國 立 交 通 大 學

統計學研究所

碩 士 論 文

動態Tobit模式之估計

Estimation of Dynamic Tobit Models by Newton's Method and SML-GHK Simulator

研 究 生 ：張宛茹

指導教授 ：周幼珍 博士

中 華 民 國 九 十 五 年 六 月

動態Tobit模式之估計

Estimation of Dynamic Tobit Models by Newton’s Method and SML-GHK Simulator

研　究　生：張宛茹　　Student： Wan-Ru Chang
指導教授：周幼珍　　Advisor： Dr. Yow-Jen Jou

國　立　交　通　大　學
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Tobit模式之估計

研究生：張宛茹　　　　　　指導教授：周幼珍 博士

國立交通大學統計學研究所

摘要

塞車是日常生活中常出現的問題，預測塞車時的車流量成為交通管理上的核心目標；我們希望預測塞車時的車流量，尤其是累積車流量的預測，將有助於交通號誌的設計與控制，進而紓解塞車情況。本文考慮一組實際偵測的交通資料，其中包含車流量與佔有率，針對有截斷（censored）且有延遲（lagged）的特性資料，提出兩種估計方法，第一是針對 Poisson 迴歸動態 Tobit 模式採取牛頓法進行參數迭代，第二是針對有潛在獨立變數的動態 Tobit 模式使用 SML-GHK 模擬法，模擬潛在變數及概似函數求得參數的最大模擬概似估計式。

Estimation of Dynamic Tobit Models by Newton's Method and SML-GHK Simulator

Student：Wan-Ru Chang　　Advisor：Dr. Yow-Jen Jou

Institute of Statistic
National Chiao Tung University

ABSTRACT

Traffic congestion is a serious problem in everybody's daily life. Prediction of traffic flow is becoming the kernel of the traffic management for oversaturated conditions. Our goal is to forecast the traffic flow, especially predict the cumulative traffic flow. This thesis considers the estimation of the parameters of censored data models with lagged dependent variable. Two methods have been proposed. NR algorithm is supplied for solving the Poisson regression dynamic Tobit model. The SML-GHK simulator is another feasible method for the estimation of dynamic Tobit model whose lagged dependent variable is latent.

誌　謝

首先，要感謝我的指導教授周幼珍老師，老師的耐心指導與付出，讓我對研究方向的架構越來越清楚，這篇論文才能夠順利完成；也謝謝老師在課業及生活上給予我許多正面的建議與鼓勵。接著，感謝運管所的卓訓榮老師與曾明德學長，在論文的資料蒐集與熟悉給予我最大的支援與幫助。也感謝口試委員周若珍、李昭勝和許英麟老師幫我校稿並提供我的寶貴意見與指教。

研究所這兩年，要謝謝我的夥伴鸞筑，謝謝妳的體貼與細心，做計畫、討論課業、當助教改作業、甚至分享生活點滴，都好高興有妳一同作伴。另外也結交了一群好姊妹，鸞筑、小宛、孟樺、秀慧、沛君、婉文與小馬，在妳們身上讓我感受到友誼的溫暖，感謝妳們陪我渡過無力與傷心的時刻，能夠預見你們是我最幸福的事情。再來，感謝我的室友栩嘉、珮琪和亦霆，這兩年的朝夕相處，妳們就像我的家人一樣，彼此分享生活經驗，撰寫論文的過程更是互相加油打氣，認識妳們讓我覺得好幸運也好開心。還要感謝Action、冠達、清峰、義家…等等，與你們相處的點點滴滴都將是我最美好而珍惜的回憶。今後必須各分東西了，但我相信堅定的友誼不會因時空而有所改變，不管距離多遠時間多長，我永遠不會忘記你們的。

我還要感謝一路走來陪在我身旁的好友，欽文、文波、育廷、素敏、雅婷、晃晃、貞君、怡珍和學姊，感謝妳們總是陪在我身旁，陪我渡過無數個開心與挫折的日子，真的好感謝老天讓我認識了你們，因為你們讓我的生活更健康也更充實。最後，我要感謝最愛的父母，提供我無憂無慮的學習機會，也謝謝我的家人，哥哥、大嫂、倍源、阿專和恩恩，給我永遠的支持與勇氣，身為家人的我，感到無比的榮耀與驕傲。

在此，僅以此篇論文獻給我的家人與好友，謝謝你們。

張宛茹　　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

List of Tables

List of Figures

Estimation of Dynamic Tobit Models by Newton's Method and SML-GHK Simulator

Student: Wan-Ru Chang Advisor: Dr. Yow-Jen Jou

Institute of Statistics

National Chiao Tung University

Hsinchu, Taiwan

Abstract

Traffic congestion is a serious problem in everybody's daily life. Prediction of traffic flow is becoming the kernel of the traffic management for oversaturated conditions. Our goal is to forecast the traffic flow, especially predict the cumulative traffic flow. This thesis considers the estimation of the parameters of censored data models with lagged dependent variable. Two methods have been proposed. NR algorithm is supplied for solving the Poisson regression dynamic Tobit model. The SML-GHK simulator is another feasible method for the estimation of dynamic Tobit model whose lagged dependent variable is latent.

Key words: Censoring; Dynamic Tobit model; Poisson regression; NR algorithm; GHK simulator; Simulated likelihood estimator

1. Introduction

Traffic congestion is becoming a serious problem in everybody's daily life. A roadway system is operating in saturated and oversaturated conditions. However, most of the existing control algorithms are developed for undersaturated conditions and would lead to misleading results if they are used to control oversaturated conditions. The appropriation traffic modeling and management procedures for oversaturated conditions are required imminently and become even more pressing with the deployment of more efficient traffic control systems within the intelligent transportation systems (ITSs) complex.

In oversaturated conditions, the subjects of signal control, queue management, and traffic management are closely related – and in fact they become one – more so than in undersaturated conditions. Generally, in congested and oversaturated conditions, maximizing system output is a critical issue. The signal timing algorithm needs to be dynamic and must be capable of responding to and evaluating various queue management strategies. The traffic flow prediction is our goal, the most important objective is the cumulative traffic flow prediction, however, due to the subjects of signal control.

