

國立交通大學

統計學研究所

碩士論文

動態計數時間序列之估計

Estimation of Dynamic Model of Time Series
Count Data with Application to Traffic Flow Forecast

研究生：黃鶯筑

指導教授：周幼珍 博士

中華民國九十五年六月

動態計數時間序列之估計

Estimation of Dynamic Model of Time Series
Count Data with Application to Traffic Flow Forecast

研究生：黃薦筑

Student: Yan-Chu Huang

指導教授：周幼珍 博士

Advisor: Dr. Yow-Jen Jou

國立交通大學理學院

統計學研究所

碩士論文

A Thesis
Submitted to Institute of Statistics
College of Science
National Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of Master
in
Statistics
June 2006
Hsinchu, Taiwan, Republic of China

中華民國九十五年六月

動態計數時間序列之估計

研究生：黃鶯筑

指導教授：周幼珍 博士

國立交通大學統計所研究所

中文摘要

本篇論文描述分析從 latent process 得到的計數資料的 parameter-driven 模式之方法，且 latent process 與計數資料具有相關性。而此模型將產生十分複雜的概似函數，modified EM 演算法是將 EM 演算法中的 E 步驟作了一些修改，即將 E 步驟中給定 y 的條件期望值用邊際期望值替代。我們將用兩種時間序列的例子來說明我們的方法：車流量資料和 Zeger (1988) 的小兒麻痺發生率序列。藉由架設在路邊的偵測器所收集到的資料，我們可以估計、配適及預測交通網絡中的交通狀態，而這些訊息將是信號控制以及交通車隊管理的關鍵。

Estimation of Dynamic Model of Time Series Count Data with Application to Traffic Flow Forecast

Student: Yan-Chu Huang

Advisor: Dr. Yow-Jen Jou

Institute of Statistics
National ChiaoTung University



This thesis describes the methodology for analyzing in parameter-driven models for time series of count data generated from latent process that characterize the correlation structure. These models result in very complex likelihoods. A modified EM algorithm is proposed which we replace the marginal expectation with the conditional expectation given y in the E step of the EM algorithm. We illustrate our method by two time series: the traffic flow data and Zeger's polio incidence series. Through the data collected by the detectors mounted on the road, we can estimate, smooth and predict the traffic condition about the network, these information are critical to signal control and traffic queue management.

誌 謝 辭

首先，我要超級感謝這兩年給與我指導的周幼珍老師，在這過程中，逐漸清楚了論文的方向，並且在生活上也給與了我很多正面的影響與鼓勵，期望在未來的日子裡依然能夠與周幼珍老師維持良好的互動。接下來要感謝的是運管所的卓訓榮老師、博士班的曾明德學長和藍建綸學長，在論文的資料蒐集上與程式的撰寫盡力的給與我最大的幫忙。還要謝謝所上的所有老師和郭姐這兩年來對我的照顧，還有溫先生。再來要特別感謝的是我的好伙伴兼好朋友宛茹，一起作計畫的時候我們相互扶持、一起努力完成成果的喜悅，還有一起改作業的日子，也感謝妳總是體諒我早上無法太早出現在學校，都是我珍貴的回憶喔！大宛、秀慧、小宛、婉文、沛君、孟樺、小馬，感謝老天讓我在這裡遇見了妳們，一起度過的每個日子，還有鼓勵我的每一句話，對我來說是碩士生活中最美好的回憶，即使今後必須各分西東，相信堅定的友誼是不會變的喔，保持聯絡。還有Action、冠達、清峰……等等，不時的給與我鼓勵的聲音，一點一滴都是我增加信心的力量，謝謝你們。

最後，要感謝我的室友，在我沮喪、無助還有壓力很大的時候總是不斷的鼓勵我，給我支持的力量，讓我能夠再度向前。還有我的父母，雖然很多話我都不會向家裡說，但是我總以能成為你們的驕傲作為目標，這也是支持我很大的一股力量。在此僅以此篇論文獻給在我身旁的你們。

黃鶯筑 謹誌于
國立交通大學統計研究所
中華民國九十五年六月

CONTENTS

1. Introduction	1
2. Literature Review	4
3. Model Specifications and Methodology	
3.1 Parameter-Driven Models for Count Data and the Application of the modified EM Algorithm	
3.1.1 The Model	7
3.1.2 Modified EM Algorithm.....	8
3.2 Fitting, Prediction and the Estimate Information Matrix	12
4. Numerical Example	
4.1 Polio Incidence	14
4.2 Traffic Flow	
4.2.1 Data Descriptions.....	23
4.2.2 Data Analysis.....	27
5. Conclusion	36
Appendix	37
References	40

LIST OF TABLES

Table 1 Monthly numbers of U.S. cases of poliomyelitis for 1970 to 1983	14
Table 2 Coefficients and Standard Errors.....	18



LIST OF FIGURES

Figure 1 Monthly Numbers of Polio Cases. Dots represent observations. The thicker solid line connects the fitted data simulated from method 1. The thinner solid line connects the fitted data simulated from method 2.....	16
Figure 2 $Q(\bullet \theta)$ for Iterations 0 Through 690.....	19
Figure 3 Change for AR(1) coefficient (ρ) estimates.....	19
Figure 4 Change for noise variance (σ_e^2) estimates	19
Figure 5 Change for covariate vector coefficients (α) estimates.....	20
Figure 6 $Q(\bullet \theta)$ for Iterations 0 Through 473.....	22
Figure 7 Monthly Numbers of Polio Cases. Dots represent observations. The thicker solid line connects the fitted data simulated from method 1. The thinner solid line connects the fitted data simulated from method 2.....	22
Figure 8 RTMS in Forward-Looking Mode	24
Figure 9 Occupancy-Volume Plot. The majority of the data set occurs at the small values of occupancy and roughly shows a positively correlated relationship between volume and occupancy. The plot conforms to the Figure 10 which is the relationship of flow and density in Traffic Theory.....	25
Figure 10 Flow (q) and Density (k); Density =constant \times Occupancy.....	25
Figure 11 Data of Traffic Flow (Complete Data).....	26
Figure 12 Data of Traffic Flow (First 200 time sequence of Complete Data)	26
Figure 13 Time sequence of the traffic flow. Dots represent observations. The thinner solid line connects the fitted data simulated from method 1. The thicker solid line connects the fitted data simulated from method 2.	28
Figure 14 Cumulative Traffic Flow. Dots represent observations. The upper figure shows the fitted data simulated from method 1. The lower figure shows the fitted data simulated from method 2. It is clear that the fitted data from method 1 is closer the traffic flow than that one from method 2, this	

unstable situation maybe come from generator which produces the latent process $\{W_t\}$	29
Figure 15 $Q(\bullet \theta)$ for Iterations 0 Through 150	30
Figure 16 Change for AR(1) coefficient (ρ) estimates	31
Figure 17 Change for noise variance (σ_ε^2) estimates	31
Figure 18 Change for covariate vector coefficients (α) estimates	32
Figure 19 $Q(\bullet \theta)$ for Iterations 0 Through 147	33
Figure 20 Time sequence of the traffic flow. Dots represent observations. The thinner solid line connects the fitted data simulated from method 1. The thicker solid line connects the fitted data simulated from method 2.	34
Figure 21 Cumulative Traffic Flow. Dots represent observations. The upper figure shows the fitted data simulated from method 1. The lower figure shows the fitted data simulated from method 2. It is clear that the fitted data from method 1 is closer the traffic flow than that one from method 2, this unstable situation maybe come from generator which produces the latent process $\{W_t\}$	35

1 Introduction

Traffic congestion is a daily occurrence in most urban areas and is becoming a serious problem in many urban and sub-urban areas. A roadway system is operating in saturated and oversaturated conditions. A signalized intersection is said to be “oversaturated” if the demand volume exceeds the capacity of the intersection. Over-saturation refers to conditions where traffic queues persist from cycle to cycle due to insufficient green splits or because of blockage. In such conditions queues along signalized arterials may block upstream intersections thus exacerbating an already bad condition. It is critical that appropriate signal control and queue management procedures are in place. Otherwise, excessive queues and spillbacks into upstream intersections would lead to gridlock with serious disruptive consequences on system operations. The need for appropriate traffic management procedures becomes even more pressing with the deployment of more efficient traffic control systems within the intelligent transportation systems (ITSs) complex which handles information, including traffic flow prediction, on a real-time base.

Several queue management and signal control strategies have been described in literature. However, to design effective signal control schemes, the traffic flows occur in the next cycles are required. Traffic signal, as an essential element of an urban transportation network plays a critical rule in the operation of urban street network and could not be neglected along with the movement of ITSs. Over the past few decades the study on signalized intersections has been carried out by various methods, however, still it is well accepted the benefit of signal control has not been fully realized. In general, current researches on traffic signal control can be grouped into two classes from the view of flow-capacity situation, i.e. under-saturation and

congestion control. The model concerning with the under-saturated flow condition serve well from isolate intersection to network control. However for a congested or over-saturated network, most of the procedures were based upon experience or simple analysis and were not efficient in easing the congestion situation. To achieve more efficient management of the traffic, it is better to keep the road system in free-flow condition as long as possible. Therefore, a dynamic model with covariates is used to describe the time series of traffic flow counts.

In some special time, like raining day, the day before long holidays or rush hours, the level of oversaturated traffic condition will become more serious. So we hope to control the daily traffic before the rush hours effectively. If we are able to release the traffic flow of the section effectively, it may be helpful to release the jammed traffic after that. However, so far most of the traffic sign in Taiwan are still controlled by the police when the traffic is heavy. Although it will effectively control at once, it is difficult to maintain the high level performance of the signals due to variability existing in personal manipulation. Moreover, the range of personal control and the vision are restricted at the located intersection. Even we transmit messages by wireless; we are not able to effectively achieve traffic flow and the continuance of traffic sign. Therefore, if we can predict the traffic flow precisely, it may help us a lot. We hope to invent a method to calculate the capacity of roads in Taiwan, and the first step of the work is to get the model of the traffic flow curve. Therefore, the purposes of this thesis are to use a more realistic model and an appropriate approach to estimate and predict the future traffic flow.

This thesis describes the methodology for estimating the parameter-driven models for time series of counts which are generated from latent unobservable process that characterize the correlation structure. It is different from previous procedures in many respects. It is a new

general formulation for an under-saturated network.

The rest of the thesis is organized as follows. Literature on modeling time series with count data are briefly reviewed in Section 2. Section 3 consists of two subsections. In the first subsection we describe the formulas of the modified EM algorithm to first-order parameter-driven model for time series of count data; and in the second subsection we describe two methods to smooth and predict, and to estimate information matrix. Two real data applications, including the traffic flow data and Zeger's polio incidence series, are presented in Section 4. Finally, we discuss the advantages and shortcomings in Section 5. Proofs are collected in the Appendix.



2 Literature Review

Let $\{Y_t\}_{t=1}^N$ be a time series of count data and let $\{U_t\}_{t=1}^N$ be an observed vector-valued covariate process. We are interested in modeling the contemporaneous dependence of Y_t on the covariates in U_t , as well as its dependence on past values. In order to describe $\mu_t = E(Y_t)$ as a function of a $p \times 1$ vector of covariates, U_t , classical linear models are regression models with the assumption that the observations are normally distributed with a covariance matrix that is known up to a constant. Generalized linear model (Nelder and Wedderburn, 1972) extended the classical linear model by allowing for non-normal noise such as the case when the response variable is binary or integer-valued. With independent observations, logistic and log linear models can be used for this purpose. If the response variables are a time series, serial dependence should be incorporated in the model.

Because of the nature of the traffic flow, a discrete time series model will be utilized to describe its behavior. Discrete time series occur in many contexts, often as counts of events, objects or individuals in consecutive intervals or at consecutive points in time. Modelling discrete time series is the most challenging, as yet, least well developed of all areas of research in time series. For time series data, Cox (1981) suggested two classes of models of time-dependent data: parameter-driven and observation-driven models. In an observation-driven, the conditional distribution of observation is specified as a function of past observations. Autoregressive models for Gaussian series are example for continuous data. Integer-valued autoregressive moving average models and Poisson autoregressive (PAR), introduced by Al-Osh and Alzaid (1987), Freeland and McCabe (2004) applied the model to describe monthly count data set of claimants for wage loss benefit. Davis et al. (2003) are

examples for discrete data. In a parameter-driven model, which is conceptually more elegant, serial correlation is introduced through an unobserved latent process. Both specifications are becoming increasingly popular because of their ability to handle serial correlation and over-dispersion in the data.

Recently, much research effort has been devoted to the theoretical development and refinement of parameter-driven models. An important approach to the incorporation of covariates was initiated by Zeger (1988) who was concerned with testing for a falling trend in monthly U.S. polio incidence data. He used Poisson regression to model trend and seasonality explicitly, and a stationary process $\{\varepsilon_t\}$ with unit mean and autocovariance function $Cov(\varepsilon_t, \varepsilon_{t-k}) = \rho_\varepsilon(k)\sigma^2$ to model the serial correlation. Campbell (1994) extended this approach to higher orders of dependence in modeling occurrences of sudden infant death syndrome and related trends to environmental factors. Brannas and Johansson (1996) further extended to panel data. Chan and Ledolter (1995), and Jorgensen et al. (1996), among others, also tried to tackle the problem of modeling discrete time series. Their common assumption is that the observations are conditionally independent given the latent process. Without assuming the full information on the latent process, Zeger (1988) estimated the parameters by quasi-likelihood estimating equation. Chan and Ledolter (1995) model the polio data as a dynamic GLM in which the latent process, $\{W_t\}$, is an Gaussian AR(1) process, i.e. $W_t = \rho W_{t-1} + \varepsilon_t$, where $\{\varepsilon_t\}$ is iid $N(0, \sigma^2)$. Explicitly, conditional on trend and seasonal covariates, $\{U_t\}$, $(y_t | W_t, U_t) \sim Poisson(\lambda_t)$, where $\log \lambda_t = \alpha' U_t + W_t$. The authors treated the latent process as missing values and used a variant of the EM algorithm to maximize the likelihood. In their algorithm, the E-step is performed indirectly using Monte Carlo simulation. The approach is direct to estimate the parameters but suffered from, as the authors stated in

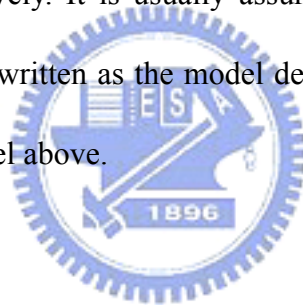
the paper, noisy simulated log-likelihood of the complete data. In this thesis, we will basically follow the formulation but modify the estimation process.

