac{1}{[1 + \gamma\Delta\exp(W_I(t)) - \phi\Delta]}, \]  
(B.1.11)

and
\[ \exp(W_I(t+1)) = \exp(W_I(t)) \cdot \left[ 1 - \gamma\Delta + \frac{(\beta\Delta\lambda_N)\exp(W_S(t))}{1 + \exp(W_S(t)) + \exp(W_I(t))} \right] \cdot \frac{1}{[1 + \gamma\Delta\exp(W_I(t)) - \phi\Delta]} \]  
(B.1.12)

Taking logarithms on both sides of (B.1.11) and (B.1.12), for \( t = 1, 2, \ldots \), we obtain,
\[ W_S(t+1) = W_S(t) + \log \left[ 1 + \frac{\phi\Delta}{\exp(W_S(t))} - \frac{(\beta\Delta\lambda_N)\exp(W_I(t))}{1 + \exp(W_S(t)) + \exp(W_I(t))} \right] \]
\[ + \log \left[ \frac{1}{1 + \gamma\Delta\exp(W_I(t)) - \phi\Delta} \right], \]  
(B.1.13)
and

\[ W_I(t + 1) = W_I(t) + \log \left[ 1 - \gamma \Delta + \beta \Delta \lambda_N \exp \left( W_S(t) \right) \frac{1 + \exp(W_S(t)) + \exp(W_I(t))}{1 + \gamma \Delta \exp(W_I(t)) - \phi} \right]. \tag{B.1.14} \]

Then (B.1.13) and (B.1.14) are used to define \( \mu^W(t) \) in the nonlinear autoregressive structure given by (3.2.17), which captures the uncertainties in the hidden epidemic process.

\section*{B.2}

Here, we give the details of how to derive a well calibrated linear process \( \{ \mu^{LW}(t) \} \), as defined by (3.3.2) and (3.3.3), to approximate \( \{ \mu^W(t) \} \) given by (3.2.18) and (3.2.19). Specifically, for \( t = 1, 2, ..., \{ \hat{A}_i(t) : i = 1, ..., 10 \} \) are initializations of the nonlinear components in equations (3.2.18) and (3.2.19), as shown in Table 3.1.

Consider \( \mu_S^W(t) \) given by (3.2.18) and use a Taylor-series expansion to second order. The second term on the right-hand side is:

\[
\log \left( 1 + \frac{\phi \Delta}{e^{W_S(t)}} - \frac{\beta \Delta \lambda_N e^{W_I(t)}}{1 + e^{W_S(t)} + e^{W_I(t)}} \right)
\]

\[
= \log(A_1(t)) + \log \left[ 1 + \left( 1 + \frac{\phi \Delta}{e^{W_S(t)}} - \frac{\beta \Delta \lambda_N e^{W_I(t)}}{(1 + e^{W_S(t)} + e^{W_I(t)})} - 1 \right) \right]
\]

\[
\approx \log(\hat{A}_1(t)) + \left( 1 + \frac{\phi \Delta}{e^{W_S(t)}} - \frac{\beta \Delta \lambda_N e^{W_I(t)}}{(1 + e^{W_S(t)} + e^{W_I(t)})} \right) \hat{A}_1(t)
\]

\[
- \frac{1}{2} \left( 1 + \frac{\phi \Delta}{e^{W_S(t)}} - \frac{\beta \Delta \lambda_N e^{W_I(t)}}{(1 + e^{W_S(t)} + e^{W_I(t)})} \right)^2. \tag{B.2.1}
\]
Now, \( \frac{e^{W_I(t)}}{1 + e^{W_S(t)} + e^{W_I(t)}} \) can be further expanded using a Taylor series to second order:

\[
\frac{e^{W_I(t)}}{1 + e^{W_S(t)} + e^{W_I(t)}} = 1 - \frac{1}{1 - \left( -\frac{e^{W_I(t)}}{1 + e^{W_S(t)}} \right)} \\
\approx 1 - \left[ \frac{1}{1 - \hat{A}_7(t)} + \frac{-e^{W_I(t)} - \hat{A}_7(t)}{(1 - \hat{A}_7(t))^2} + \frac{-\frac{e^{W_I(t)}}{1 + e^{W_S(t)}} - \hat{A}_7(t)}{(1 - \hat{A}_7(t))^3} \right].
\]