Vehicle detector is used for detecting the presentation of the queue. It is installed on links which traffic flow has key impact on critical intersection. Because of the restriction of detecting, it may be well imagined that the presentation of the queue has unobserved data. The real traffic flow will be unobservable when the quantity is too heavy and the ceiling of the observed flow will be determined according to the location of detector. To adequately model the dynamic and censored feature of our variables of interests, the Poisson dynamic Tobit model is considered for the traffic flow and the Normal distribution is used as an approximation.

The Tobit model is frequently used by economists to analyze limited dependent variable (LDV) models. In order to make the inference of the dynamic Tobit model, including estimating parameters and predicting latent vectors, the conventional technique first used in the estimation of Poisson regression dynamic Tobit model is Newton-Raphson (NR) optimization algorithm. In general, a direct solution to the Tobit model is not possible, although NR iteration method is supplied for solution by computing the gradient and Hessian matrix of the relevant log-likelihood. Secondly, the technique used in the estimation of dynamic Tobit model with lagged latent dependent variable is the simulated maximum likelihood (SML) method through procedures based on a recursive algorithm formulated by Geweke-Hajivassiliou-Keane (GHK) simulator. The SML is feasible for the estimation of such model, and GHK is a smooth, asymptotically unbiased, and consistent simulator for this likelihood function.

The rest of the paper is organized as follows. Section 2 is a review about the researches relative to the Tobit model, NR algorithm, and SML-GHK simulator. Section 3 illustrates the construction of the dynamic Tobit model based on the NR algorithm for Poisson regression and SML-GHK simulator for lagged latent dependent variable. We also construct valid asymptotic inference for the parameters of interests. The numerical example with traffic congested and oversaturated conditions is described in section 4. Comparing the effect if using different methods and choosing distinct forms for the LDV model, a few concluding remarks are made in section 5.

2. Literature Review

A growing proportion of travel is occurring in congested and oversatured conditions around the world. Oversaturation refers to conditions where traffic queues persist from cycle to cycle either due to insufficient green splits or because of blockage. The procedure for dynamic design and evaluation of traffic management strategies in oversaturated conditions was presented by Ghassan and Benekohal (2003).

A very common problem in economic data is the censoring of the dependent variable. When the dependent variable is censored, values in a certain range are all transformed to a single value. The censored regression model was first introduced by Tobin (1958) in which the dependent variable is normal but truncated to the left of zero. An iterative procedure starting from a certain initial estimator was proposed in his paper. The dependent variable of such regression often has a known lower or upper bound; it takes on the boundary values for a number of observations and takes a wide range of values for the remaining observations. This model has been widely applied in economics up to the present and is usually called 'Tobit' model.

Amemiya (1973) extended the work of Tobin, proved the consistency and asymptotic normality of the maximum likelihood estimator in the regression case. He pointed that nonlinear normal equations generally have multiple roots and for guaranteeing the consistency of the root produced iteratively, it is necessary to start from an initial consistent estimator. Amemiya also showed the second-round estimator which taking Newton iteration from this initial estimator is asymptotically the same as the maximum likelihood estimator. Olsen (1978) indicated that the likelihood function for the Tobit model has a single maximum and this property is proved in detail.

An adaptation for linear regression is Poisson regression model. It has been given considerable attention in the statistics and econometrics literatures. The analysis of economic behavior often leads to the study of characteristics taking a small number of positive values. The assumption of normality of the disturbances can't be made, because the endogenous variables take a small number of values with positive probabilities. Advances in the theory of estimation can be found in Gourieroux et al. (1984). Formulae for the gradient and Hessian matrix of the relevant log-likelihood function were given by Terza (1985) and detailed a Newton-Raphson (NR) optimization algorithm for Poisson regression model with a dependent variable subject to censoring. The asymptotic covariance matrix of the estimator was derived as well. The statistical inference on this model was introduced in Brännäs (1992). He illustrated that the efficiency of Poisson maximum likelihood estimator is lost due to the censoring and the consistent property also disappeared.

The impact of simulation methods on the analysis of LDV models is profound. Various simulation methods and procedures for drawing random variables have also been proposed in the statistics and econometrics literatures. The simulated maximum likelihood (SML) method was introduced by Lerman and Manski (1981). SML is also called maximum simulated likelihood (MSL) on ocassion, and both are the same. Simulated pseudo-maximum likelihood (SPML) method was indicated by Gourieroux and Monfort (1993). McFadden (1989) presented the method of simulated moments (MSM), and Hajivassiliou and McFadden (1990) performed the method simulated scores (MSS). For the estimation of limited dependent variable models, the maximum likelihood is the desirable estimation approach (Tobin, 1958).

We therefore concentrate on the SML for dynamic Tobit model in this paper.

The performance of a simulation method depends on simulator for relevant functions involving integrals. Several simulator have been offered for multinormial probabilities in LDV models. Among multivariate normal probability simulators, Hajivassiliou et al. (1996) suggested that the GHK simulator is the most reliable and accurate way to simulate multivariate normal probabilities for classical estimation. The GHK simulator was developed by Geweke (1992), Hajivassilious and McFadden (1990), and Keane (1990). Hendry and Richard (1992) provided simulation techniques for evaluating likelihood functions of dynamic latent variables models based on a sequential factorization of joint densities of observable and latent variables. The simulated probabilities required for SML can be estimated by the GHK simulator so that one can estimate SML-GHK models. The consistency and asymptotic normality of SML estimator can be obtained directly from Hajivassiliou and McFadden (1998).

Recently, this simulator is further applied in discrete choice estimation. For instance, Chang (2002) applied this to study the earnings dynamics and the impact of Title VII of the 1964 Civil Rights Act on the convergence of the black-white earnings gap. Contoyannis et al. (2004) used binary choice models to show what can be done with conventional methods and how the range of models can be expanded by using Markov Chain Monte Carlo (MCMC) algorithm and GHK simulator.

3. Model Specifications and Methodology

In LDV models, the observed data vector y is an indirect observation on a latent vector y^*. The generating process for y can be regard as an 'incomplete data' or 'partial observable' process.

Definition 1. (Censored Random Variable) *Let Y^* be a random variable from a population with c.d.f. $F(Y^*)$ and support $\boldsymbol{A}$. Let $\boldsymbol{B}$ be the support of the random variable $Y = \tau(Y^*)$ where $\tau : \boldsymbol{A} \to \boldsymbol{B}$ is not invertible. Then Y is a censored random variable.*

In such case, τ is often called the 'observation rule' and though it may not be monotonic, τ is generally piece-wise continuous. An important characteristic of censored sampling is that no observations are missing.