In the linear, Gaussian structure, state-space models are parameter-driven models in which an unobserved state-vector θ_t evolves linearly through time and only a linear function of it is observed, usually in noise. Thus, there are two defining equations: an observation equation, and evolution or system equation, typically of the form:

$$X_t = F_t \theta_t + \nu_t, \quad (1)$$

$$\theta_t = G_t \theta_{t+1} + \omega_t, \quad (2)$$

where $\{\nu_t\}$ and $\{\omega_t\}$ are independent sequences of i.i.d. r.v.s of zero mean and covariance matrices V_t and W_t , respectively. It is usually assumed that $\{F_t\}$, $\{G_t\}$, $\{V_t\}$ and $\{W_t\}$ are known. Our model can be written as the model described in Section 3.1.1, it can also be expressed as a state-space model above.



3 Model Specifications and Methodology

3.1 Parameter-Driven Models for Count Data and the Application of the modified EM Algorithm

3.1.1 The Model

The vehicle detectors equipped on the road system record the information of flow, speed, and occupancy every ten seconds. We are interested in utilizing these information to predict the future traffic flow. Let $\{Y_t\}_{t=1}^N$ denote the traffic flow process. At each specific time t , Y_t is the number of cars passing the vehicle detector and hence is discrete in nature. We now discuss a generalized regression model for a time series of count data. To account for the dependence of the count data on its past values and other relevant contemporaneous covariates, we use a simple parameter-driven model. Let $Y = y$ be the observed complete data of which Y is a measurable function. The parameter space is denoted by Ω and two arbitrary elements are denoted by θ and θ' . We assume that $\{W_t\}$ is a stationary Gaussian AR(1) latent process. Given this latent process $\{W_t\}$, the observations $\{Y_t\}$ are generated independently from a Poisson distribution with mean λ_t satisfying

$$\log \lambda_t = W_t + \alpha' U_t, \quad (3)$$

where α and U_t are both $p \times 1$ vector.

That is, conditional on $\{W_t\}$ the observations are independent Poisson random variables with means λ_t , written as

$$y_t | \lambda_t \sim \text{Poisson}(\lambda_t), \quad (4)$$

where $\{\lambda_t\}$ is defined as $\lambda_t = \exp(W_t + \alpha' U_t)$, and

$$W_t = \rho W_{t-1} + \varepsilon_t, \quad (5)$$

where $\{\varepsilon_t\}$ is iid $N(0, \sigma_\varepsilon^2)$. The parameter vector consists of the coefficients in equation (4) and equation (5); that is, $\theta = (\rho, \sigma_\varepsilon^2, \alpha')$.

Parameter-driven models specifies an unobserved latent process and results in a complex likelihood but they are more elegant and are straightforward in their interpretation of covariates on the observed count process and its demanding computational effort can be eased by the computational power nowadays.

3.1.2 Modified EM Algorithm

Let $X_t = (y_t, W_t)$ denote the complete-data random vector of interest at time t, $\mathbf{X} = (X_1, \dots, X_N)$ be the realization of the complete-data. Let $\mathbf{Y} = (y_1, \dots, y_N)$ be the observed data and $\mathbf{W} = (W_1, \dots, W_N)$ be the latent process and the unobserved data. The joint probability density function of (y_t, W_t) is given by

$$f_X(y_t, W_t | W_{t-1}) = \frac{\exp(-\exp(W_t + \alpha'U_t)) \times (\exp(W_t + \alpha'U_t))^{y_t}}{y_t!} \times \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left(-\frac{1}{2\sigma_\varepsilon^2}(W_t - \rho W_{t-1})^2\right), \quad (6)$$

and the log-likelihood of \mathbf{X} (conditional on the initial latent variable W_0 , where W_0 is

$N\left(0, \frac{\sigma_\varepsilon^2}{(1-\rho^2)}\right)$) is given by

$$l_X(\theta) = \sum_{t=1}^N (-\exp(W_t + \alpha'U_t) + y_t W_t + y_t \alpha'U_t) - \frac{n}{2} \log(\sigma_\varepsilon^2) - \frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^N (W_t - \rho W_{t-1})^2 \quad (7)$$

$$= \sum_{t=1}^n \left(-\exp(W_t) \exp(\alpha' U_t) + y_t \alpha' U_t \right) + \sum_{t=1}^n y_t W_t - \frac{n}{2} \log(\sigma_\varepsilon^2) \\ - \frac{1}{2\sigma_\varepsilon^2} \left(\sum_{t=1}^n W_t^2 - 2\rho \sum_{t=1}^n W_t W_{t-1} + \rho^2 \sum_{t=1}^n W_{t-1}^2 \right),$$

Clearly, $l_X(\theta)$ is linear in $\exp(W_t)$, $\sum_{t=1}^n y_t W_t$, $\sum_{t=1}^n W_t^2$, $\sum_{t=1}^n W_t W_{t-1}$ and $\sum_{t=1}^n W_{t-1}^2$, which are measurable functions of \mathbf{X} . Davis, Dunsmuir, and Wang (2000) find the MLE of α by ignoring the presence of a latent process in the model.

In the presence of latent process, which will be treated as missing data, only the log likelihood function $l_X(\theta|y) = \log \int f_X(y, W|\theta) dW$ is observed. Maximizing $l_X(\theta|y)$ is difficult. The EM algorithm maximizes $l_X(\theta|y)$ by iteratively maximizing $l_X(\theta)$. Each iteration of the EM algorithm consists of two steps:

E step: from $Q(\theta'|\theta) = E_\theta(l_X(\theta')|y)$;

M step: maximize $Q(\cdot|\theta)$.

$E_\theta(\cdot|y)$ denotes the conditional expectation given $\mathbf{Y}=y$, where θ is the true parameter and $l_X(\theta')$ is the log-likelihood of \mathbf{X} .

Because the conditional expectation given y of $l_X(\theta|y)$ is intractable, so we do some adjustment on EM algorithm. Then we replace the marginal expectation with the conditional expectation given y . Furthermore,

$$E_\theta(\exp(W_t)) = \exp\left(\frac{\sigma_\varepsilon^2}{2(1-\rho^2)}\right),$$

$$E_\theta(Y_t W_t) = 0,$$

$$E_\theta(W_t^2) = \frac{(1-\rho^{2(t+1)})}{(1-\rho^2)} \sigma_\varepsilon^2,$$

$$E_\theta(W_t W_{t+1}) = \frac{\rho}{(1-\rho^2)} \sigma_\varepsilon^2.$$

Hence,

$$Q(\bullet | \theta^{(r)}) = \sum_{t=1}^n \left(-\exp\left(\frac{\sigma_\varepsilon^{2(r)}}{2(1-\rho^{(r)2})}\right) \exp(\alpha'U_t) + y_t \alpha'U_t \right) - \frac{n}{2} \log(\sigma_\varepsilon^2) \quad (8)$$

$$- \frac{1}{2\sigma_\varepsilon^2} \left(\sum_{t=1}^n \frac{(1-\rho^{(r)2(t+1)})}{(1-\rho^{(r)2})} \sigma_\varepsilon^{2(r)} - 2\rho \sum_{t=1}^n \frac{\rho^{(r)}}{(1-\rho^{(r)2})} \sigma_\varepsilon^{2(r)} + \rho^2 \sum_{t=1}^n \frac{(1-\rho^{(r)2(t)})}{(1-\rho^{(r)2})} \sigma_\varepsilon^{2(r)} \right).$$

Note that $Q(\theta'|\theta)$ is concave in θ' . This is the reason for conditioning on W_0 , because without conditioning, $Q(\theta'|\theta)$ is no longer concave. After obtaining the r^{th} -step estimate of θ , denoted by $\theta^{(r)}$, the $\rho^{(r+1)}$ and $\sigma_\varepsilon^{2(r+1)}$ components of the maximizer of $Q(\bullet | \theta^{(r)})$ are given by

$$\rho^{(r+1)} = \frac{E\left(\sum_{t=1}^n W_t W_{t+1}\right)}{E\left(\sum_{t=1}^n W_t^2\right)}, \quad (9)$$

$$= \frac{\sum_{t=1}^n \frac{\rho^{(r)}}{(1-\rho^{(r)2})} \sigma_\varepsilon^{2(r)}}{\sum_{t=1}^n \frac{(1-\rho^{(r)2(t)})}{(1-\rho^{(r)2})} \sigma_\varepsilon^{2(r)}}$$


and

$$\sigma_\varepsilon^{2(r+1)} = \frac{1}{n} \left(E\left(\sum_{t=1}^n W_t^2\right) - \frac{\left[E\left(\sum_{t=1}^n W_t W_{t-1}\right) \right]^2}{E\left(\sum_{t=1}^n W_{t-1}^2\right)} \right) \quad (10)$$

$$= \frac{1}{n} \left(\sum_{t=1}^n \frac{(1-\rho^{(r)2(t+1)})}{(1-\rho^{(r)2})} \sigma_\varepsilon^{2(r)} - \frac{\left[\sum_{t=1}^n \frac{\rho^{(r)}}{(1-\rho^{(r)2})} \sigma_\varepsilon^{2(r)} \right]^2}{\sum_{t=1}^n \frac{(1-\rho^{(r)2(t)})}{(1-\rho^{(r)2})} \sigma_\varepsilon^{2(r)}} \right).$$

The $\alpha^{(r+1)}$ components of the maximizer of $Q(\bullet | \theta^{(r)})$ is estimated by Iterative Reweighed Least Square (IRLS).

The complete procedure of the modified EM algorithm is summarized below:

Step 1 Initialize by giving the initial values $\theta_0 = (\rho_0, \sigma_{\varepsilon_0}^2, \alpha_0)'$

Step 2 The *E step*:

$$E_{\theta}(\exp(W_t)) = \exp\left(\frac{\sigma_{\varepsilon}^{2(r)}}{2(1-\rho^{(r)2})}\right)$$

$$E_{\theta}(Y_t W_t) = 0$$

$$E_{\theta}(W_t^2) = \frac{(1-\rho^{(r)2(t+1)})}{(1-\rho^{(r)2})} \sigma_{\varepsilon}^{2(r)}$$

$$E_{\theta}(W_t W_{t+1}) = \frac{\rho^{(r)}}{(1-\rho^{(r)2})} \sigma_{\varepsilon}^{2(r)}$$

Step 3 The *M step*: maximize $Q(\bullet | \theta^{(r)})$, i.e.,

$$\rho^{(r+1)} = \frac{\sum_{t=1}^n \frac{\rho^{(r)}}{(1-\rho^{(r)2})} \sigma_{\varepsilon}^{2(r)}}{\sum_{t=1}^n \frac{(1-\rho^{(r)2(t)})}{(1-\rho^{(r)2})} \sigma_{\varepsilon}^{2(r)}}$$

$$\sigma_{\varepsilon}^{2(r+1)} = \frac{1}{n} \left(\sum_{t=1}^n \frac{(1-\rho^{(r)2(t+1)})}{(1-\rho^{(r)2})} \sigma_{\varepsilon}^{2(r)} - \frac{\left[\sum_{t=1}^n \frac{\rho^{(r)}}{(1-\rho^{(r)2})} \sigma_{\varepsilon}^{2(r)} \right]^2}{\sum_{t=1}^n \frac{(1-\rho^{(r)2(t)})}{(1-\rho^{(r)2})} \sigma_{\varepsilon}^{2(r)}} \right)$$

$\alpha^{(r+1)}$ is estimated by IRLS to maximize $Q(\bullet | \theta^{(r)})$

Step 4 If $|\rho^{(r)} - \rho^{(r+1)}| < \text{error}$, $|\sigma_{\varepsilon}^{2(r)} - \sigma_{\varepsilon}^{2(r+1)}| < \text{error}$ and $|\alpha^{(r)'} - \alpha^{(r+1)'}| < \text{error}$ stop

Step 5 Update the parameters $(\rho^{(r)}, \sigma_{\varepsilon}^{2(r)}, \alpha^{(r)'})' = (\rho^{(r+1)}, \sigma_{\varepsilon}^{2(r+1)}, \alpha^{(r+1)'})'$

Step 6 Go to Step 2

3.2 Fitting, Prediction and the Estimate Information Matrix

We consider the model in equation (4) and equation (5) of Section 3.2.1, and the EM algorithm of Section 3.1.2 can be readily applied. Furthermore, because

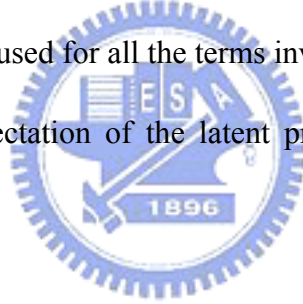
$$E_{\theta}(\lambda_t | y) = \exp(\alpha' U_t) E_{\theta}(\exp(W_t) | y) \quad (11)$$

for all t , fitting and prediction are easily carried out. The l -step prediction can be obtained from

$$E_{\theta}(\lambda_{N+l} | y) = \exp\left(\alpha' U_{N+l} + \frac{(1-\rho^{2l})}{2(1-\rho^2)} \sigma_{\varepsilon}^2\right) E(\exp(\rho' W_N) | y). \quad (12)$$

This approach is easily modified to handle missing values.