\( \text{(B.2.2)} \)

Then we expand the remaining nonlinear component in (B.2.2), \( \frac{e^{W_I(t)}}{1 + e^{W_S(t)}} \), using a Taylor series to second order:

\[
\frac{e^{W_I(t)}}{1 + e^{W_S(t)}} = \left[ e^{(W_I(t) - W_S(t))} \right] \cdot \left[ \frac{1}{1 - (-e^{-W_S(t)})} \right] \\
\approx \left[ e^{\hat{A}_4(t)} + e^{\hat{A}_4(t)}(W_I(t) - W_S(t) - \hat{A}_4(t)) + \frac{e^{\hat{A}_4(t)}}{2}(W_I(t) - W_S(t) - \hat{A}_4(t))^2 \right] \\
\cdot \left[ \frac{1}{1 - \hat{A}_9(t)} + \frac{(-e^{-W_S(t)} - \hat{A}_9(t))}{(1 - \hat{A}_9(t))^2} + \frac{-e^{W_S(t)} - \hat{A}_9(t)}{(1 - \hat{A}_9(t))^3} \right],
\]

\( \text{(B.2.3)} \)

and

\[
e^{-W_S(t)} \approx e^{\hat{A}_5(t)} + e^{\hat{A}_5(t)} (-W_S(t) - \hat{A}_5(t)) + \frac{e^{\hat{A}_5(t)}}{2} (-W_S(t) - \hat{A}_5(t))^2.
\]

\( \text{(B.2.4)} \)

Upon substituting (B.2.4) into (B.2.3), we obtain:

\[
\frac{e^{W_I(t)}}{1 + e^{W_S(t)}} \approx B_0(t) + B_1(t)W_S(t) + B_2(t)W_I(t) \\
+ o(W_S(t)^2) + o(W_I(t)^2) + o(W_S(t)W_I(t)),
\]

\( \text{(B.2.5)} \)
where \( \{B_i(t) : i = 0, 1, 2\} \) are defined in equations (3.3.10)-(3.3.12) and (3.3.16), which are repeated here for completeness:

\[
B_0(t) = e^{\hat{A}_1(t)}(1 - \hat{A}_4(t))B^*(t)
\]

\[
B_1(t) = \frac{e^{(\hat{A}_5(t)+\hat{A}_1(t))}(1 - \hat{A}_4(t)) - e^{\hat{A}_1(t)}B^*(t)}{(1 - \hat{A}_9(t))^2}
\]

\[
B_2(t) = e^{\hat{A}_1(t)}B^*(t),
\]

for

\[
B^*(t) = \frac{1}{1 - \hat{A}_9(t)} + \frac{e^{\hat{A}_5(t)}(\hat{A}_5(t) - 1) - \hat{A}_9(t)}{(1 - \hat{A}_9(t))^2}.
\]

Hence, combining (B.2.1), (B.2.2), and (B.2.5), we can finally approximate

\[
\log \left( 1 + \frac{\phi \Delta}{\exp(W_S(t))} - \frac{\beta \Delta \lambda_N e^{W_I(t)}}{1 + e^{W_S(t)} + e^{W_I(t)}} \right)
\]