Tobin (1958) first considered the regression model where truncated to the left of zero. It is the case of censored random variables. The model is called 'Tobit' model and represented as

$$y_i^* = \beta x_i + \epsilon_i$$

$$y_i = \max\{y_i^*, 0\}$$

where y_i^* is a latent dependent variable, x_i is a vector of exogenous variables, β is the corresponding vector of parameters, the disturbance ϵ_i is assumed to be independent $N(0, \sigma^2)$, and y_i represents an observed dependent variable. In above model, $y_i = y_i^*$ if $y_i^* > 0$ and $y_i = 0$ if $y_i^* \leq 0$ for each individual i.

In censored case, censoring destroys the linearity of the model, and ordinary estimation method for linear regression is not applicable. We therefore discuss two techniques to deal with this problem. The NR algorithm is considered for Poisson regression dynamic Tobit model in the first subsection below, and simulation estimation relied on GHK simulator is proposed for Tobit model with lagged latent dependent variable in the second subsection.

3.1 Poisson Regression with Newton-Raphson Algorithm

The basic model for the analysis of count data is Poisson regression. A property of this model is the variance of data to be equal to the mean, conditional on explanatory variables. The endogenous variable is assumed to have a Poisson distribution conditional upon the explanatory variables. In particular, there is no finite upper limit on the values that may be observed and the parameter of this distribution is a function of the values of the explanatory variables. The lagged observed dependent variables are considered in the Poisson dynamic Tobit model.

NR algorithm is a computational procedure that can be used generally to maximize a complicated function. The idea is to approximate the function by a linear Taylor series. NR algorithm is supplied for solving the Poisson regression models by computing the gradient and Hessian matrix of the relevant log-likelihood function.

Let conditional pmf of y_t given X_t be a Poisson distribution. The mean and the variance of y_t given X_t are equal to m_t. The basic Poisson regression model is

$$f_t(y_t|X_t) = \frac{m_t^{y_t} e^{-m_t}}{y_t!} \qquad t = 1, \ldots, T \tag{1}$$

where $m_t = \exp\{X_t'\theta\}$ is the mean of y_t given X_t, X_t is a $k \times 1$ vector of explanatory variables, and θ is a $k \times 1$ vector of unknown parameters to be estimated. The exponential function appearing in m_t is mainly justified by the positivity of m_t.

The parameter vector θ can be estimated by iterative maximum log-likelihood function via NR optimization algorithm. Given explanatory variables X_t, the dependent variables, y_t, $t = 1, \ldots, T$, are independent, and therefore likelihood function can be written as

$$L = L(\theta) = \prod_{t=1}^{T} f_t(y_t | X_t, \theta)$$

The log-likelihood function

$$l = \log L = \sum_{t=1}^{T} \big(y_t \log m_t - m_t - \log(y_t!) \big)$$

The NR update of θ at the $(i+1)^{th}$ iteration is

$$\theta^{(i+1)} = \theta^{(i)} - H(\theta^{(i)})^{-1} G(\theta^{(i)}) \tag{2}$$

where $G(k \times 1)$ is the gradient and $H(k \times k)$ is the Hessian matrix with

$$\begin{aligned} G(\theta) &= \frac{\partial l}{\partial \theta} = \sum_{t=1}^{T} (y_t - m_t) X_t \\ H(\theta) &= \frac{\partial l^2}{\partial \theta \partial \theta'} = -\sum_{t=1}^{T} m_t X_t X_t' \end{aligned} \tag{3}$$

Detailed account of the censored model is given below. Suppose that the value of y_t is observable for the t^{th} member of the sample iff $y_t < C$ where C is a known non-negative integer. Let d_t be a indicator function defined as

$$d_t = \begin{cases} 1, & \text{if } y_t < C \\ 0, & \text{if } y_t \geq C \end{cases}$$

This is a right censoring, a count above and including C is censored, only value smaller than C is exactly observed. The log-likelihood function l of censored sample can be represented as

$$\sum_{t=1}^{T}\Big(d_t\big(y_t\log m_t - m_t - \log(y_t!)\big) + (1-d_t)\log\big(Pr(y_t \geq C)\big)\Big) \tag{4}$$

Let $Pr(y_t \geq C) = F_t(C)$. The NR iteration defined in Eq.(2) can be used to maximize Eq.(4) after respectively replacing the gradient and Hessian matrix. The calculating process be detailed as follows.

$$G(\theta) = \sum_{t=1}^{T}\Big(d_t(y_t - m_t)X_t + (1-d_t)\frac{\partial}{\partial\theta}\log\big(F_t(C)\big)\Big)$$
$$H(\theta) = \sum_{t=1}^{T}\Big(-d_t m_t X_t X_t' - (1-d_t)\frac{\partial}{\partial\theta'}\frac{1}{F_t(C)}\sum_{i=0}^{C-1} f_t(i)(i-m_t)X_t\Big)$$

In which G and H matrix have following properties

$$\frac{\partial}{\partial\theta}\log\big(F_t(C)\big) = -\frac{1}{F_t(C)}\sum_{i=0}^{C-1} f_t(i)(i-m_t)X_t$$

$$\frac{\partial}{\partial\theta'}\frac{1}{F_t(C)}\sum_{i=0}^{C-1} f_t(i)(i-m_t)X_t$$
$$= \sum_{i=1}^{C-1}\frac{1}{F_t(C)^2}\Big(F_t(C)\frac{\partial}{\partial\theta'}f_t(i)(i-m_t) - f_t(i)(i-m_t)\frac{\partial}{\partial\theta'}F_t(C)\Big)X_t$$
$$= \sum_{i=1}^{C-1}\frac{1}{F_t(C)^2}\Big(F_t(C)\Big((i-m_t)\frac{\partial}{\partial\theta'}f_t(i) + f_t(i)\frac{\partial}{\partial\theta'}(i-m_t)\Big) - f_t(i)(i-m_t)\frac{\partial}{\partial\theta'}F_t(C)\Big)X_t$$
$$= \sum_{i=1}^{C-1}\frac{1}{F_t(C)}f_t(i)\Big((i-m_t)^2 - m_t\Big)X_t X_t' + \sum_{i=1}^{C-1}\frac{1}{F_t(C)^2}\Big(f_t(i) - (i-m_t)\Big)^2 X_t X_t'$$

