University of Alberta

Spatio-Temporal Prediction Modeling of Clusters of Influenza Cases in Edmonton

by

Weiyu Qiu

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Epidemiology

Department of Public Health Sciences

Fall 2011
Edmonton, Alberta

Dedication

To my dearest husband, Mu Lin, and greatest parents, Zhicheng Qiu and Li Li.

Abstract

Timely, accurate predictions of potential influenza epidemics are essential for healthcare providers and policy makers as the epidemics can result in heavy demands for health services. Current statistical modeling of surveillance data has limited prediction abilities and often fails to respond effectively to the outbreaks. The first part of this thesis, a collaboration with Alberta Health Services, aims at predicting clusters of influenza cases in Edmonton weeks in advance, using real-time data collected from emergency-department visits by Alberta Real Time Syndromic Surveillance Net. The 2004-2009 data are analyzed by spatio-temporal modeling and predictions are cross-validated. In the second part of this thesis, a related theoretical work on multivariate modeling, with spatio-temporal modeling as a potential application, is presented, proving that every conditional second moment is linear in the empirical second moment of the conditioning vector if and only if the distribution belongs to the multivariate Pearson type VII family.

Key Words: influenza prediction, generalized linear mixed model, pseudo-likelihood, cross-validation, multivariate Pearson Type VII family.

Acknowledgement

This thesis is under kind support by Alberta Ingenuity Center for Machine Learning (AICML).

I hereby sincerely express my deepest gratitude to my thesis examining committee, Dr. Irina Dinu, Dr. Josè Miguel Martinez, Dr. Peng Zhang, and Dr. Yutaka Yasui, for their expertise and support for this academic endeavor.

I would like to express my greatest thankfulness to my supervisor, Dr. Yutaka Yasui for all his invaluable advice and guidance throughout this thesis. His rigorous attitude toward the research work has set a great example for me and my future work.

The influenza prediction modeling part of this thesis is a collaboration with Dr. Shihe Fan at Alberta Health and Services, and I would like to thank him for his suggestions and provision of the data. My sincerest thank you to Dr. Anamaria Savu and Dr. Biao Wu for their valuable suggestions and patient help with the methodological part of MP VII distributions. I would also like to acknowledge my colleges in Dr. Yasui's team for their insights and comments on this thesis.

Finally, I want to express my sincere thankfulness to my dearest husband and parents for their love, encouragement and support.

Table of Contents

List of Figures

List of Abbreviations

AHS - Alberta Health Services

AHW - Alberta Health and Wellness

AICML - Alberta Ingenuity Center for Machine Learning

ARTSSN - Alberta Real Time Syndromic Surveillance Net

BeII - Beta Type II

CHICA - Community and Hospital Infection Control Association

EC - Elliptical Distribution

ED - Emergency Department

EDIS - Emergency Department Information

GISN - Global Influenza Surveillance Network

GLMM - Generalized Linear Mixed Model

GMVt - Generalized Multivariate t

HL - Health Link

LOOCV - Leave-one-out Cross-validation

MAD - Mean Absolute Deviation

MP VII - Multivariate Pearson Type VII

MVN - Multivariate Normal

MVt - Multivariate t

PDI - Patient Demographic Information

WHO - World Health Organization

1 Introduction

Pandemic of influenza can result in a heavy demand for health services. Over the past 30 years, surveillance networks for real-time monitoring of influenza activity have been developed worldwide to detect epidemics rapidly and to estimate the annual impact of influenza. However, these networks provide information on present, not future, activity. Therefore, it is essential to make timely and accurate prediction based on statistical modeling of surveillance data, so that healthcare providers and policy makers can have proper preparation and potential alleviation of the potential epidemics.

A large body of work has been devoted to the real-time detection of influenza outbreaks, defined as some increase above a historical baseline threshold (Le Strat & Carrat, 1999; Hashimoto *et al*., 2000). However, only a limited range of approaches has been developed to predict the spread of the epidemic process. These approaches fall into two categories: those that model the diffusion mechanisms and those that model the epidemic curve. Baroyan et al. first used the first approach in 1971 to describe the time and geographic spread of influenza. Longini *et al*. (1986), Flahault *et al*. (1988), and Flahault *et al*. (1994) also did similar applications based on this approach. Application of these models to retrospective data provided useful insight into the diffusion mechanisms but was not proven efficient for prospective forecasting.

The second approach is based on time-series modeling of the epidemic curve. Autoregressive seasonal linear models (Box & Jenkins, 1976) have previously been applied to influenza surveillance data (Stroup, *et al*., 1988; Quenel & Dab, 1998). However, these models did not take into account the spatial correlations related to the influenza spread process and could not be adjusted to sudden changes in dynamics. Nevertheless, although these models were not used in the past for operational forecasts of influenza epidemics on a national or regional level, they can be considered reference models against which new methods should be tested.

My thesis work aims at predicting clusters of influenza cases in Edmonton, using real time data collected over time and space from emergency department (ED) visits by Alberta Real Time Syndromic Surveillance Net (ARTSSN). Spatio-temporal modeling and pseudo likelihood estimation of the parameters are used and a 5-fold cross-validation is performed on 5-year daily data collected over 2004 to 2009. The plan for the remainder of this thesis is as follows: in Section 2, we describe some facts about influenza and current situation of influenza surveillance. Specifically, the ARTSSN is briefly described since one of its subsets is used as our analysis data. A few statistical topics that are related to our analytical method are also described, including natural cubic splines, generalized linear mixed models (GLMMs), pseudo-likelihood estimation, and cross-validation. Section 3 is the

method part of the thesis. We start by performing a GLMM analysis on the six-year data with same natural cubic splines of daily indices for each year to depict the systematic pattern of annual influenza activities. Then a temporal-spatio model is applied on the daily residuals, with pseudo-likelihood method to estimate the parameters. In particular, unequal weighted pseudo-likelihood is used afterwards to improve the prediction of influenza peaks. Section 4 is a theoretical piece about multivariate data. In this section, we discuss a property of the multivariate Pearson type VII family and use simulated data to illustrate the application of this unique property. Finally in Section 5, we draw conclusions, and make a few suggestions for future work after some discussions about the thesis.

2 Background

We begin this section by introducing some facts about influenza, followed by describing the burden of influenza in Canada and worldwide. We then take a brief review of current work done about influenza surveillance and prediction, and address the importance of accurate and timely prediction of influenza. Then we provide a brief introduction to the ARTSSN, part of which is used as our data set for the statistical modeling and prediction of influenza visits. Finally, we introduce some statistical methods that are used in the method section, including natural cubic splines, GLMM, pseudo-likelihood estimation, and cross-validation.

2.1 Key Facts about Influenza

Influenza, commonly referred to as flu, is a contagious respiratory illness caused by influenza viruses. There are two types of influenza virus known as type A and B. These two viruses, the so called human influenza viruses that routinely spread in people, are responsible for seasonal flu epidemics each year.

The most common symptoms of influenza are chills, fever, cough, sore throat, runny or stuffy nose, muscle or body aches, headaches, and fatigue. Some people may have vomiting and diarrhea, though this is more common in children than adults.

Although influenza is often confused with other influenza-like illnesses, especially the common cold, it is a more severe disease than the latter. It can cause mild to severe illness, and at times can lead to death. Older people, young children, and people with certain health conditions, are at higher risk for serious influenza complications. In occasional cases, even for healthy young adults, influenza can cause either respiratory distress syndrome or pneumonia, which manifest with a difficulty in breathing.

Most experts believe that flu viruses spread mainly by droplets from people with influenza when they cough, sneeze or talk. These droplets can land in the mouths or noses of people nearby. Less often, a person might also get influenza by touching a surface or object that has flu virus on it and then touching their own mouth, eyes or noses.

Influenza typically lasts a week to 10 days, and the incubation period for influenza is 24 to 72 hours. It may pass to someone else before one knows he/she is sick, as well as while he/she is sick. Most healthy adults may be able to infect others beginning one day before symptoms develop and up to five to seven days after becoming sick. Some people, especially children and people with weakened immune systems, might be able to infect others for even longer time.

2.2 Burden of Influenza

Influenza epidemics and pandemic have a huge impact on society and individuals. The weight and scope of the burden of influenza varies with the age and underlying health of the patient. The disease imposes a significant burden on all individuals, but hospitalization and treatment occur more frequently in high-risk patients who are elderly and/or with certain underlying medical conditions.

According to the World Health Organization (WHO), 5 - 15% of the population are affected with upper respiratory tract infections in annual influenza epidemics. Although difficult to assess, these annual epidemics are thought to result in between three and five million cases of severe illness and between 250,000 and 500,000 deaths every year around the world. This imposes a considerable economic burden in the form of hospital and other health care costs and lost productivity. In the USA, for example, recent estimates put the annual influenza-related costs to US$ 87.1 billion (Molinari, *et al.*, 2007).

In Canada, the death rate for influenza is 500 - 1,500 cases per year. According to the Community and Hospital Infection Control Association (CHICA) - Canada, the seasonal influenza totals are increasing rapidly in the past five years, from 7,422 cases in 2004 - 2005 to 38,980 in 2009 - 2010. The increased cases in 2009 are mainly due to the pandemic H1N1 influenza virus. Dianne & Thomson (2006) also showed that the total annual costs of influenza were estimated at CA$1 billion in Canada, and this is expected to increase in coming years.

In addition to the direct costs of medical care, the indirect costs of influenza are substantial and stem largely from absenteeism and loss of work productivity. Estimates of the cost of influenza in the USA, France and Germany have shown that indirect costs can be five- to 10-fold higher than direct costs. Other intangible costs associated with influenza include impaired performance which can reduce reaction times, and adverse effects on the quality of life of patients and their families.

2.3 Current Influenza Surveillance and Prediction

2.3.1 Overview of Influenza Surveillance Network

The WHO Global Influenza Surveillance Network (GISN) serves as a global alert mechanism for the emergence of influenza viruses with pandemic potential. Its activities have contributed greatly to the understanding of influenza epidemiology. The network was established in 1952, currently with 135 institutions from 105 countries recognized by WHO as National Influenza Centers. In addition, various other laboratories have regularly submitted influenza viruses to the programme

in the past years.

Besides WHO's GISN, Canada also has its own national surveillance system called FluWatch, which monitors the spread of influenza and influenza-like illnesses on an on-going basis. FluWatch reports, posted every Friday, contain specific information for health professionals on flu viruses circulating in Canada. The program consists of a network of labs, hospitals, doctor's offices and provincial and territorial ministries of health.

2.3.2 The Alberta Real Time Syndromic Surveillance Net

In Canada, provinces and territories usually have their own surveillance network in addition to the national surveillance. In order to improve public health surveillance of infectious diseases in Edmonton and Area (The former Capital Health region of the Alberta Health Services (AHS)), the Alberta Real Time Syndromic Surveillance Net (ARTSSN) was initiated in 2006 with financial support from the Alberta Health and Wellness (AHW) and in kind contributions from the Information Systems department and the Public Health Division of Capital Health.

The goal of ARTSSN is to enhance public health surveillance through the early (real time) automated detection and tracking of disease outbreaks, environmental hazard exposures and injuries. Moreover, rapid communication of the findings to public health decision-makers, using both syndromic (i.e. pre-diagnostic and pre-laboratory confirmatory) and laboratory electronic data, through a secure intranet-based network, will facilitate more direct and timely action.

ARTSSN data sources include residents of Edmonton and Area only (records on non-residents are excluded) and a common set of patient demographic information (PDI) in all data sources. There are four electronic data sources that are currently employed in real-time (10 minutes intervals, unless stated otherwise) in ARTSSN:

1. **Health Link (HL) Alberta** is a province-wide 365/7/24 telephone health advice service. Its database contains rich information on health concerns reported by the community. ARTSSN has HL data dating back to April 1, 2003. Data fields include: call date and time, PDI, chief complaint (called protocol in HL) and disposition (advice given to patients) under 132 protocols that are relevant to public health.

2. **Emergency Department (ED) Visit Data** are contributed by nine hospitals through the Emergency Department Information System (EDIS) and the E-Triage system. ARTSSN has ED data starting in mid-2004 for all nine hospitals and in 2000 for the Northeast Community Health Center, the University of Alberta Hospital and the Stollery Children's Hospital. Data fields include

visit date and time, PDI, chief complaint, diagnosis, discharge disposition, transfer, and hospital name.