in equation (3.2.18) with

\[
\begin{align*}
\log \hat{A}_1(t) + & \frac{1}{\hat{A}_1(t)} + \phi \Delta \left[ \frac{e^{(\hat{A}_5(t))}(1 - \hat{A}_5(t))}{\hat{A}_1(t)} \right] \\
- & \frac{\beta \Delta \lambda_N}{\hat{A}_1(t)} \left( 1 - \frac{1}{1 - \hat{A}_7(t)} + \frac{B_0(t) + \hat{A}_7(t)}{(1 - \hat{A}_7(t))^2} \right) - 1 \\
+ & \left[ 1 - \frac{\phi \Delta e^{\hat{A}_5(t)}}{\hat{A}_1(t)} - \frac{\beta \Delta \lambda_N B_1(t)}{\hat{A}_1(t)(1 - \hat{A}_7(t))^2} \right] W_S(t) - \left[ \frac{\beta \Delta \lambda_N B_2(t)}{\hat{A}_1(t)(1 - \hat{A}_7(t))^2} \right] W_I(t) \\
+ & o \left( W^2_S(t) \right) + o \left( W^2_I(t) \right) + o \left( W_S(t)W_I(t) \right).
\end{align*}
\]

Furthermore, the third term, \( \log \left( 1 + \Delta \gamma e^{W_I(t)} - \Delta \phi \right), \) on the right-hand side of (3.2.18), can also be expanded in a Taylor series to second order: For \( t = 1, 2, \ldots, \)

\[
\log \left( 1 + \Delta \gamma e^{W_I(t)} - \Delta \phi \right) = \log(A_2(t)) + \log \left[ 1 + \left( \frac{1 + \gamma \Delta e^{W_I(t)} - \Delta \phi}{A_2(t)} - 1 \right) \right]
\]

\[
\approx \log(A_2(t)) + \left( \frac{1 + \gamma \Delta e^{W_I(t)} - \Delta \phi}{A_2(t)} - 1 \right)
\]

\[
- \frac{1}{2} \left( \frac{1 + \gamma \Delta e^{W_I(t)} - \Delta \phi}{A_2(t)} - 1 \right)^2.
\]

(B.2.7)
Now, $e^{W_I(t)}$ can be further expanded using a Taylor series to second order:

$$e^{W_I(t)} \approx e^{\hat{A}_6(t)} + e^{\hat{A}_6(t)}(W_I(t) - \hat{A}_6(t)) + \frac{e^{\hat{A}_6(t)}}{2} (W_I(t) - \hat{A}_6(t))^2. \tag{B.2.8}$$

Upon substituting (B.2.8) into (B.2.7), we obtain:

$$\log \left( 1 + \gamma \Delta e^{W_I(t)} - \phi \Delta \right) \approx \log \hat{A}_2(t) + \frac{1}{\hat{A}_2(t)} + \frac{\gamma \Delta}{\hat{A}_2(t)} e^{\hat{A}_6(t)}(1 - \hat{A}_6(t)) - \frac{\phi \Delta}{\hat{A}_2(t)} - 1$$

$$+ \frac{\gamma \Delta e^{\hat{A}_6(t)}}{\hat{A}_2(t)} W_I(t) + o \left( W_I^2(t) \right). \tag{B.2.9}$$

Consider $\mu^W_I(t)$ given by (3.2.19) and use a Taylor-series expansion to second order.

The second term on the right-hand side is:

$$\log \left( 1 - \gamma \Delta + \frac{\beta \Delta \lambda_N e^{W_S(t)}}{1 + e^{W_S(t)} + e^{W_I(t)}} \right)$$

$$= \log(A_3(t)) + \log \left[ 1 + \left( 1 - \gamma \Delta + \frac{\beta \Delta \lambda_N e^{W_S(t)}}{1 + e^{W_S(t)} + e^{W_I(t)}} - 1 \right) \right]$$

$$\approx \log(A_3(t)) + \left( 1 - \gamma \Delta + \frac{\beta \Delta \lambda_N e^{W_S(t)}}{A_3(t)} - 1 \right) - \frac{1}{2} \left( 1 - \gamma \Delta + \frac{\beta \Delta \lambda_N e^{W_S(t)}}{A_3(t)} - 1 \right)^2. \tag{B.2.10}$$