The gradient and Hessian matrix hence become

$$\begin{aligned} G(\theta) &= \sum_{t=1}^{T}\Big(d_t(y_t - m_t) - (1-d_t)\frac{\Phi_{1t}}{F_t(C)}\Big)X_t \\ H(\theta) &= -\sum_{t=1}^{T}\Big(d_t m_t + (1-d_t)\Big(\frac{\Phi_{2t}}{F_t(C)} + (\frac{\Phi_{1t}}{F_t(C)})^2\Big)\Big)X_t X_t' \end{aligned} \tag{5}$$

where $\Phi_{1t} = \sum_{i=0}^{C-1} f_t(i)(i - m_t)$ and $\Phi_{2t} = \sum_{i=0}^{C-1} f_t(i)\big((i - m_t)^2 - m_t\big)$. The asymptotic covariance matrix of the maximum likelihood estimator of θ is

$$\left(\sum_{t=1}^{T} \left(\big(1 - F_t(C)\big) m_t + \left(\Phi_{2t} + \frac{\Phi_{1t}^2}{F_t(C)} \right) \right) X_t X_t' \right)^{-1} \tag{6}$$

The initial value of θ can be obtained from the uncensored NR algorithm. However the uncensored NR algorithm still need a initial value. One way we can operate is from the population mean of observation y_t. The y_t can be used to estimate m_t due to the equation $\log m_t = X_t'\theta$, i.e. $\log y_t \approx X_t'\theta$.

One importance of adopting this model is to generate prediction of y_t value for each t corresponding to the vector X_t of values on exogenous variables. From the Poisson regression model, we will have

$$\hat{m}_t = \exp(X_t'\hat{\theta}) \tag{7}$$

The complete procedure of NR algorithm is summarized below.

- **Step 1 :** Initiate from $\log y_t \approx X_t'\theta$, and the LSE will be the initial vector.

- **Step 2 :** Put the vector from Step1 into the basic NR algorithm (Eq.(2) and Eq.(3)), update until $|\theta_j^{(i+1)} - \theta_j^{(i)}| < \epsilon$ (a small value) for all j.

- **Step 3 :** Put the vector from Step2 into the censored NR algorithm (Eq.(2) and Eq.(5)), update until $|\theta_j^{(i+1)} - \theta_j^{(i)}| < \epsilon$ (a small value) for all j. The parameter vector $\hat{\theta}^{(i+1)}$ will be the MLE of θ.

- **Step 4 :** Calculate the asymptotic covariance matrix by Eq.(6) and prediction y_t for each individual t by Eq.(7).

In regression model, it is important to spell out the consistent and asymptotically normal estimator. Amemiya (1973) illustrated the convergence of the iteration to the global maximum if the initial estimator is consistent. However, the consistency property of the Poisson maximum likelihood estimator is lost in the LDV models (Brännäs, 1992). The efficiency is also lost based on Poisson regression.

3.2 Gaussian Model with Simulation Estimation

The maximum likelihood approach is the desirable estimation to estimate the dynamic latent variable model. The sequential factorization of joint densities of observable and latent variables were introduced by Hendry and Richard (1992). A recursive algorithm formulated by GHK simulator provides the most reliable and accurate way to simulate multivariate normal probabilities. Once the simulated probabilities are obtained, maximum likelihood techniques are employed. The general procedure is called 'simulated maximum likelihood' (SML).

Definition 2. (SML Estimator) *Let the log-likelihood function for the unknown parameter vector θ given the sample of observations $(y_t, t = 1, \dots T)$ be*

$$l_T(\theta) \equiv \sum_{t=1}^{T} \log f(\theta; y_t)$$

and let $\tilde{f}(\theta; y, \omega)$ be an unbiased simulator so that $f(\theta; y) = E_\omega\big(\tilde{f}(\theta; y, \omega)|y\big)$ where ω is a simulated vector of R random variates, and R is the number of simulation for each t. The simulated maximum likelihood estimator is

$$\hat{\theta}_{SML} \equiv \arg\max_{\theta} \tilde{l}_T(\theta)$$

where

$$\tilde{l}_T(\theta) \equiv \sum_{t=1}^{T} \log \tilde{f}(\theta; y_t, \omega_t)$$

for some given simulation sequence $\{\omega_t\}$.

In this section, we consider the following dynamic model for the true/latent data generating process

$$\begin{aligned} y_t^* &= \lambda y_{t-1}^* + \beta x_t + \alpha + \epsilon_t \qquad t = 1, \ldots, T \\ y_t &= \begin{cases} y_t^*, & \text{if } y_t^* < C \\ C, & \text{if } y_t^* \geq C \end{cases} \end{aligned} \tag{8}$$

where C is a known value. Suppose that the error terms have a normal distribution, i.e. $\epsilon_t \sim N(0, \sigma^2)$. Under this assumption, we derive the simulated likelihood function relied on GHK simulator for a general dynamic Tobit model.

3.2.1 Likelihood Simulation

Given sample observations, let I_t be an indicator function with latent variable censoring, such that

$$I_t = \begin{cases} 1, & \text{if } y_t^* < C \\ 0, & \text{if } y_t^* \geq C \end{cases}$$

Suppose the joint conditional density of observable and latent dependent variable for this model is $f(y_t, y_t^* | y_{t-1}, y_{t-1}^*)$. The classical likelihood function for the Tobit model is represented as

$$L = \int_C^\infty \ldots \int_C^\infty \prod_{t=1}^{T} f(y_t, y_t^* | y_{t-1}, y_{t-1}^*) dy_t^* \tag{9}$$

It is a high-dimensional integral with a dimension for y_t^* as the same as the number of censoring periods [1].

In order to obtain the tractable simulation estimation, the joint density function of $f(y_t, y_t^*|y_{t-1}, y_{t-1}^*)$ in Eq.(9) can be further decomposed into two products of conditional densities, the suggestions were proposed by Hendry and Richard (1992) and Lee (1999)

$$f(y_t, y_t^*|y_{t-1}, y_{t-1}^*) = g_1(y_t|y_t^*, y_{t-1}, y_{t-1}^*) \times h_1(y_t^*|y_{t-1}, y_{t-1}^*) \tag{10}$$

$$= g_2(y_t|y_{t-1}, y_{t-1}^*) \times h_2(y_t^*|I_t = 0, y_{t-1}, y_{t-1}^*) \tag{11}$$

where g_1[2] and g_2[3] is conventional Tobit likelihood function, and latent variables y_t^* can be drawn recursively from importance sampling density h_1 and h_2.