Now, we use two methods to smooth and predict the response variable. In method 1, unconditional expectations are used for all the terms involving the latent process. In method 2, estimated the conditional expectation of the latent process is substituted by the simulated results.



Method 1 (Take expectation by $E_{\theta}(\lambda_t)$)

Step 1 (Fitting of Y_t , $t=1, \dots, N$)

$$\hat{Y}_t = \hat{E}_{\theta}(\lambda_t) = \exp\left(\hat{\alpha}' U_t + \frac{\hat{\sigma}_{\varepsilon}^2}{2(1-\hat{\rho}^2)}\right) \text{ for } t=1, \dots, N \quad (13)$$

Step 2 (Prediction of Y_{N+l} , $l \geq 1$)

$$\hat{E}_{\theta}(\lambda_{N+l}) = \exp\left(\hat{\alpha}' U_{N+l} + \frac{\hat{\sigma}_{\varepsilon}^2}{2(1-\hat{\rho}^2)}\right) \quad (14)$$

Method 2 (Generate the model in equation (4) and equation (5))

Step 1 (Generate W_t , $t = 1, \dots, N$)

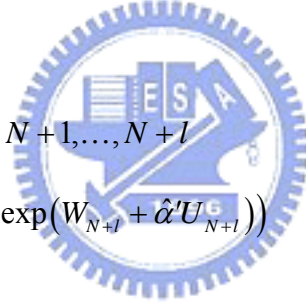
1. Generate $W_0 \sim N\left(0, \frac{\hat{\sigma}_\varepsilon^2}{(1-\hat{\rho}^2)}\right)$
2. Generate $\varepsilon_t \sim N(0, \hat{\sigma}_\varepsilon^2)$
3. Let $W_t = \hat{\rho}W_{t-1} + \varepsilon_t$
4. Repeat 2. , 3. for $t = 1, \dots, N$

Step 2 (Fitting of Y_t , $t = 1, \dots, N$)

Generate $Y_t \sim Po(\exp(W_t + \hat{\alpha}'U_t))$ for $t = 1, \dots, N$

Step 3 (Prediction of Y_{N+l} , $l \geq 1$)

1. Generate $\varepsilon_t \sim N(0, \hat{\sigma}_\varepsilon^2)$
2. Let $W_t = \hat{\rho}W_{t-1} + \varepsilon_t$
3. Repeat 1. , 2. for $t = N+1, \dots, N+l$
4. Generate $Y_{N+l} \sim Po(\exp(W_{N+l} + \hat{\alpha}'U_{N+l}))$



In order to interpret the accuracy of the estimation of the parameters, we calculate the estimated information matrix which is given by

$$I(\hat{\theta}) = E\left(-\frac{\partial^2 l_x(\theta)}{\partial \theta \partial \theta'}\right)\bigg|_{\theta=\hat{\theta}}, \quad (15)$$

and the estimated variance-covariance matrix is

$$\hat{Cov}(\hat{\theta}) = I^{-1}(\hat{\theta}). \quad (16)$$

4 Numerical Example

4.1 Polio Incidence

Table 1 Monthly numbers of U.S. cases of poliomyelitis for 1970 to 1983

	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1970	0	1	0	0	1	3	9	2	3	5	3	5
1971	2	2	0	1	0	1	3	3	2	1	1	5
1972	0	3	1	0	1	4	0	0	1	6	14	1
1973	1	0	0	1	1	1	1	0	1	0	1	0
1974	1	0	1	0	1	0	1	0	1	0	0	2
1975	0	1	0	1	0	0	1	2	0	0	1	2
1976	0	3	1	1	0	2	0	4	0	2	1	1
1977	1	1	0	1	1	0	2	1	3	1	2	4
1978	0	0	0	1	0	1	0	2	2	4	2	3
1979	3	0	0	2	7	8	2	4	1	0	2	4
1980	0	1	1	1	3	0	0	0	0	0	0	1
1981	1	0	0	0	0	0	1	2	0	2	0	0
1982	0	1	0	1	0	1	0	2	0	0	1	2
1983	0	1	0	0	0	1	2	1	0	1	3	6

We illustrate the modified EM algorithm by an example taken from Zeger (1988), who analyzed the monthly number of case of poliomyelitis from January 1970 – December 1983, and Table 1 lists the monthly numbers. A central question studied by Zeger (1988) is whether the data follow a decreasing time trend. There is seasonality in the data, and it is modeled with trigonometric components involving the first time harmonics. The covariate vector in Equation (3) is given by $U_t = (1, t/1000, \cos(2\pi t/12), \sin(2\pi t/12), \cos(2\pi t/6), \sin(2\pi t/6))'$ where p is 6.

Furthermore, we also would like to compare the results of the modified EM algorithm with those of the estimating equation method proposed by Zeger and with those of the MCEM algorithm proposed by Chan and Ledolter. Zeger's model is slightly different from our model in Section 3.2.1. Zeger assumed that $\exp(W_t)$ follows a stationary AR(1) process and used the estimating equation method to estimate the parameters, but both our model and Chan and Ledolter's model assumed that W_t is Gaussian.

We adopt the latent AR(1) model described in the previous section, and we used the modified EM algorithm to carry out the estimation. The starting values for ρ and σ_ε^2 are 0.5 and 0.5. The starting values of α are obtained by fitting a log-linear model to the data, assuming no temporal dependence, i.e. $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)'$ are $(0.554641, -4.843069, 0.136865, -0.541847, 0.455233, -0.059814)'$.

The value of $Q(\bullet | \theta)$ is maximized and the approximate ML estimates are given by

$$\hat{\rho} = 0.650877 \quad \hat{\sigma}_\varepsilon^2 = 0.384351$$

$$(0.058719) \quad (0.042059)$$

$$\hat{\alpha}' = (0.223810 , -4.798661 , 0.137132 , -0.534985 , 0.458797 , -0.069627)$$

$$(0.127307) (1.402918) (0.089482) (0.115477) (0.101469) (0.098128)$$

The figures within parentheses are estimate standard errors, obtained from Equation (16).

Fitted results in Section 3.2 are shown in Figure 1, where dots represent the observations.

The thicker solid line connects the fitted data simulated from method 1, and the average error with observations and fitted data is $\frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| = 1.150442$. The thinner solid

line connects the fitted data simulated from method 2, and the average error with observations and fitted data is $\frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| = 1.535714$.

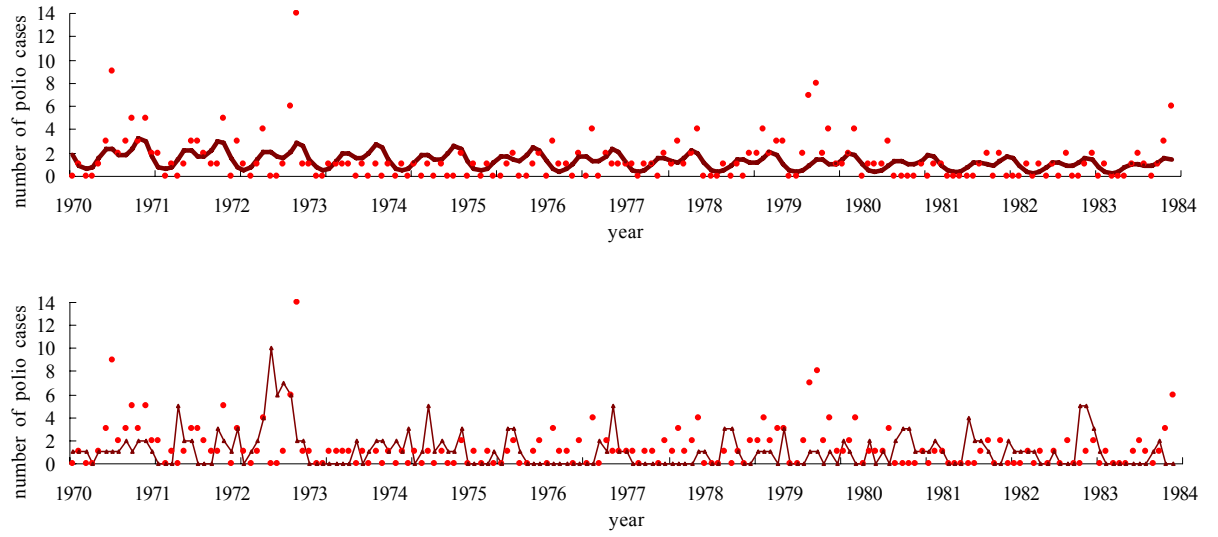


Figure 1 Monthly Numbers of Polio Cases. Dots represent observations. The thicker solid line connects the fitted data simulated from method 1. The thinner solid line connects the fitted data simulated from method 2.



Because Zeger (1988) modeled $\xi_t = \exp(W_t)/E_\theta(\exp(W_t))$, his and our results are not directly comparable. The following calculations shed light on their relationships. Because we have calculated $E_\theta(\exp(W_t)) = \exp(\sigma_\varepsilon^2/2(1-\rho^2))$ in Section 3.1.2, and because ξ_t in Zeger's model is normalized to have mean 1, we need to subtract $\sigma_\varepsilon^2/2(1-\rho^2)$ from our intercept estimate to make them comparable. Letting σ_ξ^2 and $\rho_\xi(1)$ be the variance and the first lag autocorrelation of ξ_t , we obtain σ_ξ^2 and $\rho_\xi(1)$ as follow:

$$\sigma_\xi^2 = \exp\left(\frac{\sigma_\varepsilon^2}{1-\rho^2}\right) - 1$$

$$\rho_{\xi}(1) = \frac{\exp\left(\frac{\rho\sigma_{\varepsilon}^2}{1-\rho^2}\right) - 1}{\exp\left(\frac{\sigma_{\varepsilon}^2}{1-\rho^2}\right) - 1}.$$

Base on those expressions, the ML estimate for σ_{ε}^2 and $\rho_{\xi}(1)$ can be obtained.

Our results, as well as those reported by Zeger and Chan and Ledolter, are displayed in Table 2. In general, the estimates are fairly similar; however, a few differences should be pointed out:

1. The slope estimate and the coefficient of $\cos(2\pi t/6)$ in our model are bigger, in absolute value, than those reported by both Zeger and Chan and Ledolter. Zeger reported t-ratios of 1.62 and 0.14, which are not significant at the 0.05 level, but the one reported by Chan and Ledolter and ours are significant.

2. The coefficient of $\sin(2\pi t/6)$ reported by Zeger is bigger, in absolute value, than that reported Chan and Ledolter and in our model. Chan and Ledolter reported a t-ratio of 0.4 and we have a t-ratio of 0.7, which are not significant at the 0.05 level, but the one reported by Zeger is significant.

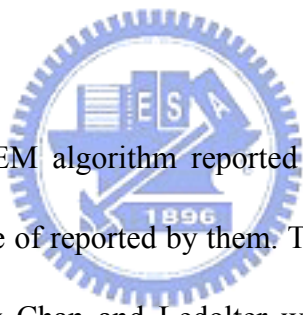
3. Our ML estimate of the noise variance is the maximum of the three methods, and is much bigger than both Zeger's estimate and the Chan and Ledolter's estimate.

4. Our ML estimate of the autoregressive coefficient is smaller than both Zeger's estimate and Chan and Ledolter's estimate. Moreover, our ML estimate is more farther from 1 than both their estimates.

5. Our standard errors for the ML estimates are smaller than the ones obtained by Zeger and similar with the ones obtained by Chan and Ledolter's.

Table 2 Coefficients and Standard Errors

	Zeger's model		Chan and Ledolter's model		Our model	
	$\hat{\theta}$	Standard error	$\hat{\theta}$	Standard error	$\hat{\theta}$	Standard error
Intercept	—	—	0.21	0.13	0.22	0.13
Trend	-4.35	2.68	-4.62	1.38	-4.80	1.40
$\cos(2\pi t/12)$	-0.11	0.16	0.15	0.09	0.14	0.09
$\sin(2\pi t/12)$	-0.48	0.17	-0.50	0.12	-0.53	0.12
$\cos(2\pi t/6)$	0.20	0.14	0.44	0.10	0.46	0.10
$\sin(2\pi t/6)$	-0.41	0.14	-0.04	0.10	-0.07	0.10
$\hat{\sigma}_\varepsilon^2$	0.77	—	0.54	0.04	0.95	0.04
$\hat{\rho}_\varepsilon$	0.77	—	0.88	—	0.57	0.06



We also applied the MCEM algorithm reported by Chan and Ledolter, and we have compared our results with those of reported by them. The starting values for ρ and σ_ε^2 are the same with the reported by Chan and Ledolter which are 0 and 1. The Markov chain sample size, m , is initially set to 200. The starting values of α are obtained by fitting a log-linear model to the data, assuming no temporal dependence, i.e. $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)'$ are $(0.554641, -4.843069, 0.136865, -0.541847, 0.455233, -0.059814)'$. The time sequence plot of the estimated log-likelihood (of the observed data) for 690 iteration is shown in Figure 2. After about 350th iterations, most coefficients stabilize. The log-likelihood is maximized at the 505th iteration and subsequently moves into a slightly lower likelihood region. Changes for selected coefficients are shown in Figure 3, Figure 4 and Figure 5.