3. **School Absenteeism Data** targets all elementary schools offering kindergarten to grade 6 education in the districts of the Edmonton Public Schools, Edmonton Catholic Schools, and the Parkland School Division #70. The data holdings go back to November 14, 2007 for the Edmonton Public Schools. Data fields in this source include student PDI, absence date, absence duration, school name, and school postal code.

4. **Laboratory Data** are contributed by the Provincial Laboratory for Public Health of Alberta and the laboratories of DynaLifeDX Inc. They contain positive results for notifiable communicable diseases. The data holdings are estimated to go back nearly 15 years. The data fields include PDI, specimen source and type, collection date and time, order date and time, receiving date and time, result date and time, test procedure, test result, organism tested for, antibiotic resistance of the organism and ordering date and time. This data source is still under development.

2.3.3 Current Influenza Prediction

Much work has been devoted to the real-time detection of influenza outbreaks, defined as some increase above a historical baseline threshold. However, a very limited range of approaches has been developed to predict the spread of the epidemic process, especially those that model the epidemic curve.

Modeling of epidemic curves is usually based on time-series approach. Autoregressive seasonal linear models (Box & Jenkins, 1976) have previously been applied to influenza surveillance data by Stroup *et al.* (1988), and Quenel & Dab (1998). Šaltytėbenth & Hofoss (2008) further improve model fit by analyzing the squared residuals after fitting a seasonal time-series model, and show that the squared residuals can reveal the presence of the remaining seasonal variation which is not exhibited by the analysis of residuals.

However, these models did not take into account of the spatial correlations which are quite important in predicting regional influenza activity. As an improvement, Viboud *et al.* (2003) apply the method of analogues to forecast the regional spread of influenza-like illnesses, which is a nonparametric approach first developed by Lorenz (1969) to forecast meteorologic time series. Temporal and spatial correlations are both considered, however, when predicting well ahead of time is attempted, influenza peaks, which are key in prediction, are not well captured. Moreover, due to the computational complexity, this method cannot be easily generalized to researchers in non-mathematical fields.

Therefore, it is important to introduce a relatively simple model which captures both temporal and spatial correlations in influenza prediction, and can accurately predict both non-peaks and peaks of influenza activities well ahead of time.

2.4 Review of Statistical Methods

In this section we review the following topics: natural cubic splines, the generalized linear mixed model (GLMM), pseudo likelihood and cross validation.

2.4.1 Natural Cubic Splines

In mathematics, cubic splines are usually used to capture the non-linear effect of a covariate X using a flexible curve without knowing the shape of the curve. In order to fit a natural cubic spline curve, we need to first decide the number and locations of knots. While the number of knots determines the smoothness of the spline curve, their locations are not as critical as the number. Suppose that $\{(x_k, y_k)\}_{k=0}^{n}$ are $n+1$ knots, where $a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b$. The function $S(x)$ is called a cubic spline if there exists n cubic polynomials $S_k(x)$ with coefficients $s_{k,0}, s_{k,1}, s_{k,2}$, and $s_{k,3}$ that satisfy the following properties:

1. $S(x) = S_k(x) = s_{k,0} + s_{k,1}(x - x_k) + s_{k,2}(x - x_k)^2 + s_{k,3}(x - x_k)^3$ for $x \in [x_k, x_{k+1}]$ and $k = 0, 1, \ldots, n-1$.

2. $S(x_k) = y_k$ for $k = 0, 1, \ldots, n$. i.e. The spline curve passes through each knots.

3. $S_k(x_{k+1}) = S_{k+1}(x_{k+1})$ for $k = 0, 1, \ldots, n-2$. i.e. The spline forms a continuous function over $[a, b]$.

4. $S_k'(x_{k+1}) = S_{k+1}'(x_{k+1})$ for $k = 0, 1, \ldots, n-2$. i.e. The spline forms a smooth function.

5. $S_k''(x_{k+1}) = S_{k+1}''(x_{k+1})$ for $k = 0, 1, \ldots, n-2$. i.e. The second derivative is continuous.

There exists a unique cubic spline with free boundary conditions $S''(a) = 0$ and $S''(b) = 0$, and such splines are called natural cubic splines. In particular, the spline functions are specified as following:

$$\begin{cases} S_0(x) = x, \text{ for } k = 0; \\ S_k(x) = (x - x_{k-1})_+^3 - (x - x_n)_+^3 \frac{(x_{n+1} - x_{k-1})}{x_{n+1} - x_n} + (x - x_{n+1})_+^3 \frac{x_n - x_{k-1}}{x_{n+1} - x_n}, \\ \text{for } k = 1, 2, \ldots, n-1. \end{cases}$$

Here $(A)_+ = \max(0, A)$.

2.4.2 Generalized Linear Mixed Model

Fixed effects models, which assume that all observations are independent of each other, are not appropriate for analysis of several types of correlated data structures, in particular, for clustered and/or longitudinal data. In clustered designs, subjects are observed nested within larger units, e.g. schools, hospitals, neighborhoods, workplaces, etc. In longitudinal designs, repeated observations are nested within subjects. For analysis of such data, random cluster/suject effects can be added into the regression model to account for the correlation of the data. The resulting model is an extension to the generalized linear model in which the linear predictor contains random effects in addition to the usual fixed effects. These random effects are usually assumed to follow normal distributions. This kind of models are called generalized linear mixed models (GLMMs). Denote the response measurement of the j^{th} subject/observation of the i^{th} cluster/subject as Y_{ij}, then a GLMM can be represented as follows:

Systematic Part
$h\left(E\left[Y_{ij}|b_{0i}, b_{1i}, \ldots, b_{qi}\right]\right) = \beta_0 + \beta_1 X_{1ij} + \ldots + \beta_p X_{pij}$
$+b_{0i} + b_{1i} Z_{1ij} + \ldots + b_{qi} Z_{qij},$ where $Z^{'}$s are subsets of $X^{'}$s, $q < p$.

Random Part
$(Y_{ij}|b_{0i}, b_{1i}, \ldots, b_{qi}) \sim$ Normal, Binomial, Poisson, etc.
$(b_{0i}, b_{1i}, \ldots, b_{qi}) \sim MVN\left(\mathbf{0}, \mathbf{\Sigma}\right)$ for all clusters/subjects.

In the above model, $\beta_0 + \beta_1 X_{1ij} + \ldots + \beta_p X_{pij} + b_{0i} + b_{1i} Z_{1ij} + \ldots + b_{qi} Z_{qij}$ is the linear predictor, and h is the link function, which is used to map the value of the linear predictor to the conditional mean of the distribution function of Y_{ij}'s. The link function can take different forms depending on the conditional distribution of the responses. The identity function, logit function, and $\log$ function are canonical links when $(Y_{ij}|b_{0i}, b_{1i}, \ldots, b_{qi})$ follows a normal distribution, a binomial distribution, and a Poisson distribution, respectively.

The GLMM is especially popular when the interest is in estimating subject-specific effects. This kind of models uses other clusters/subjects' information as well as the i^{th} cluster/subject's information to estimate the i^{th} cluster/subject-specific effect. Although small sample size per cluster/subject cannot provide stable estimates on cluster/subject-specific effects, use of the overall average across all subjects can help overcome this problem.

2.4.3 Cross-Validation

Cross-validation is a technique for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. Cross-validation is also a commonly used approach to avoid

model overfitting. One round of cross-validation involves partitioning a sample of data into complementary subsets, performing the analysis on one subset (the training set), and validating the analysis on the other subset (called the validation set or testing set). To reduce variability, multiple rounds of cross-validation are performed using different partitions, and the validation results are averaged over the rounds.

There are four common types of cross-validation: K-fold cross-validation, $k \times 2$ cross-validation, repeated random sub-sampling validation, and leave-one-out cross-validation (LOOCV). In the presented work, the LOOCV is applied where each year's observation are treated as one unit. As the name suggests, LOOCV involves using a single observation from the original sample as the validation data, and the remaining observations as the training data. This is repeated such that each observation in the sample is used once as the validation data. This the same as the K-fold cross-validation with K equal to the number of observations in the original sample.

For correlated data, like the ARTSSN data we used, the validity of the common LOOCV for independent cases is proved by Hart & Vieu (1990), who showed that the LOOCV is reasonably effective unless the data are very highly correlated. In fact, the cross-validation is still asymptotically optimal when they leave out only one unit at each time.

2.4.4 Pseudo-Likelihood

In some estimation problems the joint likelihood involves an analytically-intractable normalizing function of the parameters. In such cases the method of maximum likelihood is not easy to use. An alternative is to replace the joint likelihood by a suitable product of ratios of likelihoods of subsets of the variables. We call these products pseudo-likelihoods, e.g. the product of various conditional densities is a pseudo-likelihood that does not involve the normalizing function. The practical use of it is that it can provide an approximation to the likelihood function of a set of observed data which may either provide a computationally simpler problem for estimation, or may provide a way of obtaining explicit estimates of model parameters. Under regularity conditions, pseudo-likelihood estimators are consistent and asymptotically normal.

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ be i.i.d random variables with probability distribution F_Θ where Θ is a vector of unknown variables. Let the set E be the dependencies between these random variables, where $\{X_i, X_j\} \in E$ implies X_i is conditionally dependent of X_j given $X_j^{'}$s neighbors. Then the pseudo-likelihood of

$\boldsymbol{X} = \boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ is:

$$\mathcal{PL}\left(\Theta|\,\boldsymbol{x}\right) = \prod_i P_\Theta\left(x_i|x_j \text{ for all } j \text{ where } \{X_i, X_j\} \in E\right),$$

More often, we use the $\log$ function of pseudo-likelihood instead, known as the pseudo-log-likelihood:

$$pl = \log \mathcal{PL}\left(\Theta|\,\boldsymbol{x}\right) = \sum_i \log P_\Theta\left(x_i|\,x_j \text{ for all } j \text{ where } \{X_i, X_j\} \in E\right).$$

In particular, consider the special case that X_i fits into the auto-normal scheme, i.e. the conditional distribution of X_i is normal with conditional mean

$$\mathrm{E}\left[X_i|x_j \text{ for all } j \text{ where } \{X_i, X_j\} \in E\right] = \mu_i + \sum_j \beta_{i,j}\left(x_j - \mu_j\right),$$

where $\beta_{i,i} = 0$, and conditional variance

$$\mathrm{Var}\left[X_i|x_j \text{ for all } j \text{ where } \{X_i, X_j\} \in E\right] = \sigma_i^2.$$

This also holds if X_i follows an elliptical distribution, or equivalently an auto-normal scheme after an affine transformation. Besag (1975) shows that the pseudo-log-likelihood can be reduced to the ordinary method of least squares. In fact, the maximum pseudo-likelihood estimates $\hat{\Theta}$ can be obtained by minimizing

$$pl = \sum_i \left\{ x_i - \mu_i - \sum_j \beta_{i,j}\left(x_j - \mu_j\right) \right\}^2,$$

for all j where $\{X_i, X_j\} \in E$.

3 Statistical Modeling and Prediction of Influenza ED Visits

In this section, we introduce our method for modeling and predicting the daily counts of Edmonton ED visits.

3.1 Data Preparation

In this work, we focus on emergency department (ED) visit data recorded in ARTSSN from August 1, 2004 to July 31, 2009 with a total of 1,826 days. The raw data were collected in 10 minutes interval from residents in Edmonton and Area (The former Capital Health region of the AHS). The location of residency was indicated by the first three digits of patients' residence postal codes. Since our analyses concentrate on modeling and prediction of daily ED visits in Edmonton only, we extract subjects from Edmonton according to first three digits postal codes, and aggregate data into daily counts. After the aggregation, there are 1,826 observations (days) for each of the 38 first three digits postal codes in our data. The daily ED visit pattern is represented by the yellow line in Figure 3.1, which shows that, despite daily fluctuations over time, there appears to be a systematic trend throughout the five years.

3.2 Model without Spatial or Temporal Correlation

We start describing the influenza activity in Edmonton by a systematic yearly trend. This trend can be represented with a flexible curve that is the same for every year, and this can be done by fitting a GLMM with daily influenza ED visits as the response variable and natural cubic splines of days during the five years as the explanatory variables.

To capture the annual pattern, we use cubic splines that are identical for every year. Specifically, we code the days as 1 to 365 for each year or 1 to 366 for the leap year, denoted by $\boldsymbol{X}$, i.e. $\boldsymbol{X} = (1, 2, \ldots, 365, 1, 2, \ldots, 365, 1, 2, \ldots, 365, 1, 2 \ldots, 366, 1, 2 \ldots, 365)$. Then for each year, we choose the first days of each week as the knots, resulting in 52 knots and 51 spline functions for the natural cubic splines. The spline functions are denoted by $S^{(k)}(\boldsymbol{X}), k = 0, 1, \ldots, 50$.