On the basis of the decomposition of Eq.(10) with a finite number of simulation runs R, the unbiased likelihood simulator can be provided

$$\tilde{L} = \frac{1}{R}\sum_{r=1}^{R}\prod_{t=1}^{T}\big(f(y_t|y_{t-1}, y_{t-1}^{*(r)})\big)^{I_t}\big(I_{(C,\infty)}(y_t^{*(r)})\big)^{1-I_t} \tag{12}$$

where $y_t^{*(r)}$ is drawn from h_1. This likelihood simulation is not a smooth function in the parameter space due to the indicator function. In addition, Lee (1999) showed the simulator based on Eq.(12) is inefficient in terms of the simulation variance compared with the GHK-based estimators.

A better approach is to allow sample information to feedback to simulation procedure. That is, the latent variable y_t^* is recursively drawn from the univariate conditional probability $h_2(y_t^*|I_t = 0, y_{t-1}, y_{t-1}^{*(r)})$ conditional on not only the past,

[1] Note that we assume y_0 and y_0^* to be C in this likelihood simulation.

[2] $g_1 = \big(f(y_t|y_{t-1}, y_{t-1}^*)\big)^{I_t}\big(P(I_t = 0|y_{t-1}, y_t^*)\big)^{1-I_t}$ in Eq.(10).

[3] $g_2 = \big(f(y_t|y_{t-1}, y_{t-1}^*)\big)^{I_t}\big(P(I_t = 0|y_{t-1}, y_{t-1}^*)\big)^{1-I_t}$ in Eq.(11).

but the current sample information. Thus, $y_t^{*(r)}$ is drawn from h_2 and the unbiased likelihood simulator behind the GHK simulator can be expressed as

$$\hat{L} = \frac{1}{R}\sum_{r=1}^{R}\prod_{t=1}^{T}\left(f(y_t|y_{t-1}, y_{t-1}^{*(r)})\right)^{I_t}\left(P(I_t = 0|y_{t-1}, y_{t-1}^{*(r)})\right)^{1-I_t} \tag{13}$$

This simulator is expected to be useful for the dynamic Tobit model. SML is applied to estimate the linear dynamic Tobit model through GHK simulator and using Eq.(13) in this paper.

3.2.2 The GHK Simulator

The GHK simulator was developed by Geweke(1992), Hajivassilious and McFadden (1990), and Keane (1990). It is chosen here because a recent exhaustive survey of probability simulators (Hajivassiliou et al., 1996) reaches the conclusion that the GHK simulator is the most reliable and accurate way to simulate multivariate normal probabilities for classical estimation. SML in conjunction with the GHK simulator can cope with the intractabilities of the LDV models.

According to Eq.(8), let θ be the vector of interests in the linear dynamic Tobit model, that is, $\theta = (\lambda, \beta, \alpha, \sigma)$. Suppose the total number of censoring is m and occurs at time $t_1, \ldots, t_m$. Random variables $\xi_t^{(r)}$ are drawn from the uniform random number generator on $[0, 1]$, where $t = t_1, \ldots, t_m$ and $r = 1, \ldots, R$. Those uniform random numbers are kept fixed during an optimization search of SML estimator. For the r^{th} simulation run, the latent variable at censoring period t can be simulated from

$$y_t^{*(r)} = \lambda y_{t-1}^{*(r)} + \beta x_t + \alpha + \sigma\eta_t^{(r)}(\theta) \tag{14}$$

where $y_{t-1}^{*(r)} = y_{t-1}$ if censoring does not occur at time $t-1$.

For generating y_t^* from the univariate conditional probability $h_2(y_t^*|I_t = 0, y_{t-1}, y_{t-1}^{*(r)})$, the truncated normal random variables $\eta_t^{(r)}$ are drawn recursively for censored observations. Let Φ be the cumulative standard normal function and F be the $1-\Phi$. Hence, $\eta_t^{(r)}$ can be calculated by

$$\eta_t^{(r)}(\theta) = F^{-1}\Big(\xi_t^{(r)} F\big(\frac{C - \lambda y_{t-1}^{*(r)} - \beta x_t - \alpha}{\sigma}\big)\Big) \quad \text{if } t \in \{t_1, \ldots, t_m\} \tag{15}$$

This is a mapping that takes a uniform $[0,1]$ random variable into a right truncated standard normal random variate. Then $y_t^{*(r)}$ will be simulated by Eq.(14) for each period of time t.

The simulated likelihood function Eq.(13) can be obtained through Eq.(14) and Eq.(15). Specially in Eq.(13), the first part of two products can be given as

$$f(y_t|y_{t-1}, y_{t-1}^{*(r)}) = \frac{1}{\sigma}\phi\big(\frac{y_t - \lambda y_{t-1}^{*(r)} - \beta x_t - \alpha - \sigma\eta_t^{(r)}}{\sigma}\big) \tag{16}$$

as $t \notin \{t_1, \ldots, t_m\}$, i.e. y_t is a true observation, where ϕ is the standard normal density function. The second part of two products in Eq.(13) can be calculated by

$$P(I_t = 0|y_{t-1}, y_{t-1}^{*(r)}) = 1 - \Phi\big(\frac{C - \lambda y_{t-1}^{*(r)} - \beta x_t - \alpha - \sigma\eta_t^{(r)}}{\sigma}\big) \tag{17}$$

as $t \in \{t_1, \ldots, t_m\}$. The simulated likelihood function Eq.(13) therefore will be obtained by combining Eq.(16) and Eq.(17). Furthermore, the parameter θ will be estimated by maximizing the likelihood simulator. The complete procedure of SML-GHK simulator is summarized below.

- **Step 1 :** Initiate the estimation process with the LSE of the model $y_t = \lambda y_{t-1} + \beta x_t + \alpha + \epsilon_t$, where y_t is the observation at time t.

- **Step 2 :** Generate a random variable $\xi_t^{(r)} \sim U(0,1)$ when observation y_t is censored at period of time t.