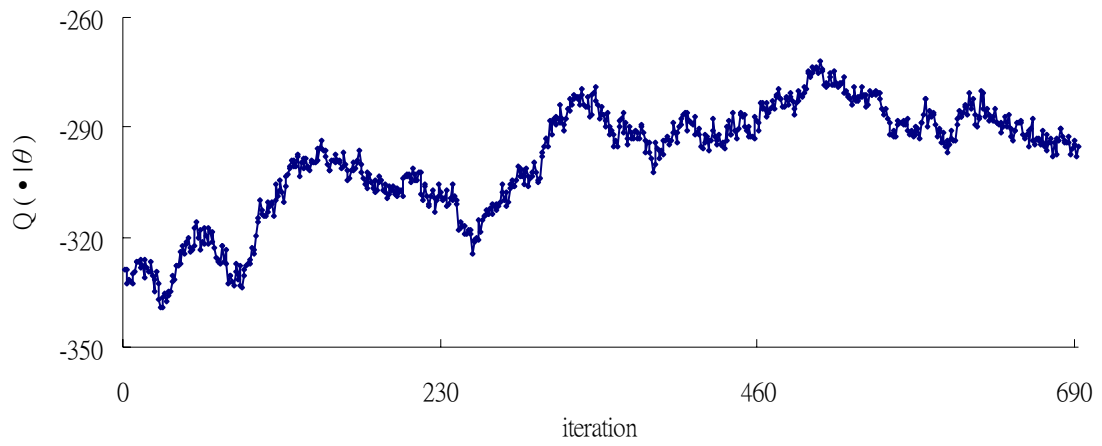


Figure 2 $Q(\bullet|\theta)$ for Iterations 0 Through 690

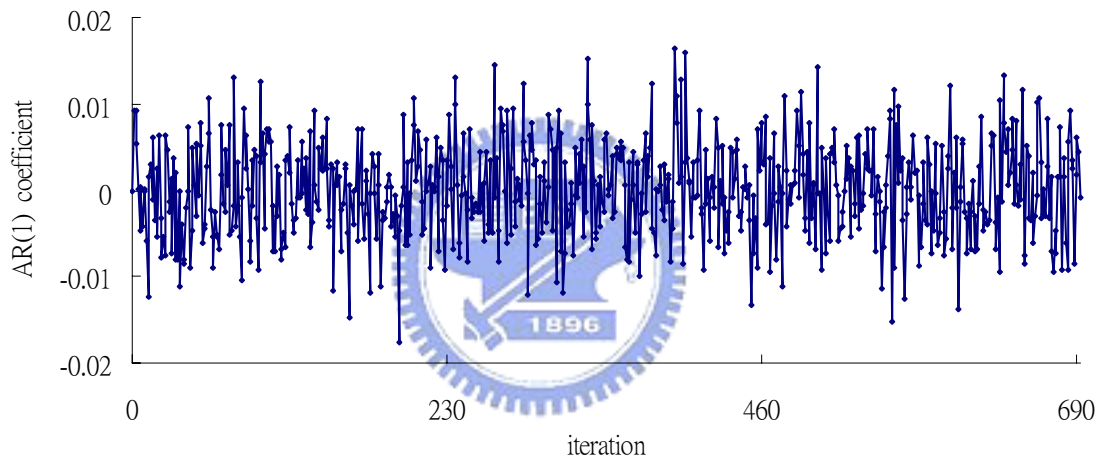


Figure 3 Change for AR(1) coefficient (ρ) estimates

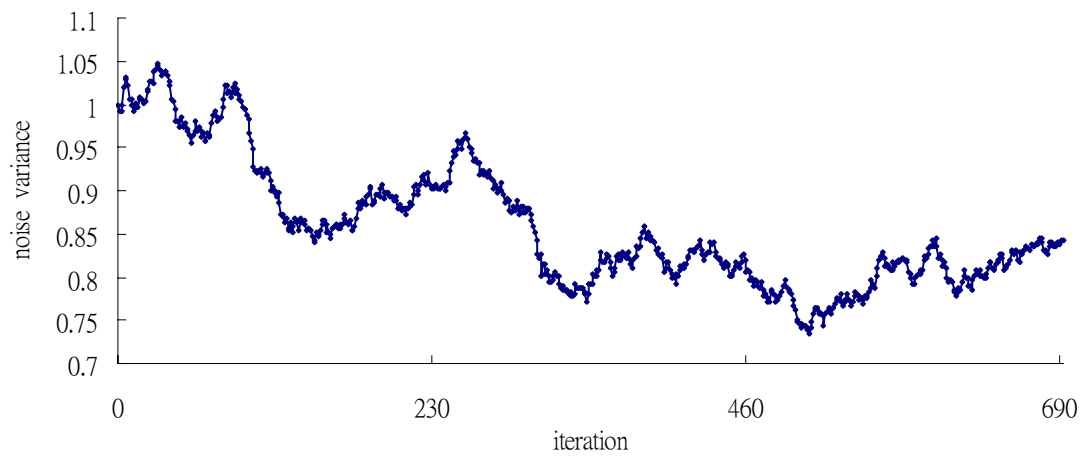


Figure 4 Change for noise variance (σ_ϵ^2) estimates

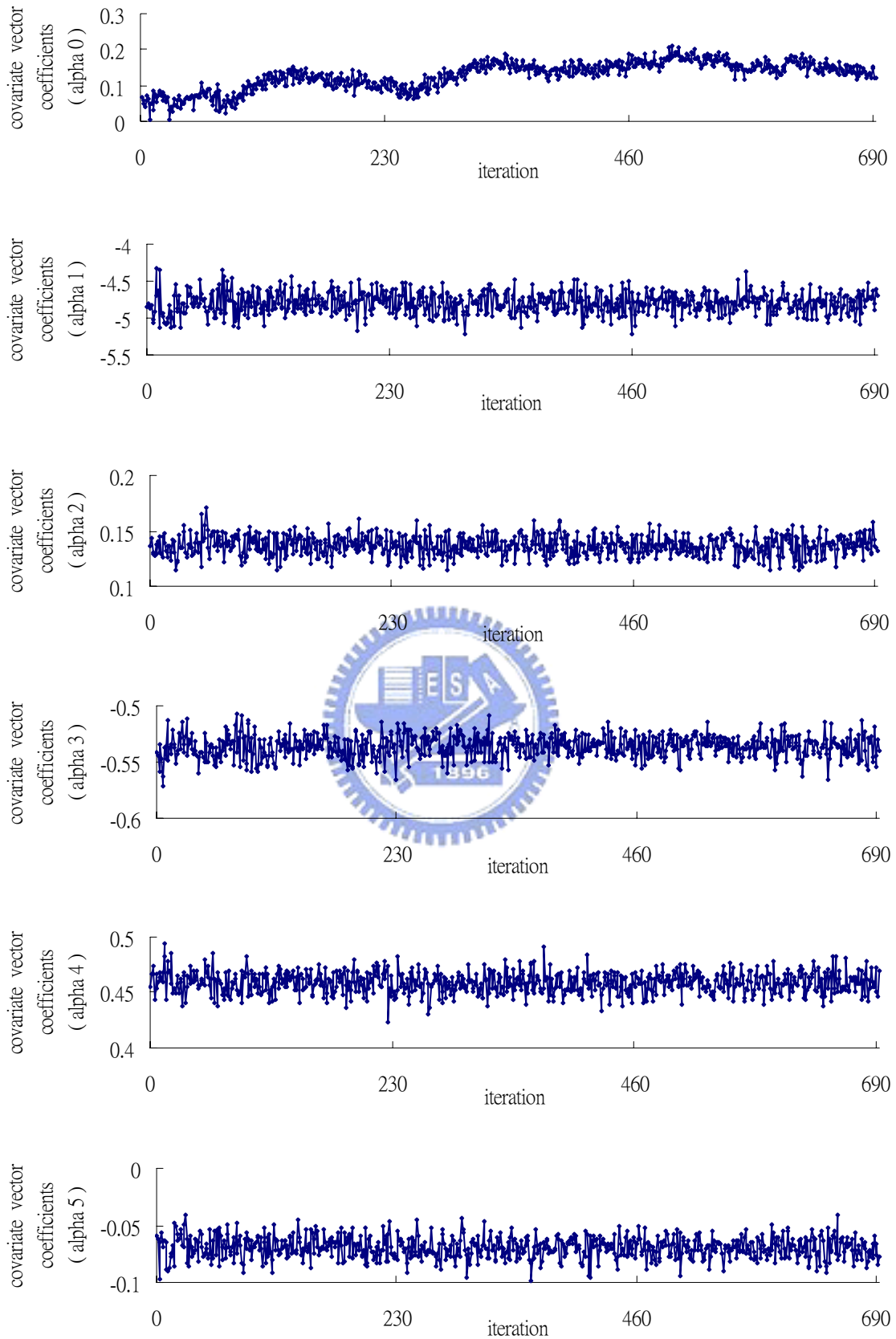


Figure 5 Change for covariate vector coefficients (α) estimates

The parameter values at the 505th iteration are used as starting values for a new MCEM iteration, i.e., $(\rho, \sigma_\varepsilon^2, \alpha')' = (-0.000208, 0.740943, 0.190889, -4.783508, 0.141291, -0.545433, 0.466094, -0.063966)'$. The Markov chain sample is increased to 2000, and time sequence plot of the estimated log-likelihood (of the observed data) for 473 iteration is shown in Figure 6. After 473 iterations, the algorithm is stopped. All estimated changes of the log-likelihood are smaller than 0.0005 in magnitudes. The approximate ML estimates are given by

$$\begin{aligned} \hat{\rho} &= 0.000489 & \hat{\sigma}_\varepsilon^2 &= 0.551862 \\ &(0.077382) & &(0.060386) \\ \hat{\alpha}' &= (0.266812 , -4.731944 , 0.134610 , -0.543219 , 0.456102 , -0.057427) \\ &(0.128069) (1.408738) (0.090010) (0.116141) (0.101956) (0.098602) \end{aligned}$$

Our results are not similar with that ones of reported by Chan and Ledolter, this may be the result of the difference between generators which produces the latent process $\{W_t\}$, thus the variation occurs in the MCE step is too large. In order to reduce such variations, we prefer the modified EM algorithm method.

Fitted results in Section 3.2 are shown in Figure 7, where dots represent the observations. The thicker solid line connects the fitted data simulated from method 1, and the average error with observations and fitted data is $\frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| = 1.145561$. The thinner solid line connects the fitted data simulated from method 2, and the average error with observations and fitted data is $\frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| = 1.601190$.

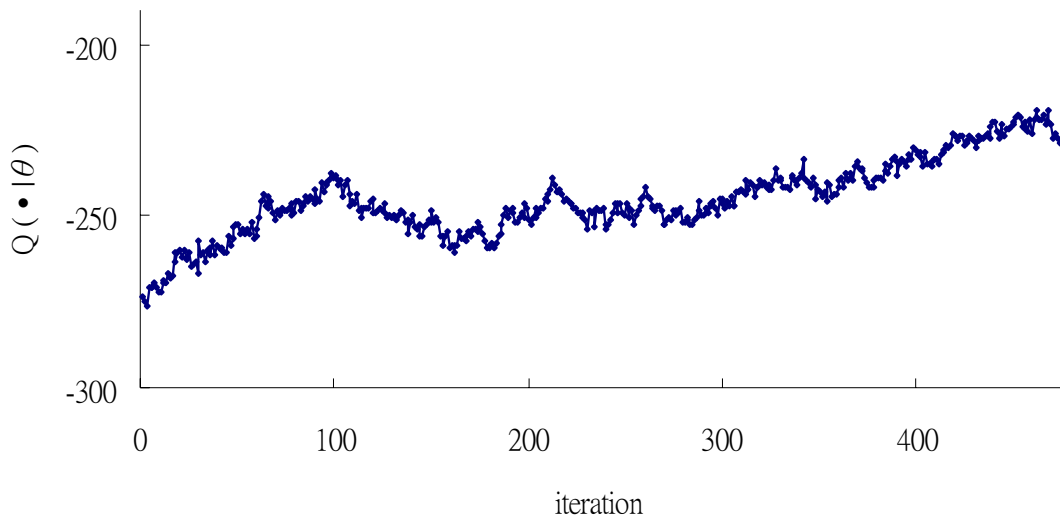


Figure 6 $Q(\bullet|\theta)$ for Iterations 0 Through 473

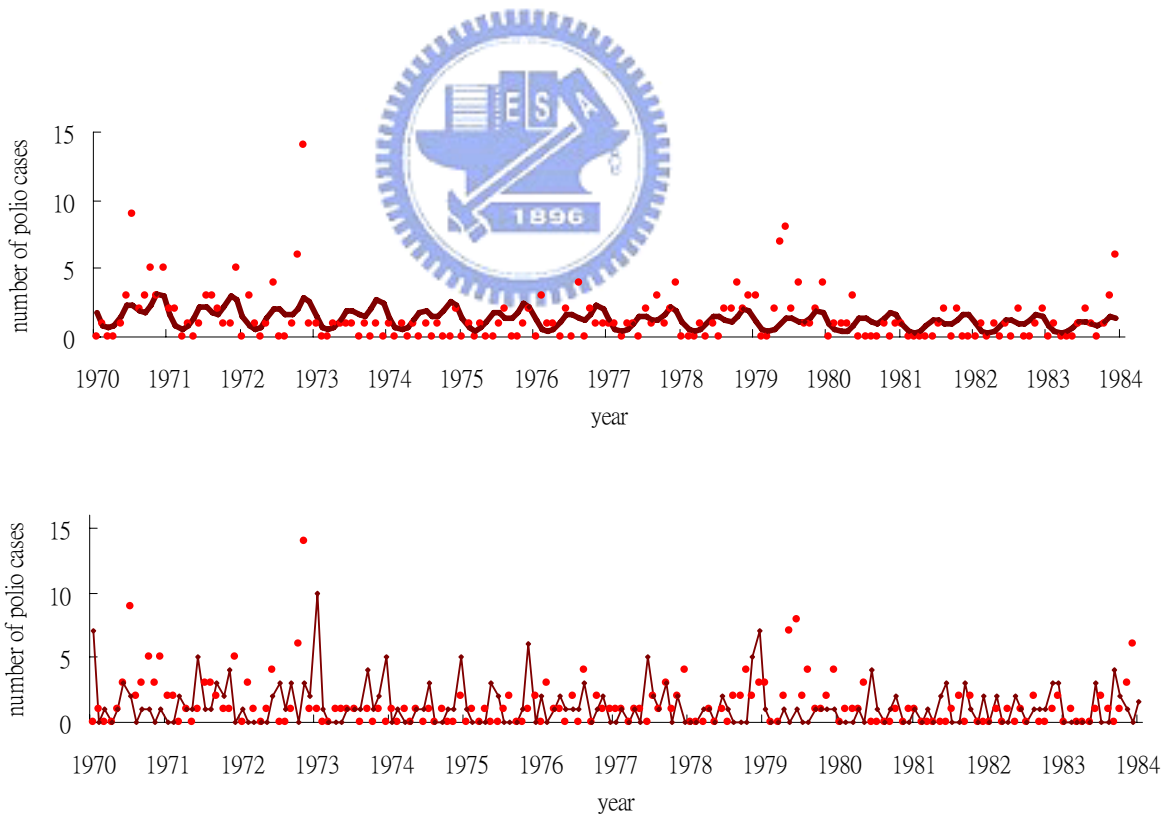


Figure 7 Monthly Numbers of Polio Cases. Dots represent observations. The thicker solid line connects the fitted data simulated from method 1. The thinner solid line connects the fitted data simulated from method 2.