Let $Y_j, j = 1, 2, \ldots, 1826$, be the daily count of ED visits, $S_j^{(k)}(\boldsymbol{X}), j = 1, 2, \ldots, 1826, k = 0, 1, \ldots, 50$, be the explanatory variables for the systematic variation represented by the cubic splines of $\boldsymbol{X}$ and are identical across five years, $r_j, j = 1, 2, \ldots, 1826$, be the daily random variations. We have the following GLMM:

$$
\begin{aligned}
&\log\left(\mathrm{E}\left[\,Y_j|\,r_j\right]\right) = \beta_0 + r_j + \beta_1 S_j^{(0)}\left(\boldsymbol{X}\right) + \beta_2 S_j^{(1)}\left(\boldsymbol{X}\right) + \ldots + \beta_{51} S_j^{(50)}\left(\boldsymbol{X}\right), \\
&\left(\,Y_j|\,r_j\right) \sim \text{Poisson}\left(\mathrm{E}\left[\,Y_j|\,r_j\right]\right), \\
&r_j \sim \mathrm{N}\left(0, \sigma^2\right), \quad j = 1, 2, \ldots, 1826.
\end{aligned} \tag{3.1}
$$

This model results in fitting the daily ED visits in Edmonton by an identical flexible curve across five years, and we show the fitted values by the blue line in Figure 3.1.

3.3 Model with Spatial and Temporal Correlations

Figure 3.1 shows that although GLMM extracts the systematic trend of influenza activities during the five years, there are still considerable variations remained unexplained, possibly because the model does not take into account of the spatial or temporal correlations. In particular, the high daily counts of ED visits are not predicted well. In this section, we will introduce an exploratory model which incorporates both spatial and temporal correlations, and is shown to improve the model fitting substantially.

3.3.1 Spatial Decomposition of Daily Variations

In order to incorporate the spatial correlations into our model, we consider the location information in Edmonton represented by the first three digits postal codes. Instead of using the 38 different first three digits postal codes directly, we partition Edmonton into 24 subareas with approximately equal total ED visits throughout the five years. We do so to balance the contributions of the subareas in the model fitting. The partition is shown in Figure 3.2 with different colors indicating different subareas. The daily ED visits are then partitioned into 24 parts accordingly. Denote p_i as the ED visits proportion in the i^{th} subarea, where

$$
p_i = \frac{\text{number of ED visits in } i^{th} \text{ subarea}}{\text{number of ED visits in Edmonton}} \text{ in the five years}, \quad i = 1, 2, \ldots, 24.
$$

Denote Y_{ij} for the observed ED visits in the i^{th} subarea on the j^{th} day, and denote $\hat{Y}_j$ for the estimated Y_j from model (3.1). Then the estimated $\hat{Y}_{ij}$ is:

$$
\hat{Y}_{ij} = p_i \hat{Y}_j, \quad i = 1, 2, \ldots, 24, j = 1, 2, \ldots, 1826. \tag{3.2}
$$

The unexplained variation in the i^{th} subarea on the j^{th} day, R_{ij} (though strictly speaking, it should be $\hat{R}_{ij}$ since it is estimated from the GLMM, but we use R_{ij} for

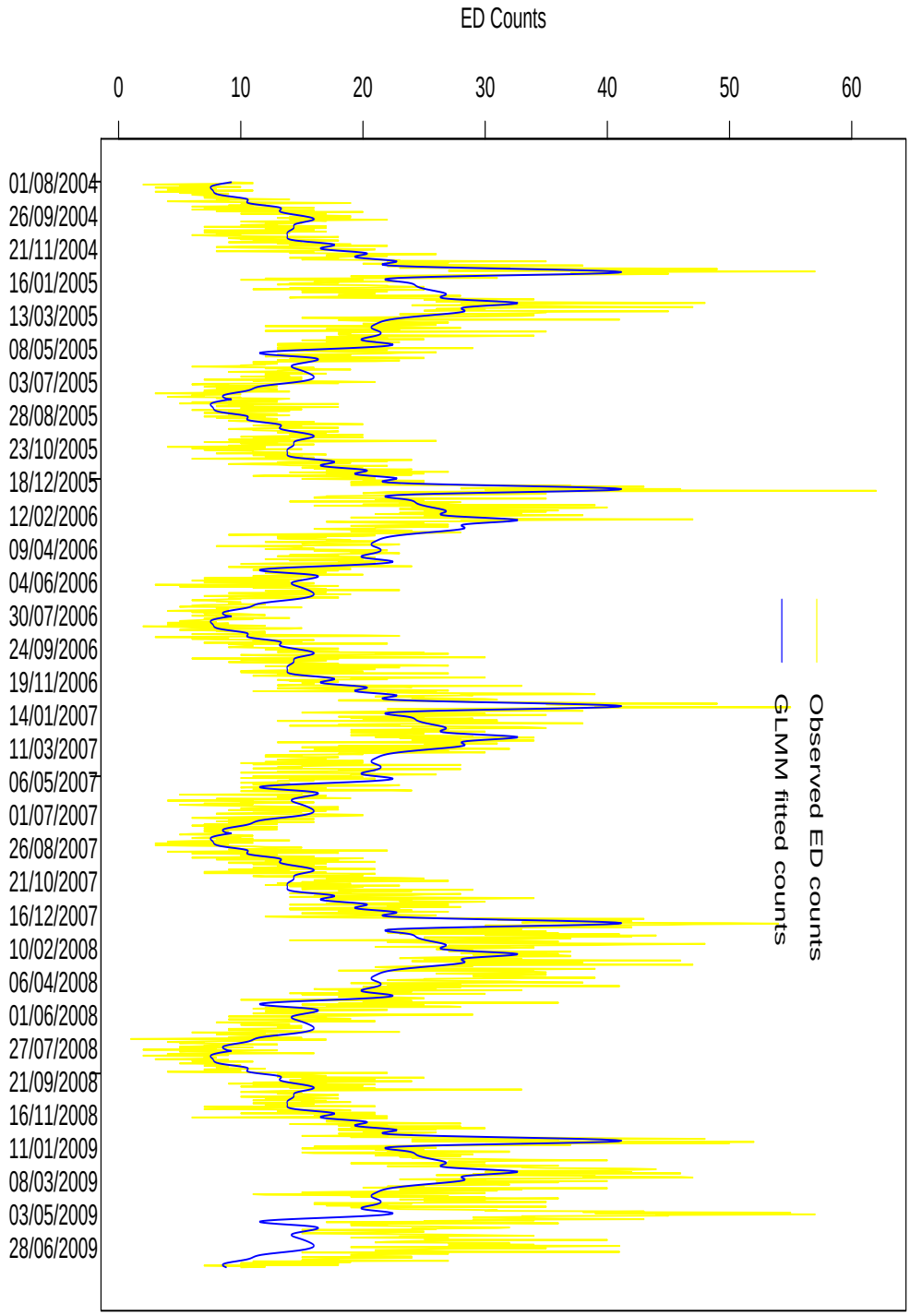

Figure 3.1: Daily ED Counts in Edmonton from Aug. 1, 2004 to July 31, 2009

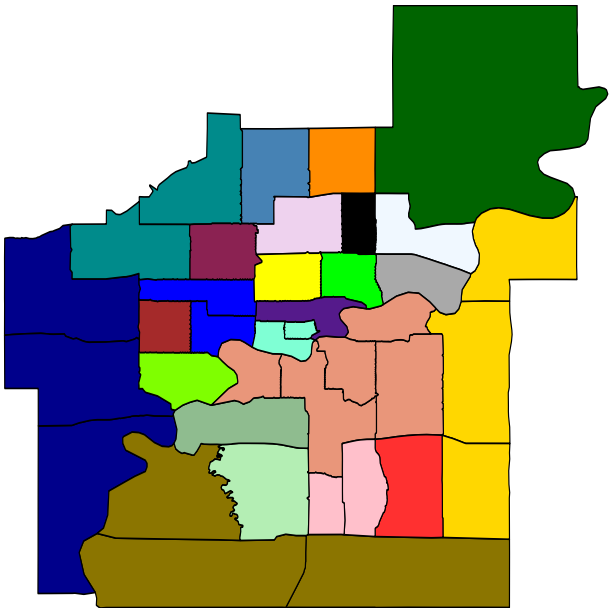

Figure 3.2: 24 Subareas in Edmonton

now for notational purposes), after fitting a GLMM, can be obtained by:

$$R_{ij} = Y_{ij} - \hat{Y}_{ij}, \quad i = 1, 2, \ldots, 24, j = 1, 2, \ldots, 1826. \tag{3.3}$$

R_{ij} can be standardized within each subarea, which is shown below:

$$R^*_{ij} = \big(R_{ij} - \overline{R}_{i\cdot}\big) / \, SD[R_{i\cdot}], \tag{3.4}$$

where $\overline{R}_{i\cdot}$ is the mean of daily variations in the i^{th} subarea across the five years, and $SD[R_{i\cdot}]$ is the standard deviation of R_{ij}'s.

3.3.2 Exploratory Modeling

It is reasonable to assume that the unexplained standardized daily variations R^*_{ij}'s are correlated with each other through both time and spatial locations. This motivates us to consider that R^*_{ij}'s can be modeled with daily variations $R^*_{i'j'}$'s of all its spatial proximities prior to the j^{th} day, and we represent it with an exploratory modeling in which every R^*_{ij} is linear in such $R^*_{i'j'}$'s with weights $w_{i'j'}$'s. The spatial proximities are represented by physical neighbors which are defined as all subareas that share at least some border with the i^{th} subarea. Denote the set of i^{th} subarea's

neighbors as $\mathcal{N}(i)$. The exploratory model is shown below:

$$\begin{aligned} &\mathrm{E}\left[R^*_{ij}\middle|\, R^*_{i'j'}\text{'s}\right] = \sum_{i'\in\{i,\mathcal{N}(i)\}}^{j'<j} w_{i'j'} R^*_{i'j'} \big/ N_i, \\ &i = 1, 2, \ldots, 24, j = 1, 2, \ldots, 1826, \end{aligned} \tag{3.5}$$

where $\{i, \mathcal{N}(i)\}$ is the set of $\mathcal{N}(i)$ and i^{th} subarea itself, and N_i is 1 plus the number of elements in $\mathcal{N}(i)$. Here we divide each component by N_i in order to standardize, since $\mathcal{N}(i)$ can have different sizes.

3.3.3 Parameter Estimation

For an exploratory model like ours, the common approach for parameter estimation, such as maximum likelihood estimation, is not plausible as there is no joint or marginal distribution assumption in the model. Thus, we need to utilize another simple estimation approach: the pseudo-likelihood method, which is introduced in Section 2.4.4. The pseudo-likelihood is defined as follows:

$$\begin{aligned} &pl\left(w_{ij}\middle|\, R^*_{ij}\text{'s}\right) = \sum_{i,j} \left(R^*_{ij} - E\left[R^*_{ij}\middle|\, R^*_{i'j'}\text{'s}\right]\right)^2, \\ &i = 1, 2, \ldots, 24, j = 1, 2, \ldots, 1826, i' \in \{i, \mathcal{N}(i)\}, j' < j. \end{aligned} \tag{3.6}$$

Then, according to model (3.5), the pseudo likelihood (3.6) becomes:

$$\begin{aligned} &pl\left(w_{ij}\middle|\, R^*_{ij}\text{'s}\right) = \sum_{i,j} \left(R^*_{ij} - \sum_{i'\in\{i,\mathcal{N}(i)\}}^{j'<j} w_{i'j'} R^*_{i'j'} \big/ N_i\right)^2, \\ &i = 1, 2, \ldots, 24, j = 1, 2, \ldots, 1826, i' \in \{i, \mathcal{N}(i)\}, j' < j. \end{aligned} \tag{3.7}$$

Consider, for example, we want to know the influenza activities in Edmonton emergency departments one week ahead so that hospital staff and policy makers can get prepared for necessary facilities. Furthermore, we may want to make predictions with records of ED visits from a month ago up to the current day. To state this scenario in general, that is, we are interested in predicting P days ahead ED influenza visits with L days of ED visits records. Using the pseudo likelihood (3.7), this general prediction scenario is equivalent to minimize (3.7) where $j' = j - P - L, j - P - L + 1, \ldots, j - P - 1$.