- **Step 3 :** Calculate a truncated normal random variable $\eta_t^{(r)}$ from Eq.(15) relied on Step1 and Step2 at each censored time t.

- **Step 4 :** Simulate the latent variable $y_t^{*(r)}$ by Eq.(14) combined with Step3.

- **Step 5 :** Calculate the likelihood simulator Eq.(13) by combining the conditional density Eq.(16) and conditional probability Eq.(17).

- **Step 6 :** The $\hat{\theta}$ will be the SML estimator of θ by maximizing the likelihood simulator Eq.(13) from Step5.

3.2.3 Asymptotic Properties

The consistency and asymptotic normality of SML estimator can be obtained directly from Hajivassiliou and McFadden (1998) since the dynamic Tobit model is a special case of a LDV model. The authors suggest choosing R bigger than $\sqrt{T}$. The following theorem, the proof can be found in their paper, develops the consistency and asymptotic normality of the maximum simulated likelihood estimator for the dynamic Tobit model.

Theorem 1. *Assume that the observations and simulators are independently and identically distributed. Assume also that the moment existence assumption in Hajivassiliou and McFadden (1998) is satisfied. Let* $\hat{\theta} \equiv \arg\max_\theta \hat{l}(\theta)$*, where* $\hat{l}(\theta)$ *is simulated by the GHK simulator with an appropriate choice of the number of sim-*

ulations specified in the proof and is obtained by requiring that the Monte Carlo random numbers are not redrawn when θ *is changed. Then,* $\hat{\theta} \rightarrow \theta^*$ *in probability, and* $(\hat{\theta} - \theta^*) \rightarrow N(0, J)$ *in distribution, where* θ^* *is the true value of* θ *in the parameter space* Θ *and* $J = -E\left(\frac{\partial^2 l}{\partial\theta\partial\theta'}\right)|_{\theta=\theta^*}$*.*

4. Empirical Illustration

4.1 Data Collection and Model Description

The data we analyze are collected near the intersection of Zhong-Hua Road (中華路) and Jing-Guang Road (警光路) of Hsinchu City by the detector, RTMS[4] ,which is in forward-looking configuration. The collected time is from 8:00 AM to 9:00 AM on April 25, 2005. These vehicles drive in the direction of the Hsinchu Train Station. There are 120 data sets and the time interval is 30 seconds. The RTMS records the volume, number of vehicles that pass over the detector at the interval of 30 seconds, and the occupancy, percentage of having vehicles detected at the interval of 30 seconds.

Time series plot of volume is shown in Figure 1, the maximum of volume is 15 and the minimum is 1. Because of the restriction of detecting, we infer that the true volume may equal or exceed the maximum value of observations. It also

[4]The RTMS (Remote Traffic Microwave Sensor) radar is a general-purpose, all-weather traffic sensor which detects presence and measures traffic parameters in multiple independent lanes. It requires no maintenance and was proven to be the only multi-zone traffic detector unaffected by any type of weather. The RTMS is a true presence traffic detector providing presence, volume, occupancy, speed and classification information in up to 8 discrete user-defined detection zones up to 60 m (200 ft.) away. It also uniquely designed for side-fired operation. In side-fired configuration, it is usually mounted on existing side-of-the-road poles, is easy to install and remove, and is fully programmable to support a variety of applications. In its forward-looking configuration it is mounted on overhead sign structures to monitor a specific lane. In many tests performed by traffic professionals worldwide, this presence radar technology has been recognized as the best for almost all traffic management applications.

indicates that the volume is censored at both time $t = 35$ (8:17:30 AM) and $t = 97$ (8:48:30 AM). Figure 2 shows the time series plot of occupancy, some close to zero percent values are found. Two reasons for the close to zero occupancy, one is indeed no vehicle crossing this section of roadway, another is in heavy traffic condition and RTMS can't detect the moving vehicles. In Figure 3 and Figure 4 we have plotted the sample ACF and sample PACF of the volume together with 95% confidence bounds. Notice those values lies outside the confidence bounds at lag 2. This statement corresponds with the signal timing algorithm, and it is needed to be dynamic.

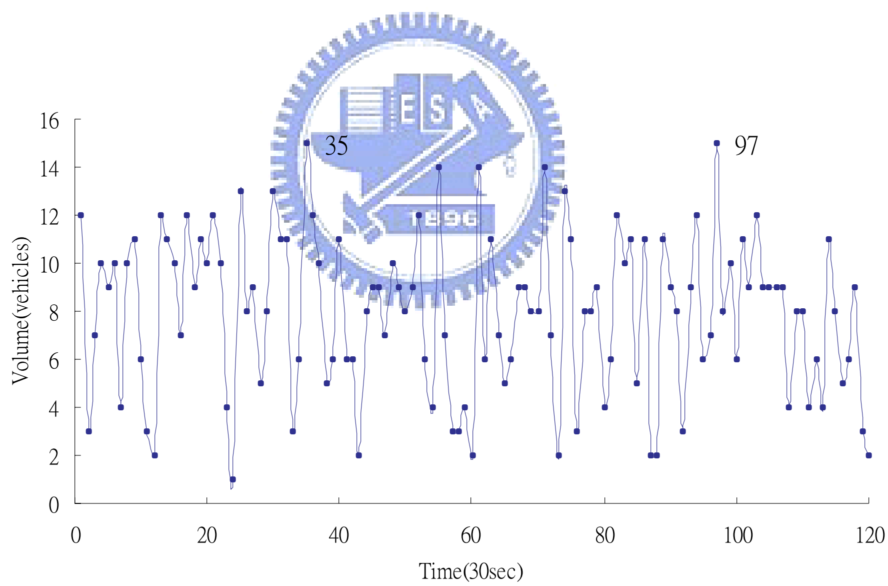

Figure 1: Time series plot of volume.