4.2 Traffic Flow

4.2.1 Data Descriptions

Vehicle detectors are used for detecting the presentation of the queue. In some special time, like raining day or the day before long holidays, the level of oversaturated traffic condition will become more serious. Therefore, accurate forecast of traffic flow from the collected data is desirable. The data we analyze are collected by the RTMS (Remote Traffic Microwave Sensor) from the intersection of Zhong-Hua Road (中華路) and Jing-Guang Road (警光路), on Apr. 25, 2005 from 14:13'28 PM to 16:59'57 to study the effect on traffic before the rush hour. The data are taken at ten-second intervals; therefore, we have a data set of sample size 1000. The RTMS (Remote Traffic Microwave Sensor) is in its forward-looking configuration and is mounted on overhead sign structures to monitor the road and collect the data of traffic flow and occupancy. The detector divides the detectable range (about 60 meters) of the lane into six sections, and collects the volume and occupancy in each section separately. The RTMS mounted on the road is shown in Figure 8. Missing values occur with unknown reasons, the maximum volume of the six sections and the occupancy of the corresponding section constitute the data at each time interval. For example, when volume is 0 means that no car passes through the detector in ten seconds, and volume is 8 means that eight cars pass through the detector in ten seconds. If occupancy is 0, the percentage of occupancy is 0 %, and if occupancy is 1, the percentage of occupancy is 100 %. The traffic flow and occupancy change with time, and the scatter plot of the data collected by the RTMS is shown in Figure 9, the plots of the complete set and a partial set first 200 observations of traffic flow are shown in Figure 11 and Figure 12, respectively. Because of undetectable

situation caused by traffic jam or some unidentified reason, we consider the existence of latent process.

The data of the traffic flow collected by RTMS are discrete data, a model for regression analysis with a time series of counts will be used. The notations used in the model are listed as follows:

Y : traffic flow (volume); this value is the variable of the detected flow, and the definition is the number of vehicles passing a point on a roadway, or other traffic way during some time interval;

u : occupancy; the percentage of flow at which vehicles traverse a point or uniform segment of a lane or roadway during a specified time period under prevailing roadway, traffic, and control conditions;

N : the number of data.

i.e.,

$$\mathbf{Y} = (y_1, y_2, \dots, y_t, \dots, y_{N-1}, y_N),$$

$$\mathbf{u} = (u_1, u_2, \dots, u_t, \dots, u_{N-1}, u_N),$$

and (y_t, u_t) is the data at time t .

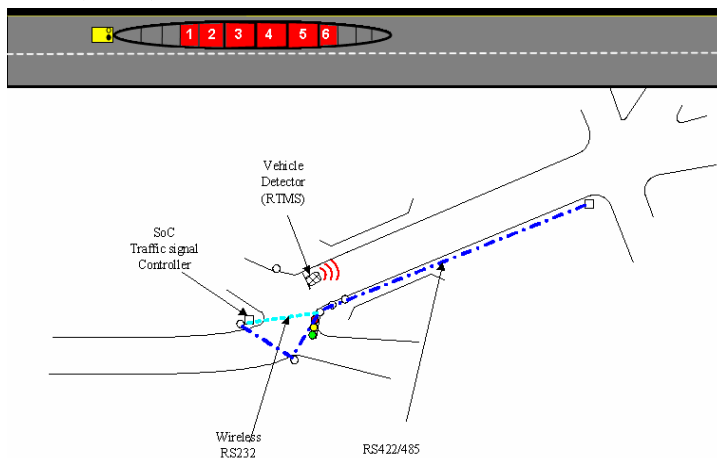


Figure 8 RTMS in Forward-Looking Mode

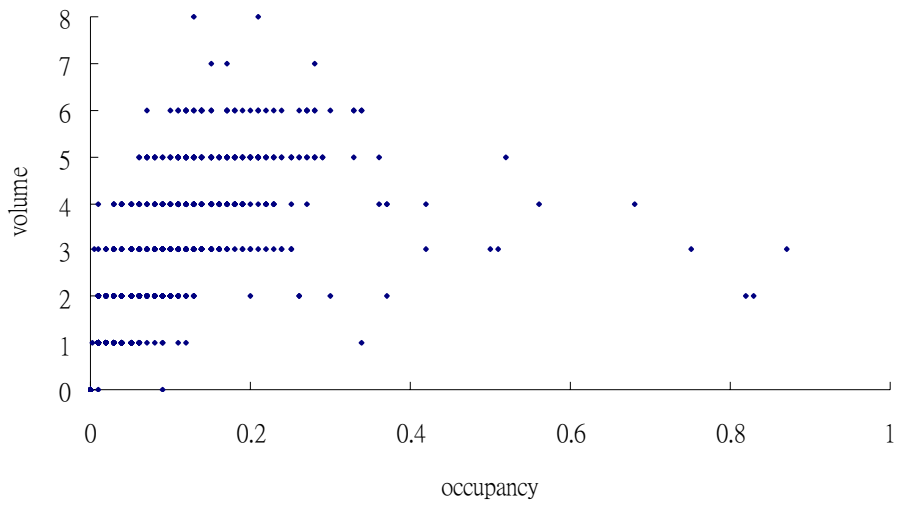
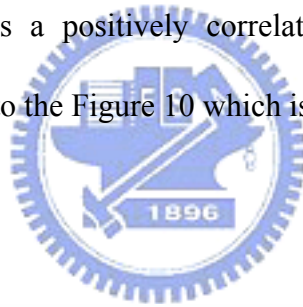


Figure 9 Occupancy-Volume Plot. The majority of the data set occurs at the small values of occupancy and roughly shows a positively correlated relationship between volume and occupancy. The plot conforms to the Figure 10 which is the relationship of flow and density in Traffic Theory.



(density, flow) diagram

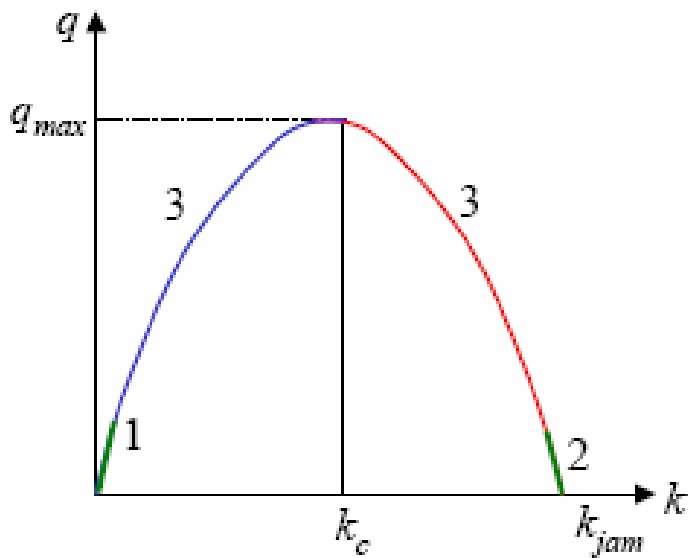


Figure 10 Flow (q) and Density (k); Density = constant \times Occupancy

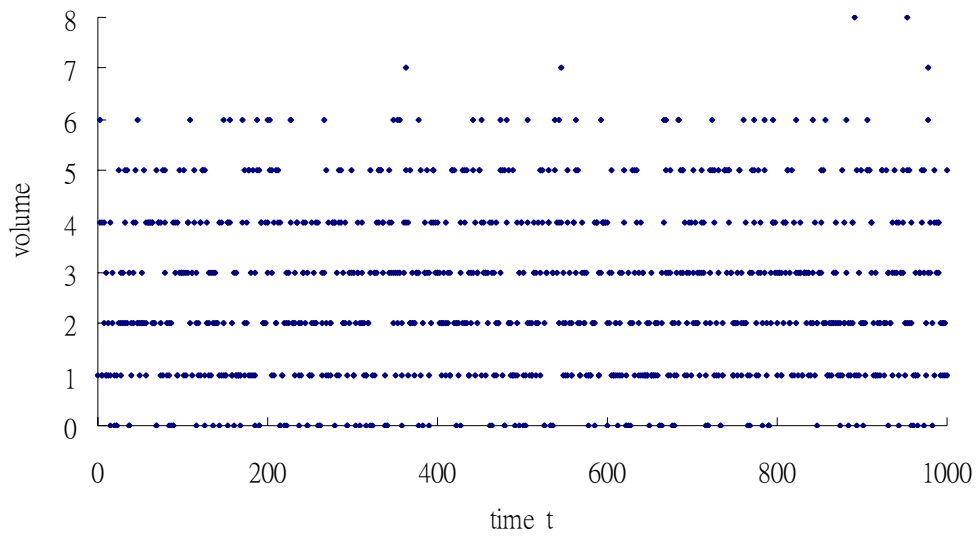


Figure 11 Data of Traffic Flow (Complete Data)

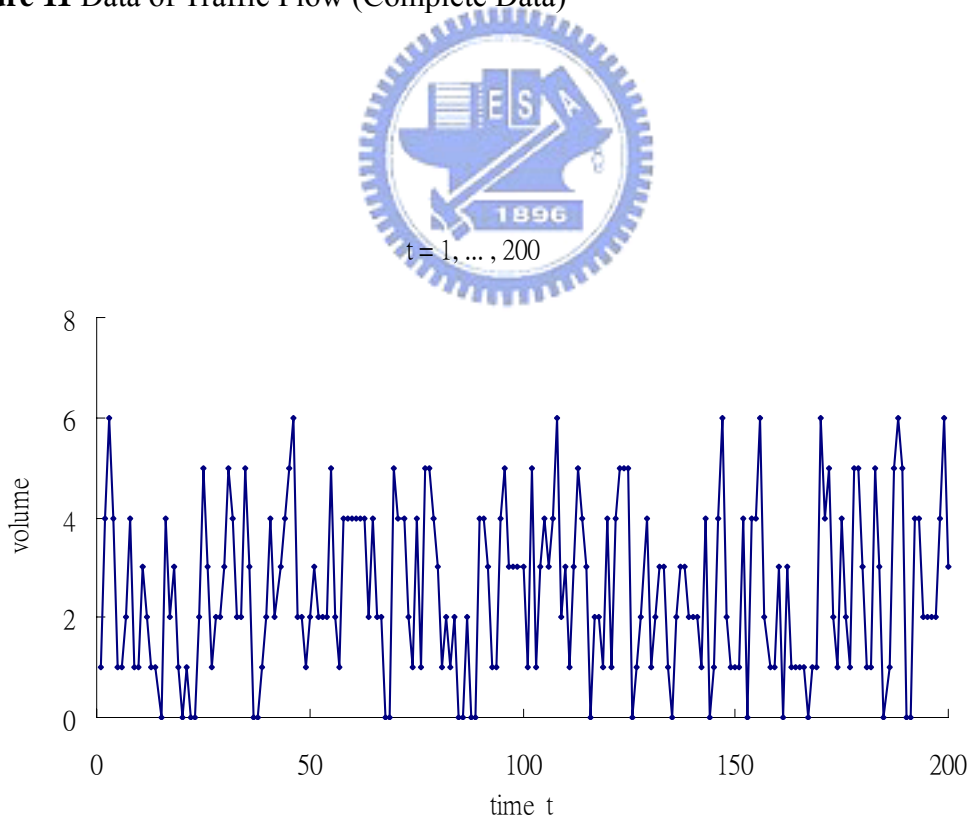


Figure 12 Data of Traffic Flow (First 200 time sequence of Complete Data)

4.2.2 Data Analysis

We have applied the parameter-driven model to analysis traffic flow as an illustration. A time sequence plot of the traffic flow is shown in Figure 11. Because the sample size is large, we can't read Figure 11 clearly and Figure 12 shown the plot of first 200 observations plot of the complete data. The covariate vector in Equation (3) is given by $U_t = (1, u_t)'$ where p is 2. We adopt the latent AR(1) model described in the previous section, and we used both MCEM and the modified EM algorithm to carry out the estimation. The starting values for ρ and σ_ε^2 are 0.5 and 0.5. The starting values of α are obtained by fitting a log-linear model to the data, assuming no temporal dependence, i.e. $(\alpha_0, \alpha_1)'$ are $(0.682636, 2.550254)'$.