In order to estimate the parameters, we re-parameterize $w_{i'j'}$'s into α_l's:

$$w_{i'j'} = \begin{cases} \alpha_{2l-1} & i' = i, \quad j' = j - P - l \\ \alpha_{2l} & i' \in \mathcal{N}(i), \quad j' = j - P - l \end{cases}, l = 1, 2, \ldots, L. \tag{3.8}$$

Then the exploratory model (3.5) becomes:

$$\mathrm{E}\left[R^*_{ij} \middle|\, R^*_{i'j'}\text{'s} \right] = \sum_{i',j',l} \alpha_l R^*_{i'j'} / \, N_i, \tag{3.9}$$

where $i = 1, 2, \ldots, 24$, $j = 1, 2, \ldots, 1826$, $i' \in \{i, \mathcal{N}(i)\}$, $j' = j - P - L, j - P - L + 1, \ldots, j - P - 1$, and $l = 1, 2, \ldots, 2L$. With the same re-parameterizations, the pseudo likelihood (3.7) is expressed as:

$$pl\left(\alpha_l |\, R^*_{ij}\text{'s} \right) = \sum_{i,j} \left(R^*_{ij} - \sum_{i',j',l} \alpha_l R^*_{i'j'} / \, N_i \right)^2, \tag{3.10}$$

where i, j, i', j', and l are defined same as equation (3.9). The estimates $\hat{\alpha}_l$'s are those that minimize the pseudo likelihood (3.10), i.e.

$$\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_{2L})' = \underset{\boldsymbol{\alpha}}{\text{argmin}} \quad pl\left(\alpha_l |\, R^*_{ij}\text{'s} \right). \tag{3.11}$$

3.4 Prediction of ED Influenza Visits

This section describes the method to predict ED influenza visits and provides the prediction results.

3.4.1 Prediction with Equal Weights

To unbiasedly estimate the prediction performance of a model in the absence of an independent validation set, we use five-fold cross-validation where each year is considered as a fold. That is, each time we exclude one year's ED visits data, and perform parameter estimation using the remaining four years' data. The predicted ED visits of the excluded year are then obtained using the estimated parameters. The detailed process is described below:

1. Exclude the first year's data, applying Equations (3.10) and (3.11) to perform parameter estimation ($\hat{\alpha}_l$'s) on the remaining four years' data.

2. Plug $\hat{\alpha}_l$'s into the re-parameterized exploratory model (3.9) on the first year's data to obtain $\hat{R}^*_{ij}$'s.

3. Obtain the estimated unstandardized daily variation in i^{th} subarea, $\hat{R}_{ij}$, through Equation (3.4).

4. Predict $\tilde{Y}_{ij}$ for the first year through Equation (3.3), where $\tilde{Y}_{ij} = \hat{Y}_{ij} + \hat{R}_{ij}$.

5. The predicted daily ED influenza visits $\tilde{Y}_j = \sum_i \tilde{Y}_{ij}$.

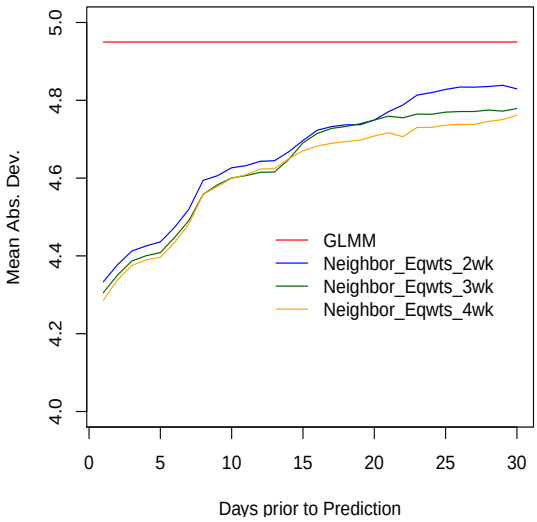

Figure 3.3: Comparison of MSE: GLMM vs. Equally Weighted Approach
All Observations

Repeat the process on the second year to the fifth year respectively to get all five years' predicted ED influenza visits.

Using model (3.9) and pseudo likelihood (3.10), we implement one-day prior to 30-day prior predictions using two-week, three-week, and four-week ED visits data, i.e. P ranges from 1 to 30, and $L = 14, 21$ and 28. The prediction performance is evaluated by a quantity called mean absolute deviation (MAD), which is the mean of the absolute distances between the observed and predicted daily ED visits:

$$\text{MAD} = \underset{j}{\text{mean}} \left| Y_j - \tilde{Y}_j \right|, \quad j = P + L + 1, P + L + 2, \ldots, 1826. \tag{3.12}$$

Figure 3.3 shows the prediction performance from Model (3.9), compared to the GLMM model (3.1). It shows that by taking both temporal and spatial correlations into account, the exploratory model (3.9) improves the prediction substantially compared to applying the GLMM (3.1) alone. The MAD decreases from about 5 in GLMM to a range of 4.3 to 4.8 for one-day prior to 30-days prior predictions, with only a few weeks information. To better illustrate, we also provide Figure 3.4 to show the fitted results from models (3.1) and (3.9). In particular, we only show the results from one-week prior prediction with four-week information for illustration, represented by the light green line in Figure 3.4.

As shown from Figure 3.4, the exploratory model improves the prediction substantially. However, the prediction performance is not quite satisfactory for days

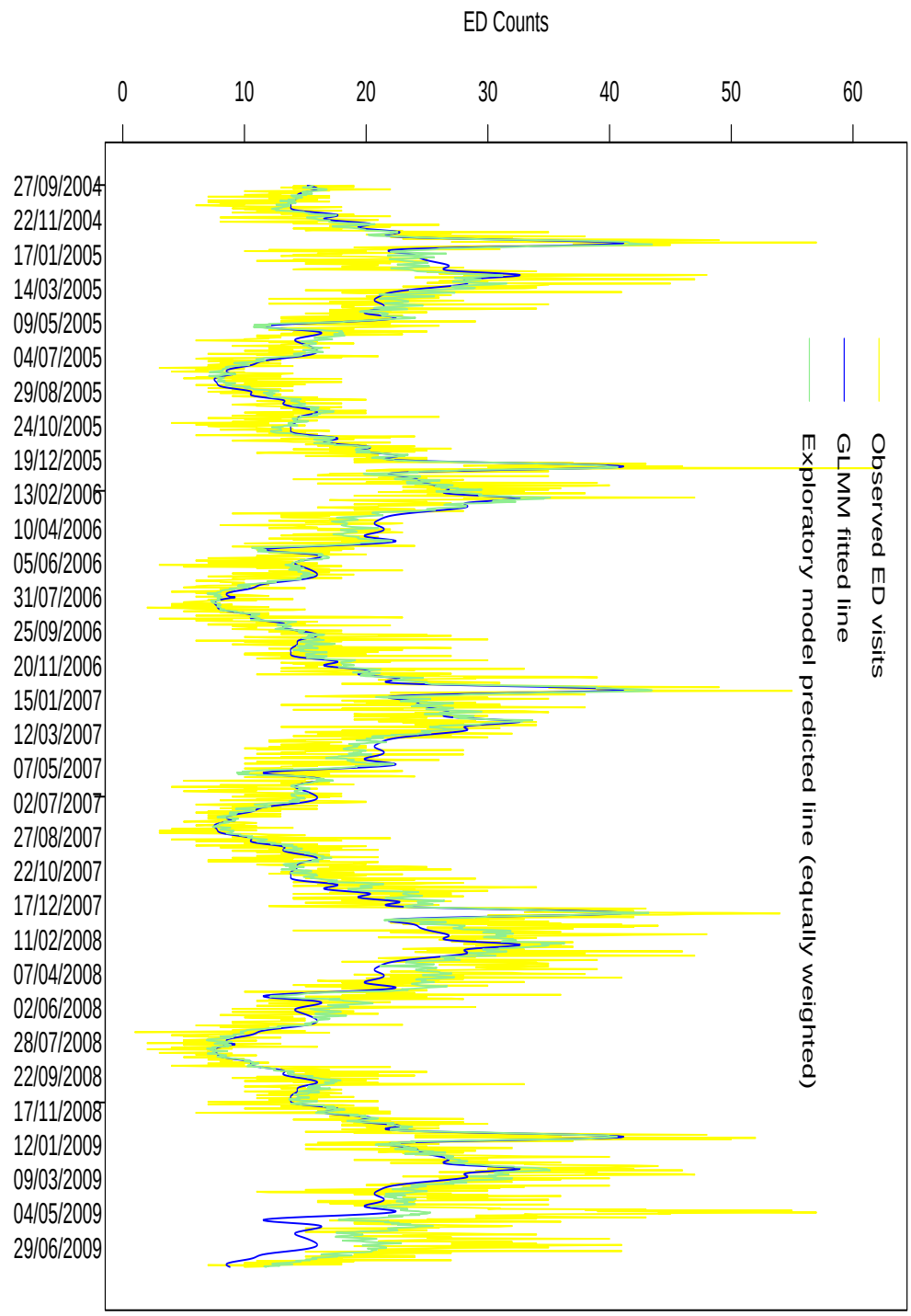

Figure 3.4: Predicted ED Counts in Edmonton: GLMM vs. Exploratory Model from Sep. 27, 2004 to July 31, 2009

with higher daily influenza counts. In fact, it tends to underestimate the ED influenza visits when there are epidemics. This is unacceptable for policy makers and health departments, since the underestimation would lead them to be unprepared when influenza activity is rampant. To better illustrate, we provide Figure 3.5 which shows the MADs of those daily influenza visits that are greater than the overall median, as only influenza activities that are more frequent than usual draw our special attention. That is, we obtain the median of daily influenza visits during the five years, and select only those daily visits that are greater than the median to calculate MADs.

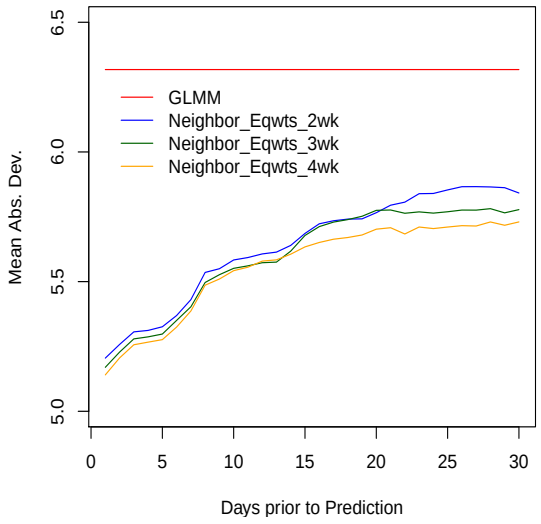

Figure 3.5: Comparison of MAD: GLMM vs. Equally Weighted Approach
Observations Greater than the Overall Median

Figure 3.5 shows that MADs of large daily visits are much greater than those of all visits, which confirms what we observe from Figure 3.4, that influenza peaks are not well predicted compared to non peaks. In order to solve this problem, we propose a modified method with unequal weights which is introduced in the next section.

3.4.2 Prediction with Unequal Weights

We introduce a modified method of using the exploratory model (3.5) by incorporating unequal weights into the pseudo likelihood estimation (3.10). Specifically, apply higher weights to the days with larger daily ED visits and lower weights to those with fewer visits, so that days with larger daily ED visits contribute more in the estimation and thus can be predicted better. The weighted pseudo likelihood is

expressed as:

$$pl\left(\alpha_l|\, R^*_{ij}\text{'s}\right) = \sum_{i,j} \left(Y_{ij} + 0.5\right)^s \left(R^*_{ij} - \sum_{i',j',l} \alpha_l R^*_{i'j'} \big/ N_i\right)^2, \tag{3.13}$$

where i, j, i', j', l are defined the same way as Equation (3.10), and s can be any positive integers. In our analyzes, we use s equal to positive integers to find the best weights. Here we add 0.5 to the actual daily counts Y_{ij} to avoid zero weights. $\hat{\alpha}_l$'s are obtained through equation (3.11).