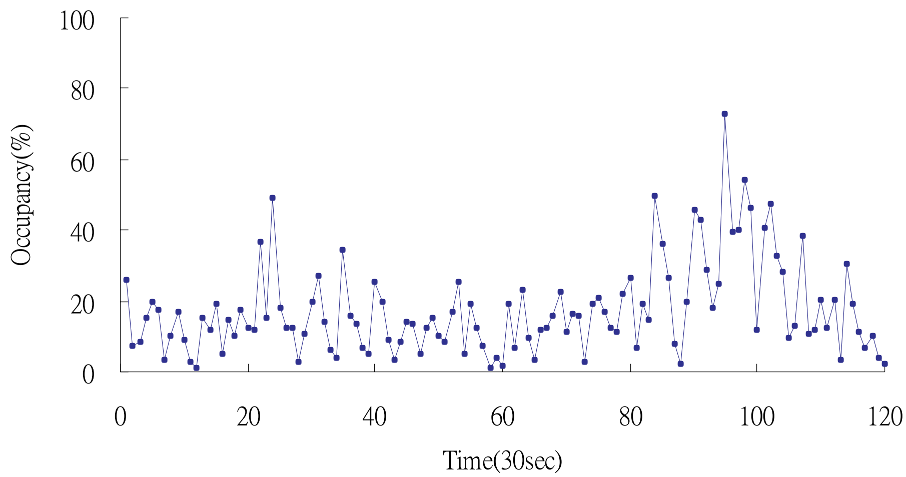

Figure 2: Time series plot of occupancy.

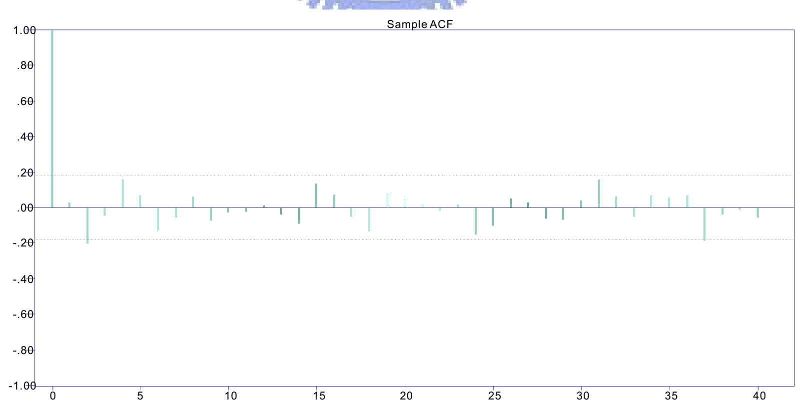

Figure 3: The sample ACF of volume (with 5% significance limits).

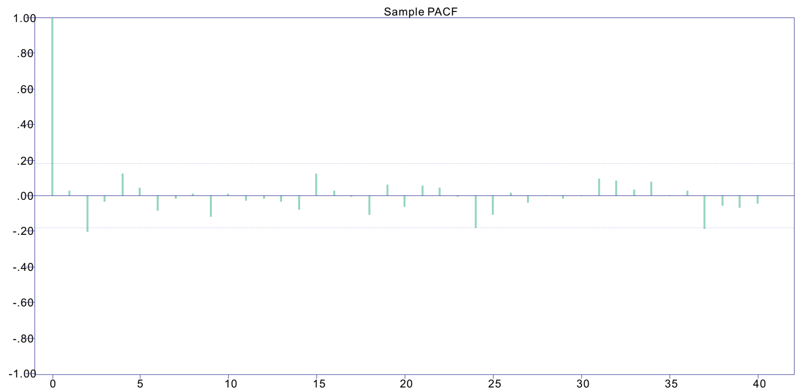

Figure 4: The sample PACF of volume (with 5% significance limits).

The volume and occupancy can be obtained at time point t by RTMS, the dynamic Tobit model is suitable for our traffic data sets

$$y_t^* = \lambda y_{t-1}^* + \beta x_t + \alpha + \epsilon_t \qquad t = 1, \ldots, 120$$

$$y_t = \begin{cases} y_t^*, & \text{if } y_t^* < 15 \\ 15, & \text{if } y_t^* \geq 15 \end{cases}$$

The notation used in the model are listed as follows:

y_t : volume, an observed dependent variable.

y_t^* : volume, the latent dependent variable.

x_t : occupancy, a vector of exogenous variables.

ϵ_t : the disturbance sequence of $i.i.d.$ random variables.

T : the number of observations detected by RTMS.

C : the maximum volume detected by RTMS.

The adopted models in this section have censored data with lagged dependent variables are linked with something previously mentioned. We propose two methods to study such intractable models. First, suppose the dependent variables are from a discrete Poisson distribution – this is reasonable due to the nature of volume. The NR algorithm is used to estimate the models under consideration. Secondly, for parameters of interests in the dynamic Tobit models with lagged latent dependent variable will be obtained by maximizing the likelihood functions simulated through procedures relied on the GHK simulator. The traffic flow prediction is our goal. The most important objective is cumulative traffic volume prediction, however, for the subjects of traffic signal control.

4.2 NR Algorithm

The density function of dynamic Poisson regression with censoring

$$f_t(y_t) = \frac{m_t^{y_t} e^{-m_t}}{y_t!} \qquad t = 1, \ldots, T$$

$$d_t = \begin{cases} 1, & \text{if } y_t < 15 \\ 0, & \text{if } y_t \geq 15 \end{cases}$$

According to the preliminary analysis of the data set, following are three models to be considered.

- **Model 1 :** $m_t = \exp\{\lambda_1 y_{t-1} + \beta x_t + \alpha\} \quad t = 1, \ldots, 120$
- **Model 2 :** $m_t = \exp\{\lambda_2 y_{t-2} + \beta x_t + \alpha\} \quad t = 2, \ldots, 120$
- **Model 3 :** $m_t = \exp\{\lambda_1 y_{t-1} + \lambda_2 y_{t-2} + \beta x_t + \alpha\} \quad t = 2, \ldots, 120$

In these cases, we stop the NR iteration's update as $|\theta_j^{(i+1)} - \theta_j^{(i)}| < 10^{-6}$ for all j. The estimates obtained from uncensored Poisson model are used as the initial values of the parameters to estimate parameters of censored model. The results are shown in Table 1. It summarizes the estimated parameter values and the corresponding SDs for each individual model. It can be found that the influence of volume y_{t-1} and y_{t-2} on present volume y_t is adverse, the effect of y_{t-1} is even weak enough to be ignored. It also shows the small SDs of each parameter.

The prediction information is presented in Table 2. It tells us that prediction means are nearly equal to observation mean but their SDs are smaller than that in the observed data. The minimum values of prediction are larger than volume of observation. It means that the predictions shake near the mean of observations and the predicted ability of high volume is better than low volume. Also, it can be found that Model 2 has less ARPE than others. Table 3 is the cumulative prediction summary including means, SDs and maximum values. It indicates those values are closed to cumulative observation and Model 1 has lesser ARCPE than two other models. The plots of predicted volume and cumulative predicted volume have been shown in Figure 5 to Figure 10.