The value of $Q(\bullet | \theta)$ is maximized and the approximate ML estimates are given by

$$\hat{\rho} = 0.707559 \quad \hat{\sigma}_\varepsilon^2 = 0.332942 \quad \hat{\alpha}' = (0.350308, 2.580916)$$

$$(0.022361) \quad (0.014899) \quad (0.025593) \quad (0.129314)$$

The figures within parentheses are estimated standard errors, obtained from Equation (16).

Fitted results in Section 3.2 are shown in Figure 13 and Figure 14, where dots represent the observations. The thinner solid line connects the fitted data simulated from method 1, and the average error between observations and fitted data is $\frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| = 1.181$. The thicker solid line connects the fitted data simulated from method 2, and the average error between observations and fitted data is $\frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| = 2.277$. We can find that the average error between observations and fitted data from method 1 is smaller than that obtained by method 2, this unstable situation may come from the variability with the latent process $\{W_t\}$.

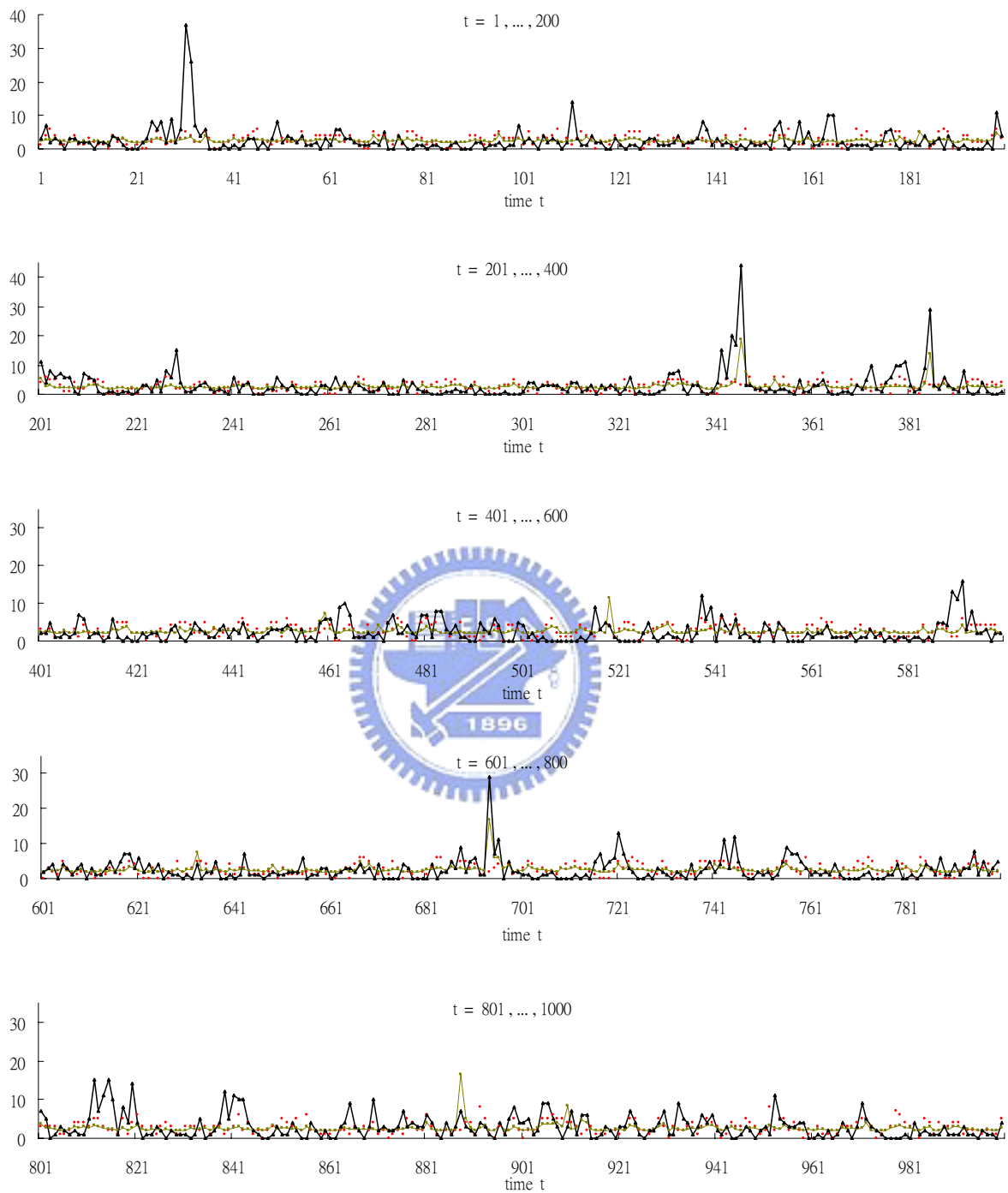


Figure 13 Time sequence of the traffic flow. Dots represent observations. The thinner solid line connects the fitted data simulated from method 1. The thicker solid line connects the fitted data simulated from method 2.

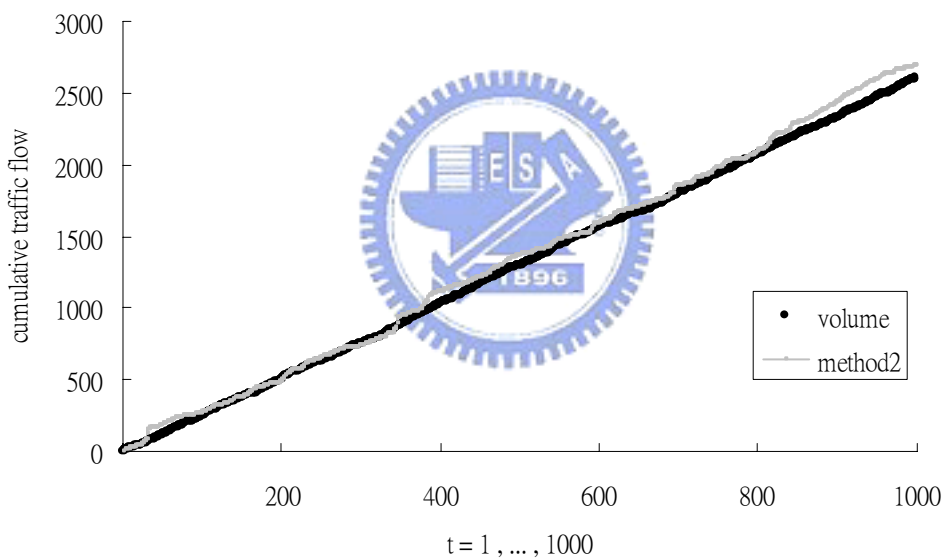
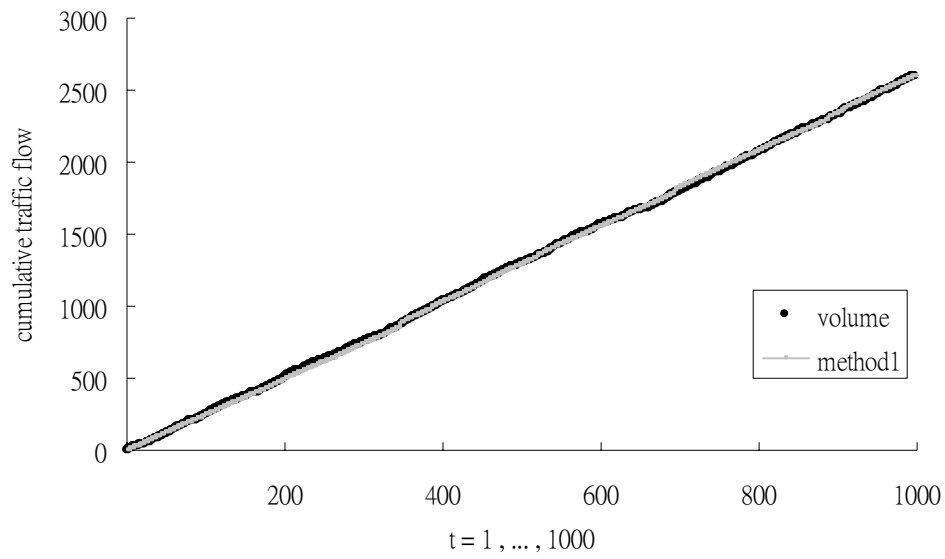


Figure 14 Cumulative Traffic Flow. Dots represent observations. The upper figure shows the fitted data simulated from method 1. The lower figure shows the fitted data simulated from method 2. It is clear that the fitted data from method 1 is closer the traffic flow than that one from method 2, this unstable situation maybe come from generator which produces the latent process $\{W_i\}$.

We apply the MCEM algorithm reported by Chan and Ledolter (1995) to our data. The starting values for ρ and σ_ε^2 are the same with the reported by them which are 0 and 1. The Markov chain sample size, m , is initially set to 200. The starting values of α are obtained by fitting a log-linear model to the data, assuming no temporal dependence, i.e. $(\alpha_0, \alpha_1)'$ are $(0.682636, 2.550254)'$. The time sequence plot of the estimated log-likelihood (of the observed data) for 150 iteration is shown in Figure 15. After about 90 iterations, all coefficients stabilize. The log-likelihood is maximized at the 132nd iteration and subsequently moves into a slightly lower likelihood region. Changes for selected coefficients are shown in Figure 16, Figure 17 and Figure 18.