After comparing the prediction performance with s equal to positive integers, we find that the prediction is best when $s = 3$. Thus, we use $s = 3$ in Equation (3.13) hereafter. However, although the influenza peaks are well predicted via inserting proportional weights into the pseudo likelihood, daily visits that are rare are poorly predicted compared to the equally weighted approach. Therefore, we combine the equally and unequally weighted approaches by choosing each predicted daily visit that is closer to the observed daily visit. Denote the equally weighted daily influenza prediction as E_j, the unequally weighted prediction as U_j, and the combined prediction as C_j, then

$$C_j = \max\left(E_j, U_j\right), \quad j = 1, 2, ..., 1826 - P - L.$$

We plot the MADs of predictions above median for both the equally weighted approach and the combined approach, which is shown in Figure 3.6. The solid lines are the equally weighted one-week, two-week, and three-week prior predictions introduced in Section 3.4.1, with two-week, three-week, and four-week daily ED visits records, respectively; the dashed lines are those predictions that combine the equally and unequally weighted approaches. However, from Figure 3.6, the combined approach does not improve the prediction for above median observations, in fact, the MADs of the combined approach are even a little larger than those of the equally weighted approach. This is because the proportionally weighted approach concentrates on accurately predicting high daily observations, i.e., the peaks, while it loosens the predictions on the non-peaks. Therefore, if we compare the MADs of daily observations that are greater than the 75%ile, we expect to see an improvement in the prediction by the combined approach, compared to the equally weighted approach. This is confirmed by Figure 3.7.

Figure 3.8 shows the one-week prior predicted line from the combined approach, using four weeks records of daily ED visits. In comparison, we also show the GLMM fitted line and the predicted line using equally weighted pseudo likelihood. The plot again shows that the combined approach predicts daily influenza visits well, with nice performance on both the peaks and non-peaks.

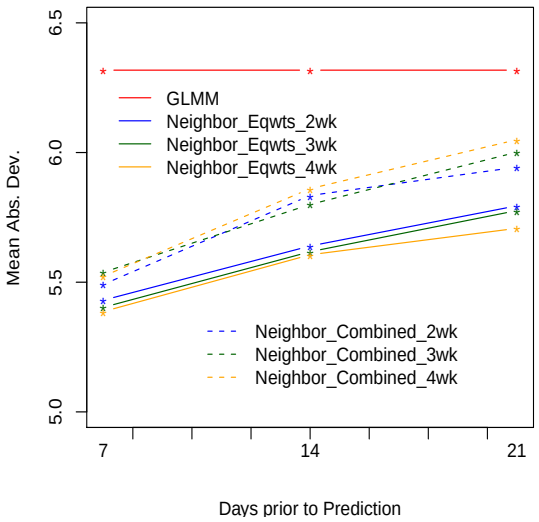

Figure 3.6: Comparison of MAD: Equally Weighted vs. Combined Approaches Observations Greater than the Overall Median

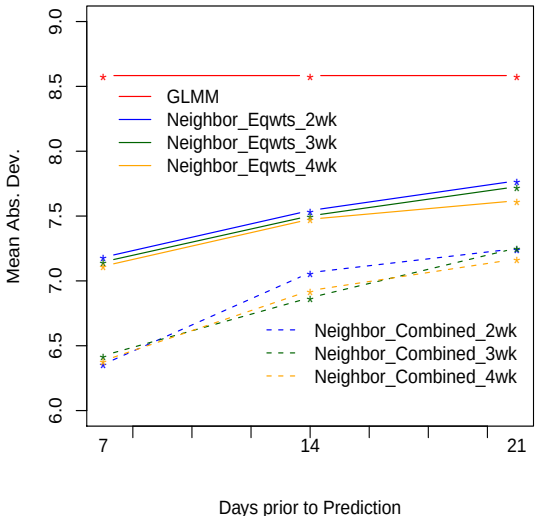

Figure 3.7: Comparison of MAD: Equally Weighted vs. Combined Approaches Observations Greater than the Overall 75%ile

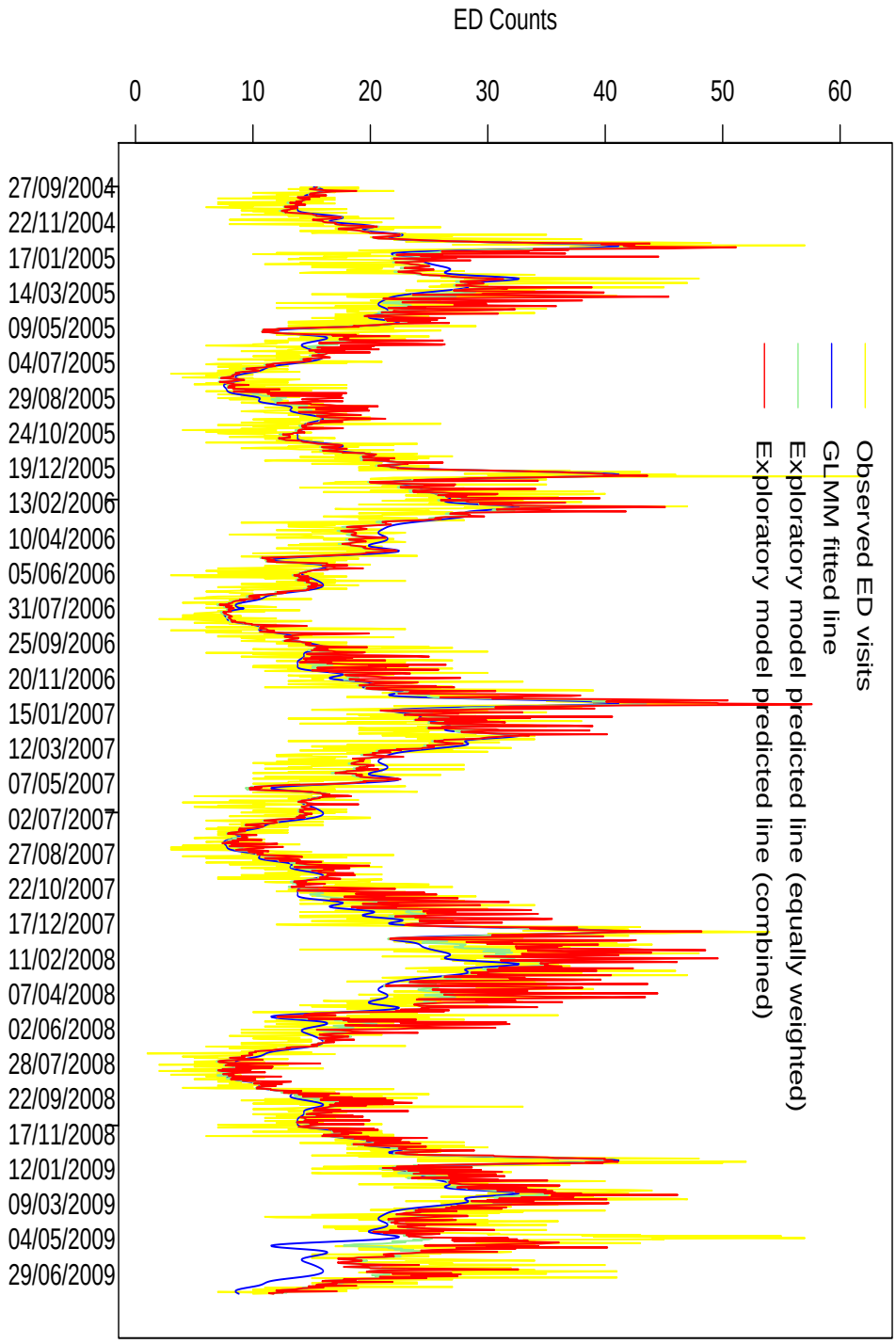

Figure 3.8: Prediction Comparison of Equal Weights vs. Combined Approach from Sep. 27, 2004 to July 31, 2009

4 Symmetric Multivariate Pearson Type VII Family and Applications

A characterization of the class of elliptical symmetric distributions, under the finite second moments assumption, is that all regressions are linear (Nimmo-Smith, 1979). Within the elliptical symmetric class, all regressions are homoscedastic if and only if the distribution is Gaussian (Kelker, 1970). Given the substantial roles the Gaussian distribution plays in multivariate statistics, it is of considerable interest to extend the family by relaxing the homoscedasticity of regressions while retaining their linearity, as one of the features of Gaussian distributions that one may regard as questionable is the constant conditional second moments. That is, realizations of the conditioning variables do not influence the magnitudes of the conditional second moments, thus may not represent real stochastic processes.

This section introduces the Multivariate Pearson Type VII (MP VII) family, a family of multivariate distributions which has the same useful feature that all regressions are linear as Gaussian distributions, and yet its conditional second moments are not constant like Gaussian. In particular, they are linear in the empirical second moment of the entire conditioning vector. The development is based on the theory of elliptical distributions (Kelker 1970; Cambanis, Huang & Simmons, 1981; Fang, Kotz & Ng, 1990; Fang & Zhang, 1990).

4.1 Introduction to MP VII Family

We first introduce the elliptical family to which the MP VII family belong, followed by results on the conditional moments and definition of MP VII family.

4.1.1 Elliptical Distribution

Suppose a $n \times 1$ random vector $\boldsymbol{Y} \in \mathcal{R}^n$ has finite second moments and the support that does not lie in any proper linear subspace. Nimmo-Smith (1979) proved that $\boldsymbol{Y}$ follows an elliptical distribution, or equivalently a spherical distribution after an affine transformation, if and only if all regressions are linear. Cambanis, Huang & Simmons (1981) showed that any elliptical distribution can be expressed uniquely by the following stochastic representation. Let $\boldsymbol{U}^{(n)}$ denote a random vector distributed uniformly on the surface of the unit sphere in $\mathcal{R}^n$. Let R be the positive generating variate of $\boldsymbol{Y}$ and is independent of $\boldsymbol{U}^{(n)}$, with the density of $T = R^2$ given by

$$\frac{\pi^{n/2}}{\Gamma\left(n/2\right)} t^{n/2-1} g\left(t\right),$$

where g satisfies

$$\frac{\pi^{n/2}}{\Gamma(n/2)} \int_0^\infty u^{n/2-1} g(u)\, du = 1. \tag{4.1}$$

Then the distribution of $\boldsymbol{Y}$ is an elliptical distribution with mean $\boldsymbol{\mu}$, positive definite characteristic matrix $\boldsymbol{\Sigma}$, and density generator g, if and only if

$$\boldsymbol{Y} \overset{d}{=} \boldsymbol{\mu} + R\boldsymbol{\Sigma}^{1/2}\boldsymbol{U}^{(n)}, \tag{4.2}$$

where the sign $\overset{d}{=}$ means the random vectors on each side of the sign has the same distribution, and $\boldsymbol{\Sigma}^{1/2}$ is the square root matrix of $\boldsymbol{\Sigma}$. The density of $\mathbf{Y}$ is

$$|\boldsymbol{\Sigma}|^{-1/2} g\left(\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right).$$

We use $\boldsymbol{Y} \sim EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ to denote this distribution. For instance, if R^2 has a χ^2 distribution with n degrees of freedom, the distribution of $\boldsymbol{Y}$ is multivariate normal (MVN) with $g(t) = e^{-t/2}$.

4.1.2 Conditional Second Moments

Consider the partition

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \tag{4.3}$$

where $\boldsymbol{Y}_1$ and $\boldsymbol{\mu}_1$ are $m \times 1$, $\boldsymbol{\Sigma}_{11}$ is $m \times m$, and $1 \leq m < n$. The conditional expectations of elliptical distributions are determined entirely by $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ and are independent of g (Kelker, 1970):

$$\mathrm{E}[\boldsymbol{Y}_1 | \boldsymbol{Y}_2 = \boldsymbol{y}_2] = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{y}_2 - \boldsymbol{\mu}_2), \tag{4.4}$$

The conditional second moments are given by Szablowski (1990) as below:

$$\mathrm{Cov}[\boldsymbol{Y}_1 | \boldsymbol{Y}_2 = \boldsymbol{y}_2] = \frac{\int_{a/2}^\infty g_{n-m}^*(2u)\, du}{g_{n-m}^*(a)} \boldsymbol{\Sigma}_{11.2}, \tag{4.5}$$

where

$$g_k^*(x) = \frac{\pi^{(n-k)/2}}{\Gamma((n-k)/2)} \int_x^\infty (u-x)^{(n-k)/2-1} g(u)\, du, \tag{4.6}$$

$$a = (\boldsymbol{y}_2 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{y}_2 - \boldsymbol{\mu}_2), \tag{4.7}$$

and

$$\boldsymbol{\Sigma}_{11.2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}. \tag{4.8}$$

4.1.3 Symmetric Multivariate Pearson Type VII Distribution

When the density of $T = R^2$ is found to be

$$\frac{1}{\lambda B\left(n/2, N - n/2\right)}\left(t/\lambda\right)^{n/2-1}\left(1 + t/\lambda\right)^{-N}, \quad t > 0,$$

i.e., R^2/λ has a beta type II distribution with parameters $n/2$ and $N - n/2$, then $\boldsymbol{Y}$ in (4.2) follows the MP VII distribution (Fang, Kotz & Ng, 1990), and,

$$g\left(t\right) = \frac{\Gamma\left(N\right)}{\Gamma\left(N - n/2\right)}\left(\pi\lambda\right)^{-n/2}\left(1 + t/\lambda\right)^{-N}. \tag{4.9}$$

REMARK A random variable B is said to have a beta type II distribution with parameters α and β if B has density

$$\frac{1}{B\left(\alpha, \beta\right)} b^{\alpha-1}\left(1 + b\right)^{-(\alpha+\beta)}, \quad b > 0,$$

and we shall denote it by $B \sim \text{BeII}\left(\alpha, \beta\right)$.