Now, $\frac{e^{W_S(t)}}{1 + e^{W_S(t)} + e^{W_I(t)}}$ can be further expanded using a Taylor series to second order:

$$\frac{e^{W_S(t)}}{1 + e^{W_S(t)} + e^{W_I(t)}} = \frac{1}{1 - \left( -\frac{1 + e^{W_I(t)}}{e^{W_S(t)}} \right)}$$

$$\approx \frac{1}{1 - \hat{A}_{10}(t)} + \frac{\left( \frac{1 + e^{W_I(t)}}{e^{W_S(t)}} - \hat{A}_{10}(t) \right)}{(1 - \hat{A}_{10}(t))^2} + \frac{\left( \frac{1 + e^{W_I(t)}}{e^{W_S(t)}} - \hat{A}_{10}(t) \right)^2}{(1 - \hat{A}_{10}(t))^3}. \tag{B.2.11}$$
Then we expand the remaining nonlinear component in (B.2.11), \( \frac{1 + e^{W_I(t)}}{e^{W_S(t)}} \), using a Taylor series to second order:

\[
\frac{1 + e^{W_I(t)}}{e^{W_S(t)}} = e^{W_I(t) - W_S(t)} + e^{-W_S(t)} \\
\approx \left[ e^{\hat{A}_4(t)} + e^{\hat{A}_4(t)}(W_I(t) - W_S(t) - \hat{A}_4(t)) + \frac{e^{\hat{A}_4(t)}}{2} (W_I(t) - W_S(t) - \hat{A}_4(t))^2 \right] \\
+ \left[ e^{\hat{A}_5(t)} + e^{\hat{A}_5(t)}(-W_S(t) - \hat{A}_5(t)) + \frac{e^{\hat{A}_5(t)}}{2} (-W_S(t) - \hat{A}_5(t))^2 \right] .
\]

(B.2.12)

We can rewrite (B.2.12) as

\[
\frac{1 + e^{W_I(t)}}{e^{W_S(t)}} \approx B_3(t) + B_4(t)W_S(t) + B_5(t)W_I(t) \\
+ o (W_S(t)^2) + o (W_I(t)^2) + o (W_S(t)W_I(t)),
\]

(B.2.13)

where \( \{ B_i(t) : i = 3, 4, 5 \} \) are defined in (3.3.13)-(3.3.15), which are repeated here for completeness:

\[
B_3(t) = e^{\hat{A}_4(t)}(1 - \hat{A}_4(t)) + e^{\hat{A}_5(t)}(1 - \hat{A}_5(t)) \\
B_4(t) = -e^{\hat{A}_4(t)} - e^{\hat{A}_5(t)} \\
B_5(t) = e^{\hat{A}_4(t)}.
\]

Hence, combining (B.2.10), (B.2.11), and (B.2.13), we can finally approximate

\[
\log \left[ (1 - \gamma \Delta) + \frac{\beta \Delta \lambda N e^{W_S(t)}}{1 + e^{W_S(t)} + e^{W_I(t)}} \right]
\]

in equation (3.2.19) with

\[
\log \hat{A}_3(t) + \frac{1 - \gamma \Delta}{\hat{A}_3(t)} + \frac{\beta \Delta \lambda N}{\hat{A}_3(t)(1 - \hat{A}_{10}(t))} - \frac{\beta \Delta \lambda N (B_3(t) + \hat{A}_{10}(t))}{\hat{A}_3(t)(1 - \hat{A}_{10}(t))^2} - 1 \\
+ \left[ \frac{-\beta \Delta \lambda N B_4(t)}{\hat{A}_3(t)(1 - \hat{A}_{10}(t))^2} \right] W_S(t) + \left[ \frac{1 - \beta \Delta \lambda N B_5(t)}{\hat{A}_3(t)(1 - \hat{A}_{10}(t))^2} \right] W_I(t) \\
+ o (W_S^2(t)) + o (W_I^2(t)) + o (W_S(t)W_I(t)).
\]

(B.2.14)
The third component on the right-hand side of (3.2.19) is identical to that of (3.2.18).

Therefore, (B.2.6), (B.2.9), and (B.2.14) yields the linear dynamical process \( \mu^{LW}(t) \) as defined in (3.3.2) and (3.3.3) in Section 3.3.1. That is, \( \mu^{LW}(t) \) approximates the nonlinear dynamical process, \( \mu^{W}(t) \), in the HSIR model defined by equations (3.2.18) and (3.2.19).