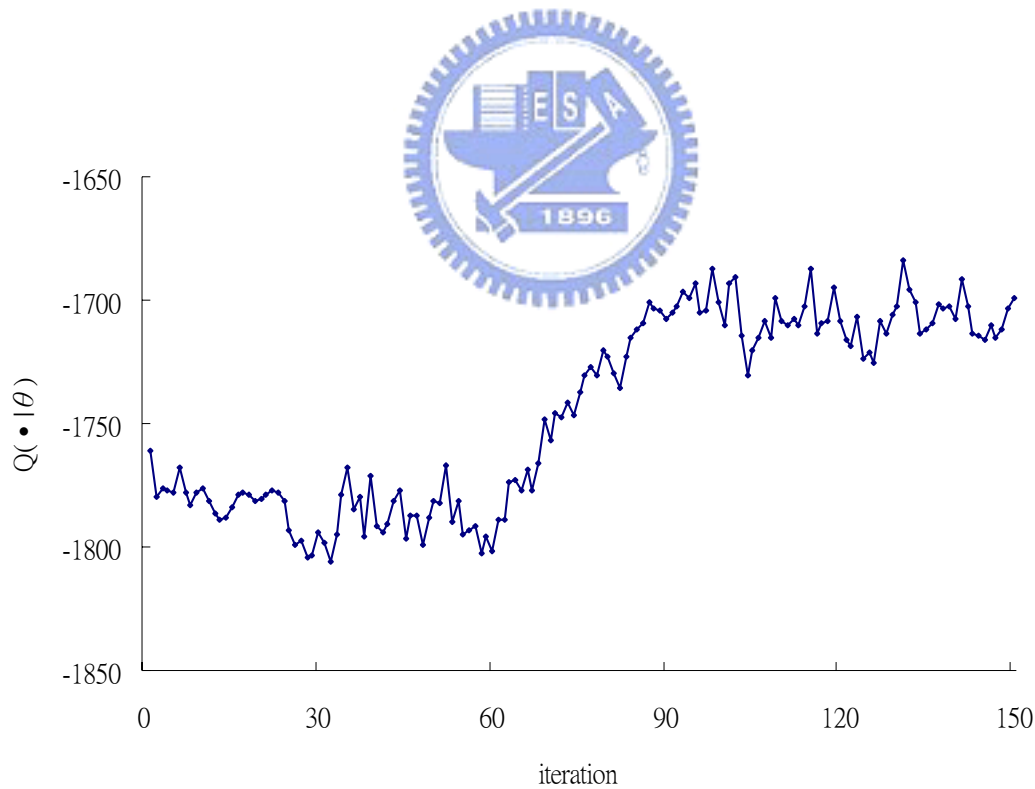


Figure 15 $Q(\bullet|\theta)$ for Iterations 0 Through 150

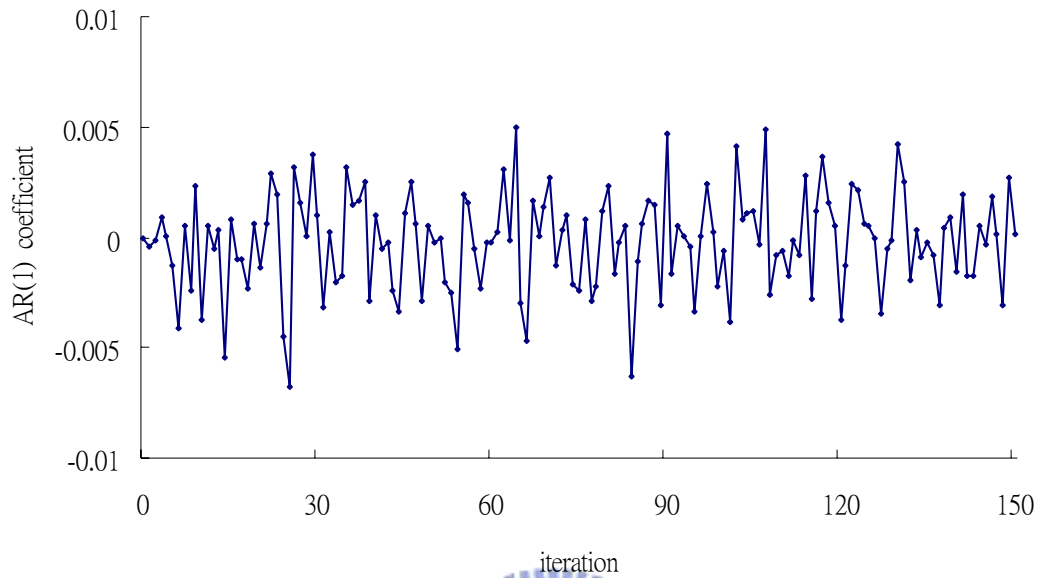


Figure 16 Change for AR(1) coefficient (ρ) estimates

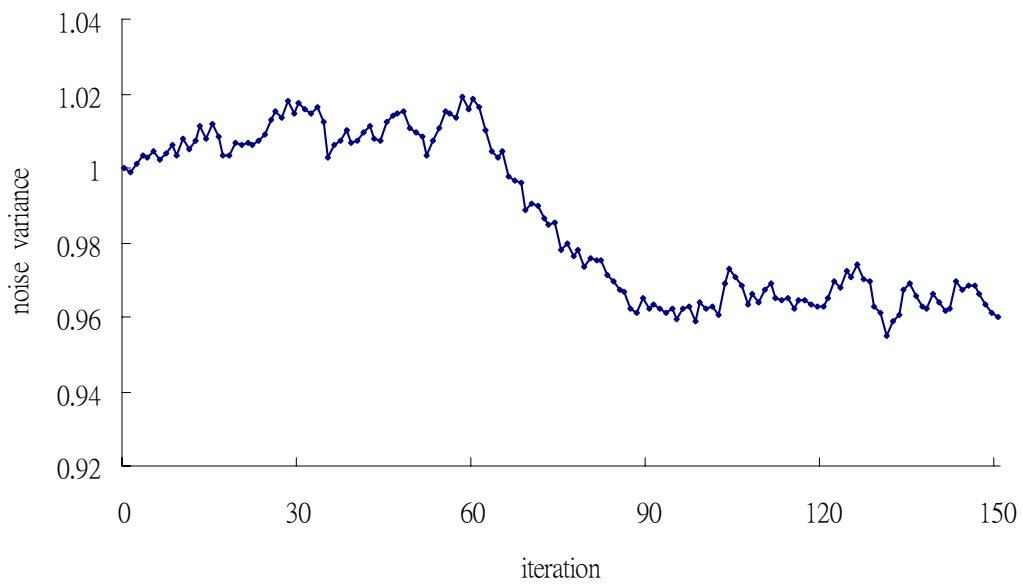


Figure 17 Change for noise variance (σ_ε^2) estimates

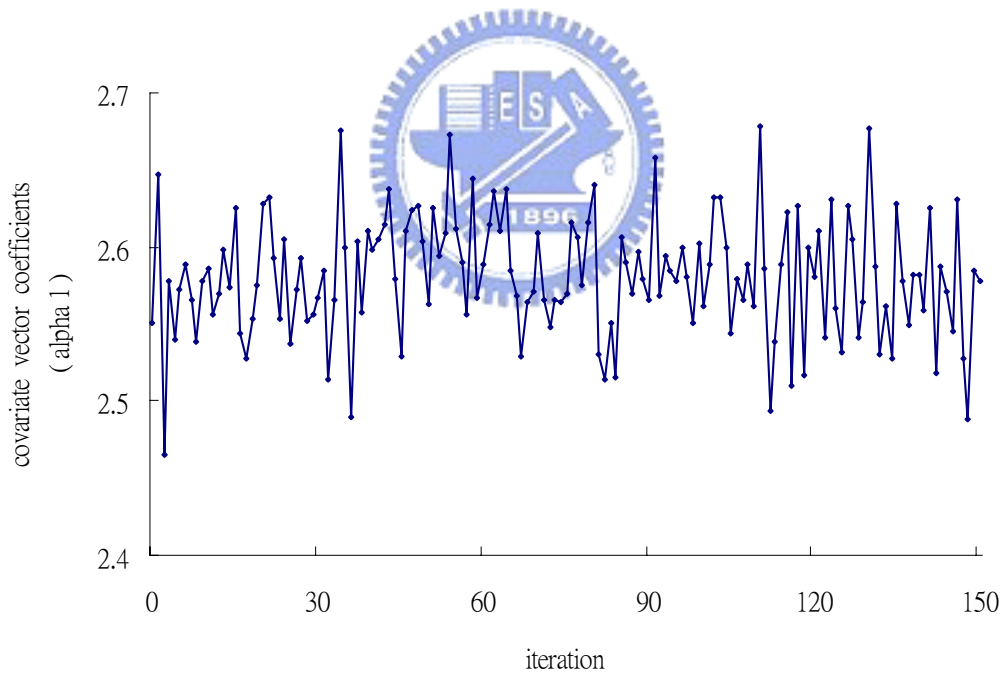
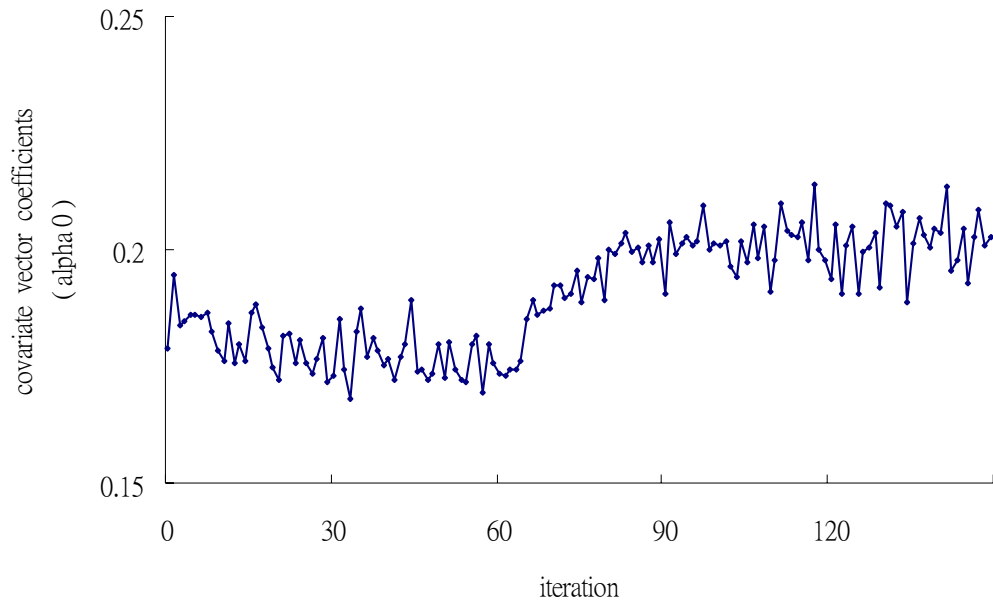


Figure 18 Change for covariate vector coefficients (α) estimates

The parameter values at the 132nd iteration are used as starting values for a new MCEM iteration, i.e., $(\rho, \sigma_\varepsilon^2, \alpha')' = (-0.001962, 0.958849, 0.209393, 2.530430)'$. The Markov chain sample is increased to 2000, and time sequence plot of the estimated log-likelihood (of the observed data) for 147th iteration is shown in Figure 19. After 147 iterations, the algorithm is stopped. All estimated changes of the log-likelihood are smaller than 0.0005 in magnitudes at 147th iteration. The approximate ML estimates are given by

$$\begin{array}{lll} \hat{\rho} = 0.001183 & \hat{\sigma}_\varepsilon^2 = 0.913263 & \hat{\alpha}' = (0.231416 , 2.578988) \\ (0.031639) & (0.040863) & (0.025537) (0.129102) \end{array}$$

Our fitted results in Section 3.2 are shown in Figure 20 and Figure 21, where dots represent the observations. The average error with observations and fitted data simulated from method 1, is $\frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| = 1.181$, and the average error with observations and fitted data simulated from method 2 is $\frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| = 2.493$. We can find that the average error with observations and fitted data from method 1 is smaller than that one from method 2.

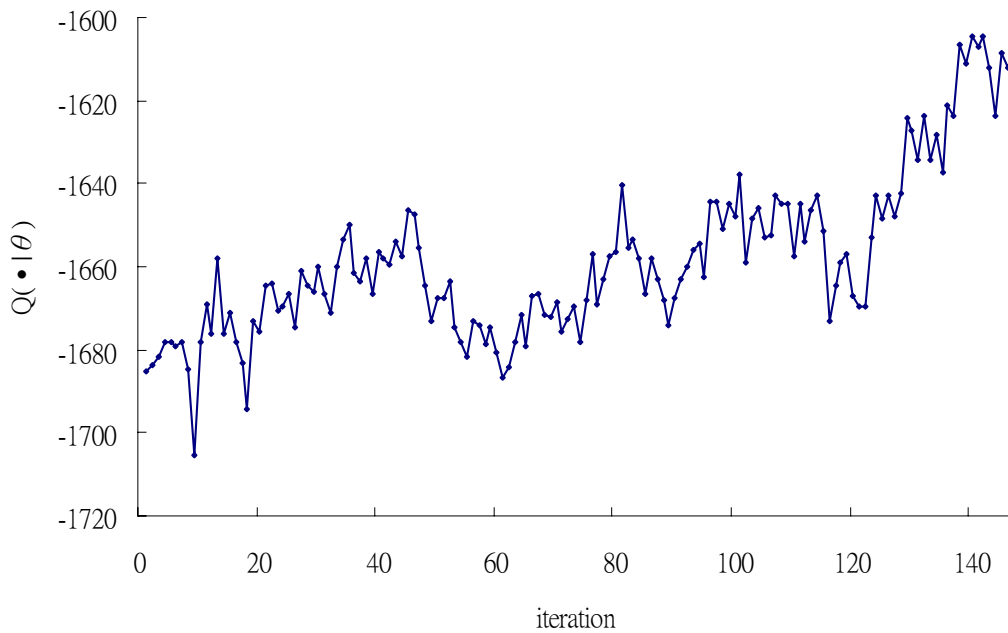


Figure 19 $Q(\bullet|\theta)$ for Iterations 0 Through 147

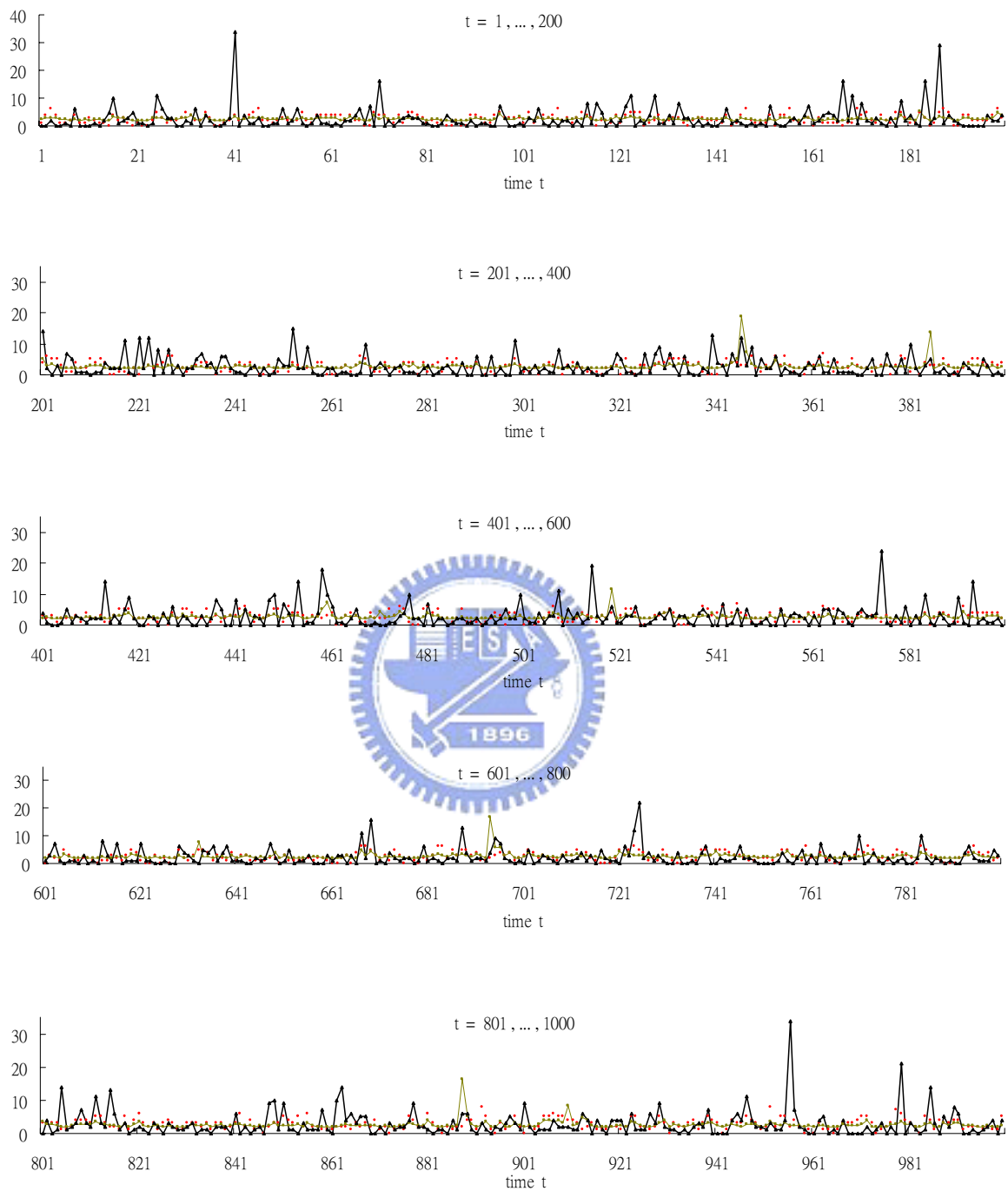


Figure 20 Time sequence of the traffic flow. Dots represent observations. The thinner solid line connects the fitted data simulated from method 1. The thicker solid line connects the fitted data simulated from method 2.