A special and commonly used distribution in MP VII family is the multivariate t-distribution (MVt), denoted as MV$t\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \lambda\right)$. It satisfies when $N = (n + \lambda)/2$ and λ is an integer.

4.2 Property of MP VII Distribution

Here we state a characterization of the MP VII family as a theorem with proof.

THEOREM Suppose $\boldsymbol{Y} \sim \text{EC}_n\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g\right)$ and $\boldsymbol{Y}$ possesses finite second moments. Consider the partition (4.3) and a defined in (4.7). The conditional second moments, Cov $[\boldsymbol{Y}_1|\boldsymbol{Y}_2 = \boldsymbol{y}_2]$, are linear in a if and only if the distribution of $\boldsymbol{Y}$ belongs to the MP VII family.

Proof Consider the sufficient part first. Apply the Liouville's integral transformation (Samko, Kilbas & Marichev, 1993)

$$I^{\alpha} f\left(x\right) = \int_x^{\infty} f\left(t\right)\left(t - x\right)^{\alpha-1} dt \Big/ \Gamma\left(\alpha\right), \tag{4.10}$$

to the right side of equation (4.6), then

$$g_k^*(x) = \pi^{(n-k)/2} I^{(n-k)/2} g(x). \tag{4.11}$$

Note that equation (4.10) has the following property

$$\frac{d}{dx}\left[I^{\alpha+1} f(x)\right] = -I^{\alpha} f(x), \tag{4.12}$$

thus the conditional second moments (4.5) have the following expression:

$$\begin{aligned}
\mathrm{Cov}\left[\boldsymbol{Y}_1|\boldsymbol{Y}_2 = \boldsymbol{y}_2\right] &= \frac{\int_a^\infty I^{m/2} g(2u)\, du}{I^{m/2} g(a)} \boldsymbol{\Sigma}_{11.2} \\
&= \frac{\frac{1}{2} I^{m/2+1} g(a)}{I^{m/2} g(a)} \boldsymbol{\Sigma}_{11.2} \\
&= -\frac{\frac{1}{2} I^{m/2+1} g(a)}{\frac{d}{da} I^{m/2+1} g(a)} \boldsymbol{\Sigma}_{11.2}.
\end{aligned}$$

Therefore, $\mathrm{Cov}\left[\boldsymbol{Y}_1|\boldsymbol{Y}_2 = \boldsymbol{y}_2\right]$ is linear in a if and only if $I^{m/2+1} g(a) \big/ \frac{d}{da} I^{m/2+1} g(a) = Aa + B$ with some constants A and B. By solving this ordinary differential equation, we obtain

$$I^{m/2+1} g(a) = C(Aa + B)^{-1/A},$$

where A, B, and C are constants. Differentiating both sides l times, where l is the largest integer which satisfies $l \leq m/2 + 1$, we have

$$(-1)^l I^{\frac{m}{2}-l+1} g(a) = (-1)^l C A^l \Gamma(1/A + l)(Aa + B)^{-1/A-l} \Big/ \Gamma(1/A).$$

Using Table 9.3 of Samko, Kilbas & Marichev (1993), we obtain the solution

$$\begin{aligned}
g(a) &= C A^{m/2+1} \Gamma(1/A + m/2 + 1)(Aa + B)^{-1/A-m/2-1} \\
&\triangleq C^* (1 + a/\lambda)^{-N},
\end{aligned}$$

where C^*, λ, and N are constants. The function g, however, has to satisfy (4.1) because the left side of the equation is the integral of a density over the support. Therefore,

$$g(a) = \frac{\Gamma(N)}{\Gamma(N - n/2)} (\pi\lambda)^{-n/2} (1 + a/\lambda)^{-N},$$

which is the density generator (4.9) of MP VII family.

Now consider the necessary part. If $\mathbf{Y}$ belongs to MP VII family, the density

generator $g(t)$ has the form in (4.9). Then, according to equation (4.6),

$$g^*_{n-m}(x) = \frac{\pi^{-(n-m)/2}\lambda^{-n/2}\Gamma(N)}{\Gamma(m/2)\,\Gamma(N-n/2)} \int_x^\infty (u-x)^{m/2-1}(1+u/\lambda)^{-N}\,dt.$$

Use integration by parts and repeat $(m/2-1)$ times, we obtain

$$g^*_{n-m}(x) = \frac{\Gamma(N-m/2)}{\Gamma(N-n/2)}(\pi\lambda)^{-(n-m)/2}(1+x/\lambda)^{-N+m/2}.$$

Then

$$\begin{aligned}
&\int_{a/2}^\infty g^*_{n-m}(2u)\,du \\
&= \frac{\Gamma(N-m/2)}{\Gamma(N-n/2)} \cdot \frac{\lambda/2}{N-m/2-1} \cdot (\pi\lambda)^{-(n-m)/2} \cdot (1+a/\lambda)^{-N+m/2+1} \\
&= g^*_{n-m}(a)\,\frac{a+\lambda}{2N-m-2}.
\end{aligned}$$

According to equation (4.5), we obtain the expression of the conditional second moment for MP VII distribution:

$$\mathrm{Cov}\left[\boldsymbol{Y}_1|\boldsymbol{Y}_2=\boldsymbol{y}_2\right] = \frac{a+\lambda}{2N-m-2}\boldsymbol{\Sigma}_{11.2}, \tag{4.13}$$

which is linear in a. This completes the proof.

4.3 Conditional Prediction Interval for MVt Distribution

We present a scenario where the conditional prediction interval is of interest, and show that if we erroneously assume MVN distribution when the underlying distribution indeed belongs to the MP VII family, the coverage probability is not nominal because the conditional second moments behave differently between the two distributions.

Consider n independent random vectors $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_n$ which are $i.i.d.$ MP VII distributed, where each $\boldsymbol{Y}_i$, $i = 1, \ldots, n$, are $n \times 1$ vectors. We are interested in constructing a prediction interval for Y_{ij} given other components in $\boldsymbol{Y}_i$ and $\boldsymbol{Y}_j^{'}s$, where $j = 1, 2, \ldots, i-1, i+1, \ldots, n$, i.e. the conditional prediction interval of $Y_{ij|-ij}$, where $Y_{ij|-ij}$ denotes Y_{ij} conditioning on Y_{-ij}, and Y_{-ij} denotes all components from the k vectors except the j^{th} component in $\boldsymbol{Y}_i$. This prediction interval is more precise than the unconditional intervals, since it incorporates the information from the known components of the same vector.

4.3.1 Multivariate t-Distribution

For computational ease, we use a special distribution within the MP VII family, called multivariate t-distribution (MVt), to illustrate the conditional prediction interval application afterwards. It satisfies when $N = (n + \lambda)/2$ and λ is an integer in equation (4.9). Denote $\boldsymbol{Y} \sim \text{MV}t_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \lambda)$, where λ is a scaler which is the degrees of freedom. The stochastic representation of MVt distribution is:

$$\boldsymbol{Y} = \boldsymbol{\mu} + \lambda^{1/2}\boldsymbol{Z} \big/ s, \tag{4.14}$$

where s follows Chi distribution with degrees of freedom λ, and $\boldsymbol{Z} \sim \text{MVN}_n(\boldsymbol{0}, \boldsymbol{\Sigma})$.

The univariate case of the MVt distribution is the student t-distribution, which satisfies when $\boldsymbol{\mu}$ and $\boldsymbol{Z}$ in equation (4.14) both reduce to scalers.

4.3.2 Generalized Multivariate t-Distribution

In order to construct the correct MVt-based prediction interval of $Y_{ij|-ij}$ and show that the coverage probability of MVN-based prediction interval is incorrect, we need to utilize some important properties of a generalized version of the MVt distribution, the so called generalized multivariate t-distribution (GMVt) (Dickey, 1967).

The stochastic representation of the GMVt distribution differs slightly from that of the MVt distribution:

$$\boldsymbol{Y} = \boldsymbol{\mu} + \lambda^{1/2}\boldsymbol{Z} \big/ s^*, \tag{4.15}$$

where the only difference from equation (4.14) is that s^* follows Chi distribution with degrees of freedom ν, which adds an additional degrees of freedom to the distribution. Denote the distribution of $\boldsymbol{Y}$ which satisfies equation (4.15) as $\boldsymbol{Y} \sim \text{GMV}t_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \lambda, \nu)$.

When $\lambda = \nu$, it follows that $\text{GMV}t_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \lambda, \lambda) = \text{MV}t_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \lambda)$, where the latter is the MVt distribution discussed in Section 4.1.

We list some important properties of the generalized multivariate t-distribution according to Arellano-Valle & Bolfarine (1995) as follows.

1. Let $\boldsymbol{X} \sim \text{GMV}t_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \lambda, \nu)$. Then,$\text{E}[\boldsymbol{X}] = \boldsymbol{\mu}$ and $\text{Cov}[\boldsymbol{X}] = (\lambda/(\nu - 2))\boldsymbol{\Sigma}$, for $\lambda > 1$ and $\nu > 2$, respectively.

2. Let $\boldsymbol{X} \sim \text{GMV}t_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \lambda, \nu)$ and consider partition (4.5), we have

   (a) $\boldsymbol{X}_1 \sim \text{GMV}t_m(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \lambda, \nu)$ and $\boldsymbol{X}_2 \sim \text{GMV}t_{n-m}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}, \lambda, \nu)$, and,

(b) $(\boldsymbol{X}_1|\boldsymbol{X}_2=\boldsymbol{x}_2)\sim \text{GMV}t_m\left(\boldsymbol{\mu}_{1|2},\boldsymbol{\Sigma}_{11.2},\lambda_{q(\boldsymbol{x}_2)},\nu_m\right)$ where, for $\boldsymbol{\Sigma}>0$,

$$\boldsymbol{\mu}_{1|2}=\boldsymbol{\mu}_1+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\left(\boldsymbol{x}_2-\boldsymbol{\mu}_2\right),\quad \boldsymbol{\Sigma}_{11.2}=\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21},$$
$$q\left(\boldsymbol{x}_2\right)=\left(\boldsymbol{x}_2-\boldsymbol{\mu}_2\right)'\boldsymbol{\Sigma}_{22}^{-1}\left(\boldsymbol{x}_2-\boldsymbol{\mu}_2\right),$$
$$\lambda_a=\lambda+a\quad\text{and}\quad\nu_m=\nu+p-1.$$

3. Let $\boldsymbol{X}\sim\text{GMV}t_n\left(\mathbf{0},\mathbf{I}_n,\lambda,\nu\right)$, where $\mathbf{I}_n$ is the n-dimensional identity matrix. Let $\mathbf{A}$ be a $n\times n$ symmetric matrix. Then we have that $\boldsymbol{X}'\mathbf{A}\boldsymbol{X}\sim\lambda\text{BeII}\left(m/2,\nu/2\right)$ (or, equivalently, $\left(m\lambda/\nu\right)F_{m,\nu}$) if and only if $\mathbf{A}^2=\mathbf{A}$ and $\text{rank}\left(\mathbf{A}\right)=m$, $1\leq m\leq n$.

4.3.3 Construction of Conditional Prediction Interval

Now we proceed to the construction of the conditional prediction interval with MVt as the underlying distribution. For simple notations, consider instead the prediction interval of $Y_{11|-11}$, and the following partition:

$$\boldsymbol{Y}_1=\begin{pmatrix}Y_{11}\\ \boldsymbol{Y}_{12}\end{pmatrix},\boldsymbol{\mu}=\begin{pmatrix}\mu_1\\ \boldsymbol{\mu}_2\end{pmatrix},\boldsymbol{\Sigma}=\begin{pmatrix}\sigma_{11}&\boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21}&\boldsymbol{\Sigma}_{22}\end{pmatrix}. \tag{4.16}$$

According to previous properties (a), (b) and (c), we obtain the following results.