**B.3**

Here we give the derivation of the nonlinear dynamical structure of \( \mu^{D}(t) \) as in (3.2.36) and (3.2.37) in Section 3.2.2.2.

Recall that in the extended HSIR model, the hidden true proportions, \( P_S(t) \) and \( P_I(t) \), are from a joint Logistic-Normal distribution conditional on the odds ratios, \( W_S(t) \) and \( W_I(t) \), and \( E_{PS|PR} \) and \( E_{PI|PR} \) are the conditional means of \( P_{S+1} \) and \( P_{I+1} \), as defined in (3.2.29) and (3.2.30). Based on equations (B.1.3) and (B.1.4) derived in Appendix B.1, we can obtain approximations to the conditional means, using the leading order of a delta method: For \( t = 1, 2, \ldots \),

\[
E_{PS|PR}(w(t+1)) \approx \frac{E_{PS|PR}(w(t))E_{PS|PR}(w(t)) - \beta \Delta \lambda \gamma E_{PS|PR}(w(t))E_{PS|PR}(w(t)) + \phi \Delta E_{PS|PR}(w(t))}{E_{PS|PR}(w(t)) + \gamma \Delta E_{PS|PR}(w(t))E_{PS|PR}(w(t)) - \phi \Delta E_{PS|PR}(w(t))} \tag{B.3.1}
\]

\[
E_{PI|PR}(w(t+1)) \approx \frac{\beta \Delta E_{PS|PR}(w(t))E_{PS|PR}(w(t)) + (1 - \gamma \Delta) E_{PS|PR}(w(t))E_{PS|PR}(w(t))}{E_{PS|PR}(w(t)) + \gamma \Delta E_{PS|PR}(w(t))E_{PS|PR}(w(t)) - \phi \Delta E_{PS|PR}(w(t))} \tag{B.3.2}
\]

In equations (B.3.1) and (B.3.2), the conditional means, \( E_{PS|PR}(w(t)) \) and \( E_{PS|PR}(w(t)) \), for \( t = 1, 2, \ldots \), can be obtained from (3.2.31) and (3.2.32) in Section 3.2.2.2. Furthermore, from the mass-balance constraint given in (3.2.7), we can obtain

\[
\frac{1}{P_R(t)} = 1 + \frac{P_S(t)}{P_R(t)} + \frac{P_I(t)}{P_R(t)}. \tag{B.3.3}
\]
Thus,

\[ E_{\frac{1}{r_r}} \mathbf{W}(t) \equiv E \left( \frac{1}{P_R(t)} | \mathbf{W}(t) \right) = 1 + E_{\frac{p_r}{r_r}} \mathbf{W}(t) + E_{\frac{p_r}{r_r}} \mathbf{W}(t). \quad (B.3.4) \]

Substituting (3.2.31), (3.2.32), and (B.3.4) into (B.3.1) and (B.3.2), for \( t = 1, 2, \ldots \), we obtain:

\[
\exp (W_S(t + 1)) = \exp (D_S(t)) \cdot \left[ 1 + \frac{\phi \Delta}{\exp (D_S(t))} - \frac{(\beta \Delta \lambda_N \exp (D_I(t)))}{1 + \exp (D_S(t)) + \exp (D_I(t))} \right] \frac{1}{[1 + \gamma \Delta \exp (D_I(t)) - \phi \Delta]} \cdot \exp \left( -\frac{1}{2} \sigma^2_P (t + 1) \right), \quad (B.3.5)
\]

and

\[
\exp (W_I(t + 1)) = \exp (D_I(t)) \cdot \left[ 1 - \gamma \Delta + \frac{\beta \Delta \lambda_N \exp (D_S(t))}{1 + \exp (D_S(t)) + \exp (D_I(t))} \right] \frac{1}{[1 + \gamma \Delta \exp (D_I(t)) - \phi \Delta]} \cdot \exp \left( -\frac{1}{2} \sigma^2_P (t + 1) \right). \quad (B.3.6)
\]