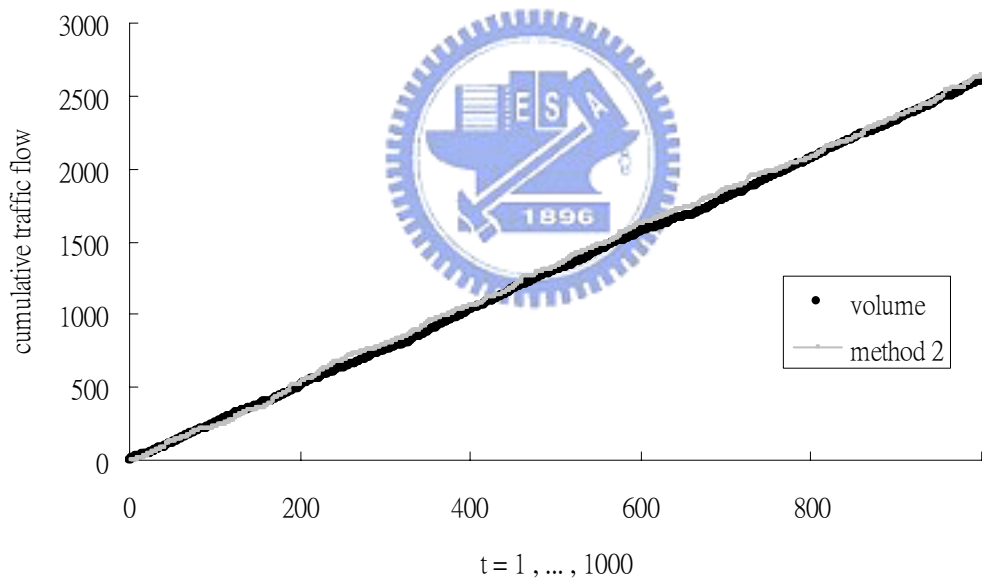
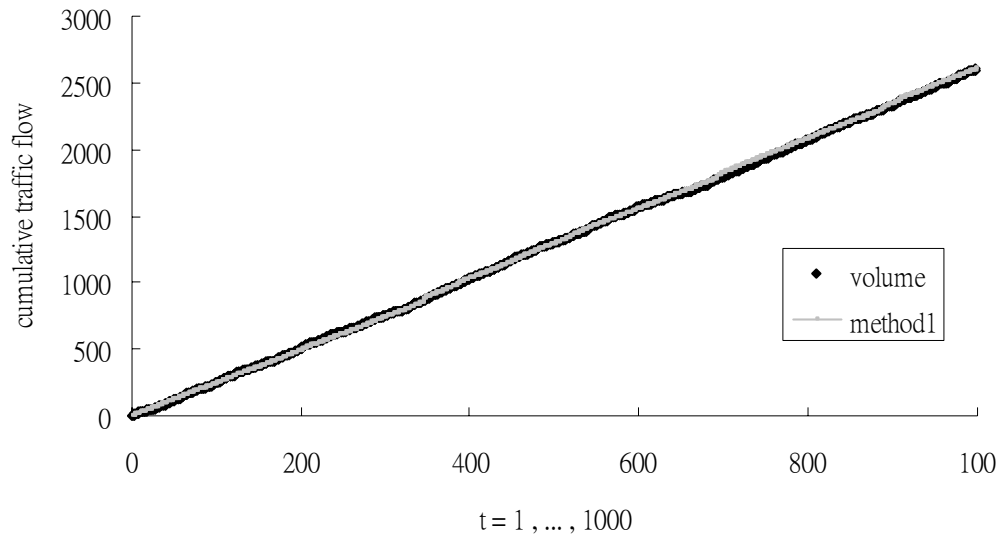


Figure 21 Cumulative Traffic Flow. Dots represent observations. The upper figure shows the fitted data simulated from method 1. The lower figure shows the fitted data simulated from method 2. It is clear that the fitted data from method 1 is closer the traffic flow than that one from method 2, this unstable situation maybe come from generator which produces the latent process $\{W_i\}$.

5 Conclusion

In this thesis, a generalized regression model for a time series of count data is used to model the relationship between the traffic flow and the occupancy, assuming the existence of an AR(1) Gaussian latent process and the modified EM algorithm. Our method and our results, a few phenomena should be pointed out:

1. If the log-likelihood function is concave and unimodal, then the sequence of iterates $\theta^{(r)}$ ($\theta^{(r)}$ is the estimate at the r^{th} iteration) converges to the ML estimate $\hat{\theta}$ of θ , in one step if the log-likelihood is a quadratic function of θ .

2. The starting value for ρ of the modified EM is 0.5 that we chose. We also tried the value closer to 0 or 1, in absolute value, and problems come up more easily during the solving process.

3. No matter which method we use, some very large fitted values pump up in Figure 13. This is because when we fitted the data, it has some relationship with the covariate variable. It is observed that fitted data could be affected by observed occupancy seriously, and the average error with observations and fitted data from method 1 is smaller than that from method 2, this unstable situation may have come from generator which produces the latent process $\{W_i\}$.

Our method and the MCEM algorithm reported by Chan and Ledolter are a few differences shown as follows:

We replace the marginal expectation which has some relation with latent process, and the variation caused by simulation could be reduced. Another problem to be overcome is that we may spend much time and need large memory as simulation.

Appendix : Proof of the *E step*

Proof of $E_t(\exp(W_t)) = \exp\left(\frac{\sigma_\varepsilon^2}{2(1-\rho^2)}\right)$. The marginal expectation of $\exp(W_t)$ is

found by making use of the following conditional expectation

$$E_{t|t-1}(\exp(cW_t)) = \exp\left(c\rho W_{t-1} + \frac{c^2\sigma_\varepsilon^2}{2}\right)$$

and

$$E_0(\exp(cW_0)) = \exp\left(\frac{c^2\sigma_\varepsilon^2}{2(1-\rho^2)}\right) \text{ where } c \text{ is a constant.}$$

Now,

$$\begin{aligned} E_t(\exp(W_t)) &= E_{t-1}E_{t|t-1}(\exp(W_t)) \\ &= E_{t-1}\left(\exp\left(\rho W_{t-1} + \frac{\sigma_\varepsilon^2}{2}\right)\right) \end{aligned}$$

Repeat the same calculation procedure and we can derive

$$E_t(\exp(W_t)) = \exp\left(\frac{\sigma_\varepsilon^2}{2(1-\rho^2)}\right). \quad \blacksquare$$

Proof of $E_t(Y_t W_t) = 0$. The marginal expectation of $Y_t W_t$ is found by making use of the following conditional expectation and marginal expectation

$$E_{t|t-1}(Y_t W_t) = (1-\sigma_\varepsilon^2)^{-3/2} \rho W_{t-1} \exp\left\{\frac{\rho^2}{1-\sigma_\varepsilon^2} W_{t-1}^2 + \alpha' U_t\right\}$$

and

$$E_t(W_t \exp\{c W_t^2\}) = 0 \text{ where } c \text{ is a constant.}$$

First, the derivative processes of conditional expectation are as follows:

$$\begin{aligned}
E(y_t W_t) &= \int \int y_t W_t \frac{e^{-W_t + \alpha' U_t} (e^{W_t + \alpha' U_t})^{y_t}}{y_t!} \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}(W_t - \rho W_{t-1})^2\right\} dy_t dW_t \\
&= \int W_t \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}(W_t - \rho W_{t-1})^2\right\} \int y_t \frac{e^{-W_t + \alpha' U_t} (e^{W_t + \alpha' U_t})^{y_t}}{y_t!} dy_t dW_t \\
&= \int W_t \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}(W_t - \rho W_{t-1})^2\right\} \times e^{W_t + \alpha' U_t} dW_t \\
&= \frac{1}{\sqrt{1-2\sigma_\varepsilon^2}} \exp\left\{\frac{\rho^2}{1-2\sigma_\varepsilon^2} W_{t-1}^2 + \alpha' U_t\right\} \\
&\quad \times \int W_t \frac{1}{\sqrt{2\pi} \sqrt{\frac{\sigma_\varepsilon^2}{1-2\sigma_\varepsilon^2}}} \exp\left\{-\frac{1}{2} \frac{\sigma_\varepsilon^2}{1-2\sigma_\varepsilon^2} \left(W_t - \frac{\rho}{1-2\sigma_\varepsilon^2} W_{t-1}\right)^2\right\} dW \\
&= (1-\sigma_\varepsilon^2)^{-3/2} \rho W_{t-1} \exp\left\{\frac{\rho^2}{1-\sigma_\varepsilon^2} W_{t-1}^2 + \alpha' U_t\right\}
\end{aligned}$$

Second, the derivations of marginal expectation are as follows:

$$E_t(W_t \exp\{c W_t^2\}) = E_{t-1} E_{t|t-1}(W_t \exp\{c W_t^2\})$$

Repeat the same calculation procedure and $E_0(W_0 \exp\{c W_0^2\}) = 0$, we can derive

$$E_t(W_t \exp\{c W_t^2\}) = 0$$

Now,

$$E_t(Y_t W_t) = E_{t-1} E_{t|t-1}(Y_t W_t)$$

$$= E_{t-1} \left((1-\sigma_\varepsilon^2)^{-3/2} \rho W_{t-1} \exp\left\{\frac{\rho^2}{1-\sigma_\varepsilon^2} W_{t-1}^2 + \alpha' U_t\right\} \right).$$

Repeat the same calculation procedure and we can derive

$$E_t(Y_t W_t) = 0. \quad \blacksquare$$

Proof of $E_t(W_t^2) = \frac{(1-\rho^{2(t+1)})}{(1-\rho^2)} \sigma_\varepsilon^2$. The marginal expectation of W_t^2 is found by

making use of the following conditional expectation

$$E_{t|t-1}(W_t^2) = \exp(\rho^2 W_{t-1}^2 + \sigma_\varepsilon^2)$$

and

$$E_0(W_0^2) = \sigma_\varepsilon^2.$$

Now,

$$\begin{aligned} E_t(W_t^2) &= E_{t-1} E_{t|t-1}(W_t^2) \\ &= E_{t-1}(\exp(\rho^2 W_{t-1}^2 + \sigma_\varepsilon^2)). \end{aligned}$$

Repeat the same calculation procedure and we can derive

$$E_t(W_t^2) = \frac{(1 - \rho^{2(t+1)})}{(1 - \rho^2)} \sigma_\varepsilon^2. \quad \blacksquare$$

Proof of $E_t(W_t W_{t+1}) = \frac{\rho}{(1 - \rho^2)} \sigma_\varepsilon^2$. The marginal expectation of $W_t W_{t+1}$ is found by making use of the following relation

$$Cov(W_t, W_{t+1}) = \frac{\rho}{(1 - \rho^2)} \sigma_\varepsilon^2$$

and

$$E_t(W_t) = 0.$$

Now,

$$\begin{aligned} E_t(W_t W_{t+1}) &= Cov(W_t, W_{t+1}) - E_t(W_t) E_{t+1}(W_{t+1}) \\ &= \frac{\rho}{(1 - \rho^2)} \sigma_\varepsilon^2. \quad \blacksquare \end{aligned}$$

Reference

- Al-Osh, M. A. and Alzaid, A. A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, Vol. 8, No. 3, 261-275.
- Bent, J. , Soren, L. C., Song, P. X. K. and Li, S. (1999). A State Space Model for Multivariate Longitudinal Count Data. *Biometrika*, Vol. 86, No. 1, 169-181.
- Brannas, K. & Johansson, P. (1994). Time series count data regression. *Commun. Statist.-Theory and Meth.* Vol. 23, No. 10, 2907-2926.
- Campbell, M. J. (1994). Time series regression for counts: an investigation into the relationship between Sudden Infant Death Syndrome and environmental temperature. *Journal of Royal Statistical Society. Series A* , Vol. 157, No. 2, 191-208.
- Chan, K. S. and Ledolter, J. (1995). Monte Carlo EM Estimation for Time Series Models Involving Counts. *Journal of the American Statistical Association*, Vol. 90, No. 429, 242-252.
- Cho, H. J. and Tseng, M. T.. A Level Set Approach to Oversaturated Traffic Network Simulation. Department of Transportation Technology and Management, National Chiao Tung University, Taiwan.
- Cox, D. R. (1981). Statistical analysis of time series: some recent developments (with discussion). *Scandinavian Journal of Statistics*, 8, 93-115.
- Davis, R. A., Dunsmuir, W. T. M. and Streett, S. B. (2003). Observation-driven models for Poisson counts. *Biometrika*, Vol. 90, no. 4, 777-790.
- Davis, R.A., Dunsmuir, W. T. M., and Wang, Y. (2000) On autocorrelation in a Poisson regression model. *Biometrika*, Vol. 87, No. 3, 491-505.
- Freeland, R. K. and McCabe, B. P. M. (2004). Analysis of Low Count Time Series Data by Poisson Autoregression. *Journal of Time Series Analysis*, Vol. 25, No.5, 701-722.
- Jorgensen, B., Lundbye, C. S., Song. X. K. and Sun. L. (1996). A longitudinal study of emergency room visits and air pollution for Prince George, British Columbia. *Statistics in Medicine*. Vol. 15, 823-836.
- Liu, H.C. and Kuwahara, M. A Study on Real-Time Signal Control for an Oversaturated

Network. Institute of Industrial Science, University of Tokyo, 4-6-1 komaba, Meguro-ku, 153-8505 Tokyo, Japan.

Louis, T. A. (1982). Finding the Observed Information Matrix when Using the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 44, No. 2, 109-286.

McKenzie, E. (2000). *Discrete Variate Time Series*. Department of Statistics & Modelling Science, University of Strathclyde, 15th August.

Meng, X. L. and Rubin, D. B. (1991). Using EM to Obtain Asymptotic Variance- Covariance Matrices: The SEM Algorithm. *Journal of the American Statistical Association*, Vol. 86, No. 416, 899-909.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A*, Vol. 135, 370-384.

Zeger, S. L. (1988). A Regression Model for Time Series of Counts. *Biometrika*, 75, 621-629.