(a') $Y_{11|-11}\sim\text{GMV}t_1\left(\mu_{1|2},\sigma_{11.2},\lambda_{q(\boldsymbol{y}_{12})},\lambda_1\right)$, where

$$\begin{aligned}&\mu_{1|2}=\mu_1+\Sigma_{12}\Sigma_{22}^{-1}\left(\boldsymbol{y}_{12}-\boldsymbol{\mu}_2\right),\quad\sigma_{11.2}=\sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21},\\ &\lambda_{q(\boldsymbol{y}_{12})}=\lambda+\left(\boldsymbol{y}_{12}-\boldsymbol{\mu}_2\right)'\Sigma_{22}^{-1}\left(\boldsymbol{y}_{12}-\boldsymbol{\mu}_2\right),\quad\lambda_1=\lambda+p-1.\end{aligned} \tag{4.17}$$

(b') $\frac{\left(Y_{11|-11}-\mu_{1|2}\right)^2}{\sigma_{11.2}^2}\sim\lambda_{q(\boldsymbol{y}_{12})}\text{BeII}\left(1/2,\lambda_1/2\right)$ (or, equivalently, $\frac{\lambda_{q(\boldsymbol{y}_{12})}}{\lambda_1}F_{1,\lambda_1}$).

Based on result (b'), we can construct the $100(1-\alpha)\%$ prediction interval of $Y_{11|-11}$ for the MVt distributed data:

$$Y_{11|-11}\in\mu_{1|2}\pm\left|t_{\lambda_1}\left(\alpha/2\right)\right|\cdot\sqrt{\lambda_{q(\boldsymbol{y}_{12})}/\lambda_1}\cdot\sigma_{11.2}, \tag{4.18}$$

where $\mu_{1|2}$, $\sigma_{11.2}$, $\lambda_{q(\boldsymbol{y}_{12})}$ and λ_1 are defined in (4.17), $t_{\lambda_1}\left(\alpha/2\right)$ is the lower $\alpha/2$ quartile of the t distribution with degrees of freedom λ_1. Here we utilize the result $F_{1,\lambda_1}\left(x\right)=t_{\lambda_1}^2\left(x\right)$. When $\boldsymbol{\mu}$ and Σ are unknown, and $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ which are the estimates of $\boldsymbol{\mu}$ and Σ are used, thus the prediction interval of $Y_{11|-11}$ becomes:

$$Y_{11|-11}\in\hat{\mu}_{1|2}\pm\left|t_{\lambda_1}\left(\alpha/2\right)\right|\cdot\sqrt{\lambda_{q(\boldsymbol{y}_{12})}/\lambda_1}\cdot\hat{\sigma}_{11.2}. \tag{4.19}$$

If the MVN family is erroneously assumed for the MVt distributed data, the

$100\,(1-\alpha)\,\%$ prediction interval of $Y_{11|-11}$ has the form:

$$Y_{11|-11} \in \hat{\mu}_{1|2} \pm |t_{k-1}\,(\alpha/\,2)| \cdot \sqrt{\lambda/\,(\lambda-2)} \cdot \hat{\sigma}_{11.2}, \tag{4.20}$$

when $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown.

4.3.4 Coverage Probability of Prediction Intervals

The coverage probability of the MVt-based prediction interval is nominal because we use correct distribution assumption to construct the intervals. For the MVN-based prediction intervals,however, because the distributional assumption is incorrect, their coverage probability is different from the nominal level, which we show in the following result.

RESULT Suppose $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_n \sim \text{MV}t\,(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \lambda)$, then, the coverage probability of the MVN-based $100\,(1-\alpha)\,\%$ prediction interval of $Y_{11|-11}$ is different from $1-\alpha$.

Proof According to (4.20), the coverage probability of the MVN-based prediction interval can be expressed as:

$$\begin{aligned} \text{P}_{\text{Coverage}} &= \text{P}\left(\left|\frac{Y_{11|-11}-\hat{\mu}_{1|2}}{\sqrt{\lambda/\,(\lambda-2)}\hat{\sigma}_{11.2}}\right| \leq |t_{k-1}\,(\alpha/\,2)|\right) \\ &= \text{P}\left(\frac{\left(Y_{11|-11}-\hat{\mu}_{1|2}\right)^2}{\frac{\lambda}{\lambda-2}\hat{\sigma}^2_{11.2}} \leq F_{1,k-1}\,(1-\alpha)\right). \end{aligned} \tag{4.21}$$

According to Results (a') and (b') in Section 4.3.3, Equation (4.21) is equivalent to:

$$\text{P}\left(\frac{\left(Y_{11|-11}-\hat{\mu}_{1|2}\right)^2}{\hat{\sigma}^2_{11.2}} \leq \left[\frac{\lambda_1}{\lambda_{q(\boldsymbol{y}_{12})}} \cdot \frac{\lambda}{\lambda-2} \cdot \frac{F_{1,k-1}\,(1-\alpha)}{F_{1,\lambda_1}\,(1-\alpha)}\right] \cdot \frac{\lambda_{q(\boldsymbol{y}_{12})}}{\lambda_1} F_{1,\lambda_{q(\boldsymbol{y}_{12})}}\right),$$

where $F_{a,b}\,(1-\alpha)$ is the lower $1-\alpha$ quartile of F distribution with degrees of freedom a and b. According to (b'), it is clear that $\text{P}_{\text{Coverage}} \neq 1-\alpha$ as the quantity in the square bracket is not equal to 1.

4.3.5 Simulation Results

Consider the following scenario. We have multivariate data for N subjects. For example, the multivariate measurements can be any characteristics of the subjects such as body mass index, gender, blood pressure, etc., and they tend to be correlated. Although in literature we usually assume MVN for the underlying distribution, in reality, such data may be close to MVt. Suppose one of the measurement of one

subject is randomly missing, and the conditional prediction interval of this missing measurement is of interest. Conditional prediction intervals are more accurate than the commonly used marginal prediction intervals, since they incorporate the observed information from the the subject him/herself in addition to observations from other subjects. However, according to our previous result, the MVN-based conditional prediction interval is incorrect as it does not possess the nominal coverage probability if the underlying distribution is MVt, and we illustrate this point using a simulation below.

The simulation study is described as follows:

1. Randomly generate N (N is large) independent p-variate $\boldsymbol{Y}_i$'s following MVt distribution: $\boldsymbol{Y}_i \overset{i.i.d.}{\sim} \text{MV}t_p\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \lambda\right), i = 1, 2, \ldots, N$.

2. Randomly select an observation $\boldsymbol{Y}_i$, and calculate the conditional prediction intervals for $\left(Y_{ij} | \, \boldsymbol{Y}_{-ij}\right)$, where $j = 1, 2, \ldots, p$, respectively, using both the MVt- and MVN-based formulae.

3. Record the bounds of the prediction interval and its width. Record also whether the true missing value lies in the constructed prediction intervals.

4. Repeat Steps 1, 2, and 3 for K times, where K is large.

In our simulation, we set $N = 1000$ and $K = 1000$. We apply different λ and p values to see how the two types of prediction intervals behave. For correctly constructed $100(1-\alpha)\%$ conditional prediction intervals, we expect the probability of simultaneously covering s components of a p-variate vector $\boldsymbol{Y}_i$ is $C_p^s \alpha^s \left(1-\alpha\right)^{p-s}$. Therefore, the expected counts of simultaneously covering s components of $\boldsymbol{Y}_i$ with K simulations are $KC_p^s \alpha^s \left(1-\alpha\right)^{p-s}$. By comparing the covered counts with the MVt- and MVN-based intervals to the expected counts, we can tell whether the MVN-based prediction intervals have nominal coverage. In particular, we show the simulation results of the 95% conditional prediction intervals in Figures 4.1 and 4.2 for $\lambda = 100$ and 3, respectively, and $p = 5$.

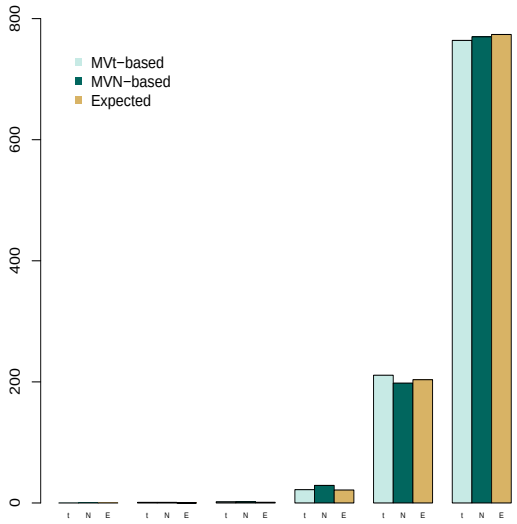

Figure 4.1: Frequencies of Covered Prediction Intervals (df=100)

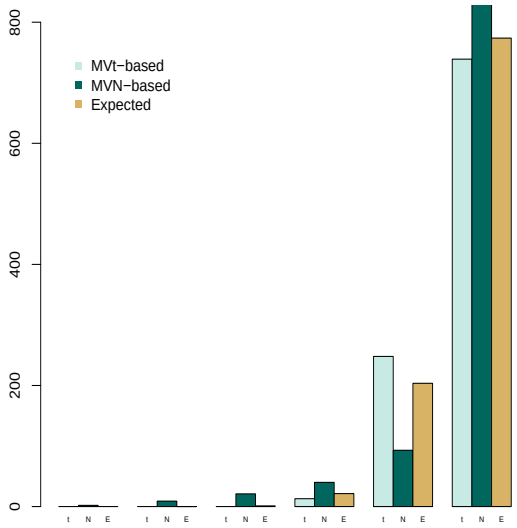

Figure 4.2: Frequencies of Covered Prediction Intervals (df=3)

In the two figures, the light and dark green bars represent the covered counts of MVt and MVN-based prediction intervals, respectively, while the brown bars represent the expected covered counts. As is shown from Figure 4.1, there is not much difference among the three when $\lambda = 100$. This is reasonable due to the fact that as λ gets large for MVt, the distribution tends to be more similar to a MVN distribution. Therefore both of the covered counts are close to the expected ones. However, when λ is small, that is, when the MVt distribution is quite different from

the MVN distribution, the MVN-based prediction intervals do not have nominal coverage probabilities, as shown in Figure 4.2.

We also compare the width of MVt-based prediction intervals with that of the MVN-based intervals. We only show interval widths for $Y_{1|-1}$ in Figures 4.3 and 4.4: others for $Y_{j|-j}$, where $j = 2, 3, \ldots, p$, look similar.

There is no noticeable difference between the two types of intervals when λ is

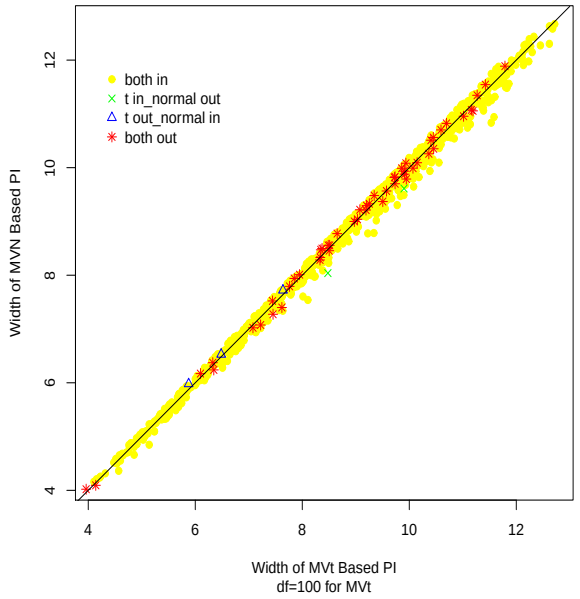

Figure 4.3: Comparison of the Widths of Prediction Intervals (df=100)

large. However, when λ is small, Figure 4.4 shows that the MVt-based intervals behave better than the MVN counterparts when the interval width is large, which shows nicely of the unique property of MVt distributions that their conditional second moments are linear in the empirical second moments of the conditioning vector, as contrasted to the constant conditional second moments of MVN distributions.