Recall that in (3.2.33) and (3.2.34),

\[ D_S(t) \equiv W_S(t) + \frac{1}{2} \sigma^2_P (t), \]

\[ D_I(t) \equiv W_I(t) + \frac{1}{2} \sigma^2_P (t), \]

are the adjusted log odds ratios of susceptible over recovered populations, and infectious over recovered populations, respectively. Taking logarithms on both sides of (B.3.5) and (B.3.6), for \( t = 1, 2, \ldots \), we obtain

\[
W_S(t + 1) = D_S(t) + \log \left[ 1 + \frac{\phi \Delta}{\exp (D_S(t))} - \frac{(\beta \Delta \lambda_N \exp (D_I(t)))}{1 + \exp (D_S(t)) + \exp (D_I(t))} \right] \frac{1}{[1 + \gamma \Delta \exp (D_I(t)) - \phi \Delta]} - \log [1 + \gamma \Delta \exp (D_I(t)) - \phi \Delta] - \left[ \frac{1}{2} \sigma^2_P (t + 1) \right], \quad (B.3.7)
\]
\[ W_I(t + 1) = D_I(t) \]
\[ + \log \left[ 1 - \gamma \Delta + \frac{(\beta \Delta \lambda_N) \exp (D_S(t))}{1 + \exp (D_S(t)) + \exp (D_I(t))} \right] \]
\[ - \log \left[ 1 + \gamma \Delta \exp (D_I(t)) - \phi \Delta \right] - \left( \frac{1}{2} \sigma^{2}_{P_{I}}(t + 1) \right), \] (B.3.8)

Then the first three terms in (B.3.7) and (B.3.8) are used to define \( \mu^D(t) \) in the nonlinear autoregressive structure given by (3.2.35), which captures the uncertainties in the hidden epidemic process.

**B.4**

Here we give the details of how to derive a well calibrated linear process \( \{ \mu^{LD}(t) \} \), as defined by (3.3.20) and (3.3.21), to approximate \( \{ \mu^D(t) \} \), given by (3.2.36) and (3.2.37). Recall that in (3.2.33) and (3.2.34), we denote \( D_S(t) \) and \( D_I(t) \) to be the adjusted log odds ratios of susceptible over recovered populations, and infectious over recovered populations, respectively, for \( t = 1, 2, \ldots \). If we compare the nonlinear dynamical process of the extended HSIR model defined by (3.2.36) and (3.2.37), to that of the HSIR model defined by (3.2.18) and (3.2.19), we notice that the only difference is that (3.2.36) and (3.2.37) depend on adjusted log odds ratios, \( D_S(t) \) and \( D_I(t) \), rather than the (unadjusted) log odds ratios, \( W_S(t) \) and \( W_I(t) \).

We can follow the same linearization procedures as illustrated in Appendix B.2, but replace \( W_S(t) \) and \( W_I(t) \) with \( D_S(t) \) and \( D_I(t) \), to obtain Taylor-series approximations to second orders, of the nonlinear components,

\[ \log \left[ 1 + \frac{\phi \Delta}{\exp (D_S(t))} - \frac{\beta \Delta \lambda_N \exp (D_I(t))}{1 + \exp (D_S(t)) + \exp (D_I(t))} \right], \]
\[ \log \left[ 1 + \gamma \Delta \exp (D_I(t)) - \phi \Delta \right], \]
and

\[
\log \left[ 1 - \gamma \Delta + \frac{\beta \Delta \lambda N \exp(D_S(t))}{1 + \exp(D_S(t)) + \exp(D_I(t))} \right].
\]

Based on those approximate values, it is straightforward to obtain the linear dynamical process \(\mu^{LD}(t)\) as defined in (3.3.20) and (3.3.21) in Section 3.3.2. That is, \(\mu^{LD}(t)\) approximates the nonlinear dynamical process, \(\{\mu^D(t)\}\), in the extended HSIR model defined by equations (3.2.36) and (3.2.37).
BIBLIOGRAPHY