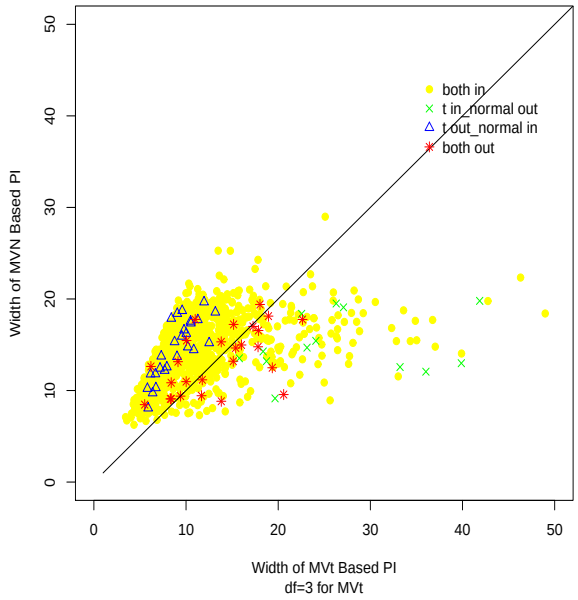

Figure 4.4: Comparison of the Widths of Prediction Intervals (df=3)

4.4 Graphical Check of MP VII Distribution

In this section we introduce a graphical method to qualitatively check the MP VII distribution vs. MVN.

This method is based on the unique conditional second moments property of the MP VII distribution, as is described in Section 4.2. Since its conditional second moment is linear in the empirical second moment of the entire conditioning vector, we expect that if we fit a LOWESS curve of the two quantities, the line should be approximately linear with a non-zero slope. Specifically, the conditional second moment can be replaced by the empirical conditional second moment, since in reality, the former is usually unknown. This can be done as follows: consider the data $\boldsymbol{Y}_i \sim \text{MP VII}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g\right), i = 1, 2, \ldots, N$, the empirical conditional second moments can be calculated as $\left(Y_{i1} - \mu_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\left(\boldsymbol{Y}_{i2} - \boldsymbol{\mu}_2\right)\right)^2$, and the empirical second moments of the conditioning vector are denoted as $\left(\boldsymbol{Y}_{i2} - \boldsymbol{\mu}_2\right)' \boldsymbol{\Sigma}_{22}^{-1}\left(\boldsymbol{Y}_{i2} - \boldsymbol{\mu}_2\right)$. We can then check the LOWESS plot to see if there is a linear relationship between the two. Here we choose the second to last components of each vector as the conditioning vector, but theoretically we should alternate the conditioning vectors and check each time to see if the conditional second moments are linear in any conditioning vectors of the given data. We show the LOWESS plot in Figure 4.5 as an example, where there are 1000 observations in the data and are simulated from a 5-variate MVt distribution with $df = 3$.

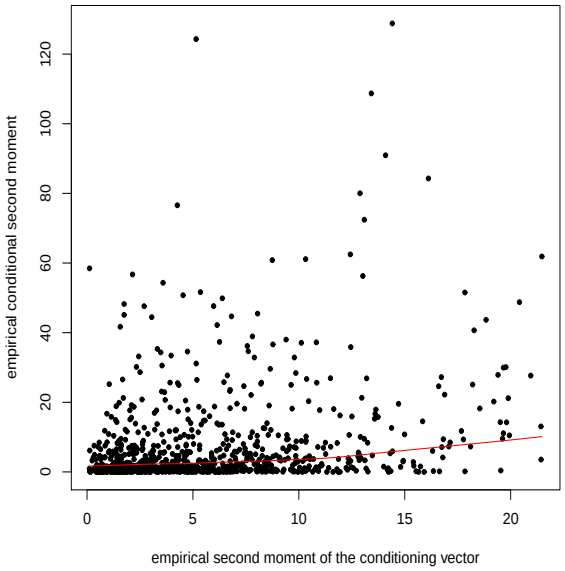

Figure 4.5: LOWESS from MVt Distribution

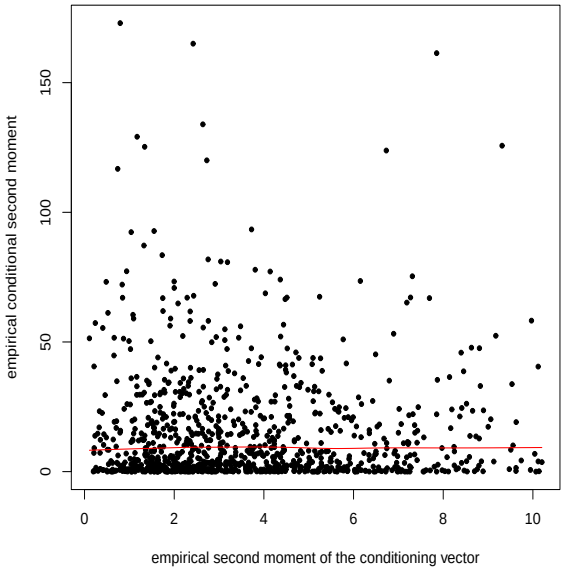

Figure 4.6: LOWESS from MVN Distribution

Figure 4.5 shows clearly a straight smoothed line, thus verifies there is a linear relationship with non-zero slope between the two variables. As a result, we can tell that the data is MP VII distributed from a LOWESS plot like Figure 4.5. For comparison purpose, we also generate $\mathbf{Y}_i^{'}$s from a MVN distribution with the same mean and covariance matrix as the MVt distributions, and the corresponding LOWESS plot is shown in Figure 4.6. It is clear that the empirical conditional second moments do not depend on the empirical second moments of the conditioning vector, which is consistent with the fact that the conditional second moments of a MVN distribution are constant.

5 Conclusion and Discussions

The presented work aims to predict daily ED influenza visits in Edmonton. The prediction is based on the ED visits data collected over 10 minutes intervals by ARTSSN, from August 1, 2004 to July 31, 2009. We first extract the systematic trend of ED influenza visits by fitting a GLMM, with natural cubic splines of the dates as explanatory variables. The daily variations can be represented by the residuals obtained from the observed daily counts and estimated daily counts from the GLMM. We then decompose the daily variations into subareas in Edmonton according to first three postal codes recorded in the data. Exploratory modeling is used to model the decomposed daily variations, assuming that the expected variation of a subarea on one day can be linearly represented by the variations of the same subarea on previous days, as well as the variations on previous days of all neighboring subareas. Parameter estimations are obtained through pseudo-likelihood method. By adding proportional weights to the pseudo-likelihood, and combining the results with those from the commonly used equally weighted pseudo-likelihood, we are able to predict the daily ED influenza visits precisely enough. Specifically, the performance of the prediction of peaking visits is good. The prediction of the peaks plays a more important role since policy makers and healthcare providers need to be better prepared for influenza epidemics.

Our prediction method takes both temporal and spatial correlations into consideration, thus provides more accurate predictions. The modeling and analytical methods are also of computational ease, and therefore can be utilized by scientists in various fields. We expect the proposed prediction method to inform healthcare providers and policy makers weeks in advance regarding the potential epidemics of influenza for proper preparation and potential alleviation of influenza epidemics.

Despite the merits mentioned above, we also have a few suggestions regarding future work. First of all, we will seek non-parametric methods to construct confidence intervals or confidence bands for the predictions. Since ARTSSN has other data sources including health link telephone calls and school absenteeism data, we could apply similar models to them to get predictions for each of the data sources. After that, we can predict the actual daily influenza cases (lab confirmed influenza cases) using prior information combined from all other three data sources, with sufficient days ahead to be useful for AHS. Finally, the prediction method can be applied to other communicable diseases.

References

[1] Arellano-Valle, R.B. and Bolfarine, H. (1995). On some characterizations of the t distribution. *Statistics and Probability Letters*, Vol. 25, pp. 79-85.

[2] Arnold, B.C. and Strauss, D. (1991). Pseudolikelihood estimation: some examples. *Indian Statistical Institute*, Vol. 42, pp. 233-243.

[3] Besag, J. (1975). Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society*, Vol. 24, pp. 179-195.

[4] Box, G.E.P. and Jenkins, G.M. (1976). Time series analysis: forecasting and control. *San Francisco, CA: Holden Days*.

[5] Cambanis, S., Huang, S., and Simons, G. (1981). On the theory of elliptically contoured distributions. *Journal of multivariate analysis*, Vol. 11, pp. 368-385.

[6] Dianne, L.G. and Thomson, D.J. (2006). Incidence of influenza in Ontario following the Universal Influenza Immunization Campaign. *Vaccine*, Vol. 24, pp.5245-5250.

[7] Fang, K.T., Kotz, S., and Ng, K.W. (1989). Symmetric multivariate and related distributions. *Chapman & Hall*, London.

[8] Fang, K.T. and Zhang, Y.T. (1990). Generalized multivariate analysis. *Springer-Verlag*, Berlin.

[9] Flahault, A., Deguen, S., and Valleron A.J. (1994). A mathematical model for the European spread of influenza. *European Journal of Epidemiology*, Vol. 10, pp. 471-474.

[10] Flahault, A., Letrait, S., Blin, P., et al. (1988). Modelling the 1985 influenza epidemic in France. *Statistics in Medicine*, Vol. 7, pp. 1147-1155.

[11] Gong, G. and Francisco J.S. (1981). Pseudo maximum likelihood estimation: theory and applications. *The Annals of Statistics*, Vol. 9, pp. 861-869.

[12] Hart, J.D. and Vieu, P. (1990). Data-driven bandwidth choice for density estimation based on dependent data. *Annals of Statistics*, Vol. 18, pp. 873-890

[13] Hashimoto, S., Murakami, Y., and Taniguchi, K., *et al*. (2000). Detection of epidemics in their early stage through infectious disease surveillance. *International Journal of Epidemiology*, Vol. 29, pp. 905-910.

[14] Hjorth, J.S.U. (1994). Computer intensive statistical methods. *Chapman & Hall*, London.

[15] Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *The Indian Journal of Statistics*, Vol. 32, pp. 419-430.

[16] Le Strat, Y. and Carrat F. (1988). Monitoring epidemiology surveillance data using hidden Markov models. *Statistics in Medicine*, Vol. 18, pp. 3463-3478.

[17] Longini, I.M. Jr, Fine, P.E., and Thacker, S.B. (1986). Predicting the global spread of new infectious agents. *American Journal of Epidemiology*, Vol. 123, pp. 383-391.

[18] Lorenz, E.N. (1969). Atmospheric predictability as revealed by naturally occurring analogies. *Journal of the Atmospheric Sciences*, Vol. 26, pp. 636-646.

[19] Molinari, N.A., Ortega-Sanchez, I.R., and Thompson, W.W., *et al.* (2007). The annual impact of seasonal influenza in the US: measuring disease burden and costs. *Vaccine*, Vol. 25, pp. 5086-5096.

[20] Nimmo-Smith, I. (1979). Linear regressions and sphericity. *Biometrika*, Vol. 66, pp. 390-392.

[21] Quenel, P. and Dab, W. (1998). Influenza A and B epidemic criteria based on time-series analysis of health services surveillance data. *European Journal of Epidemiology*, Vol. 14, pp. 275-285.

[22] Šaltytėbenth, J. and Hofoss, D. (2008). Modelling and prediction of weekly incidence of influenza A specimens in England and Wales. *Epidemiology and Infection*, Vol. 136, pp. 1658-1666.

[23] Samko, S.G., Kilbas, A.A., and Marichev, O.I. (1993). Fractional integrals and derivatives. *Gordon & Breach*, Amsterdam.

[24] Stroup, D.F., Thacker, S.B., and Herndon, J.L. (1988). Application of multiple time series analysis to the estimation of pneumonia and influenza mortality by age 1962Ű1983. *Statistics in Medicine*, Vol. 7, pp. 1045-1059.

[25] Szablowski, P.J. (1990). Expansions of $E(X|Y + \varepsilon Z)$ and their applications to the analysis of elliptically contoured measures. *Computers & Mathematics, with Applications*, Vol. 19, pp. 75-83.

[26] Viboud, C., Boëlle, P.Y., and Carrat, F., *et al*. (2003). Prediction of the Spread of Influenza Epidemics by the Method of Analogues. *American Journal of Epidemiology*, Vol. 158, pp. 996-1006.

Appendix

- Qiu, W., Savu, A., and Yasui, Y. Oral presentation: Elliptical Distributions with Linear Conditional Second Moments: Multivariate Pearson Type VII Family and Applications. *2011 Statistical Society of Canada Meetings*, Wolfville, Nova Scotia.

- Qiu, W., Savu, A., and Yasui, Y. Oral presentation: Elliptical Distributions with Linear Conditional Second Moments: Multivariate Pearson Type VII Family and Applications. *2011 Joint Statistical Meetings*, Miami Beach, Florida, US.