The results of residual analysis, listed in Table 4, indicate that each individual model are adequate. The time series plots of residuals for each model have been plotted in Figure 11 to Figure 13.

Table 1: Estimated parameter values and the corresponding standard deviations (SD) with NR algorithm.

Parameter	Model 1	Model 2	Model 3
λ_1	-0.0058 (0.0101)		-0.0065 (0.0101)
λ_2		-0.0304 (0.0099)	-0.0294 (0.0099)
β	0.0109 (0.0024)	0.0111 (0.0024)	0.0112 (0.0024)
α	1.9172 (0.0896)	2.1070 (0.0930)	2.1469 (0.1158)

* The values in the parentheses are SDs corresponding its parameter values.

Table 2: Descriptive statistics of prediction compared with observation by NR algorithm.

		Model 1	Model 2	Model 3
	Observation	Prediction	Prediction	Prediction
Mean	7.9250	7.9454	7.9150	7.9159
SD	3.3460	1.2598	1.4682	1.4810
Max	15	14.2285	13.9707	13.8898
Min	1	6.6345	5.5323	5.5353
ARPE[a]		0.5576	0.5370	0.5387

[a] Average relative prediction error $= \frac{1}{T}\sum_{t=1}^{T}\frac{|y_t-\hat{y}_t|}{y_t}$.

Table 3: Descriptive statistics of cumulative prediction compared with observation by NR algorithm.

		Model 1	Model 2	Model 3
	Cumulate Observation	Cumulate Prediction	Cumulate Prediction	Cumulate Prediction
Mean	491.4083	486.0201	461.0155	460.2781
SD	277.8814	276.6035	274.0158	274.0276
Max	951	953.4527	941.8832	941.9922
ARCPE[a]		0.0622	0.0693	0.0706

[a] Average relative cumulative prediction error $= \frac{1}{T}\sum_{t=1}^{T}\frac{|Y_t-\hat{Y}_t|}{Y_t}$.

Table 4: Residual analysis for Model 1, 2, and 3 with NR algorithm.

Testing Method	Model 1	Model 2	Model 3
Ljung-Box[a]	24.847 (0.2074)	13.744 (0.8432)	13.843 (0.8383)
McLeod-Li[a]	16.330 (0.6960)	22.667 (0.3054)	20.270 (0.4411)
Jarque-Bera(normality)[b]	2.7725 (0.2500)	2.4705 (0.2908)	2.9830 (0.2250)

[*] The values in the parentheses are p-value corresponding its χ^2 statistic.
[a] The test statistic has asymptotically a χ^2 distribution with 20 degrees of freedom.
[b] The test statistic has asymptotically a χ^2 distribution with 2 degrees of freedom.

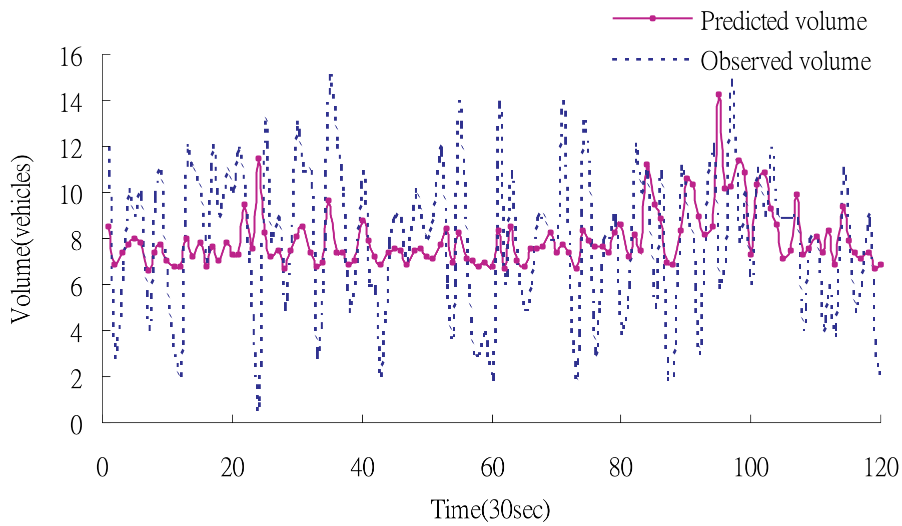

Figure 5: Time series plot of prediction compared with observation for Model 1 (ARPE = 0.5576).

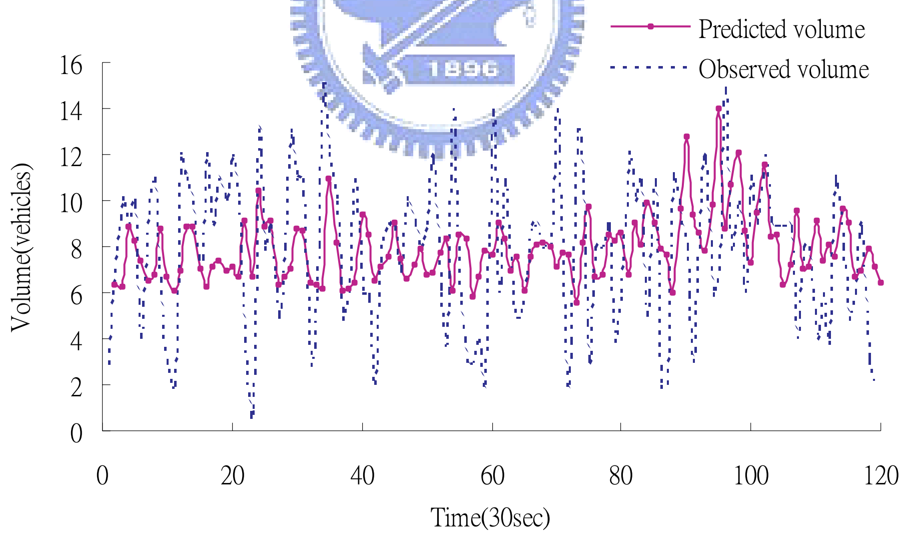

Figure 6: Time series plot of prediction compared with observation for Model 2 (ARPE = 0.5370).

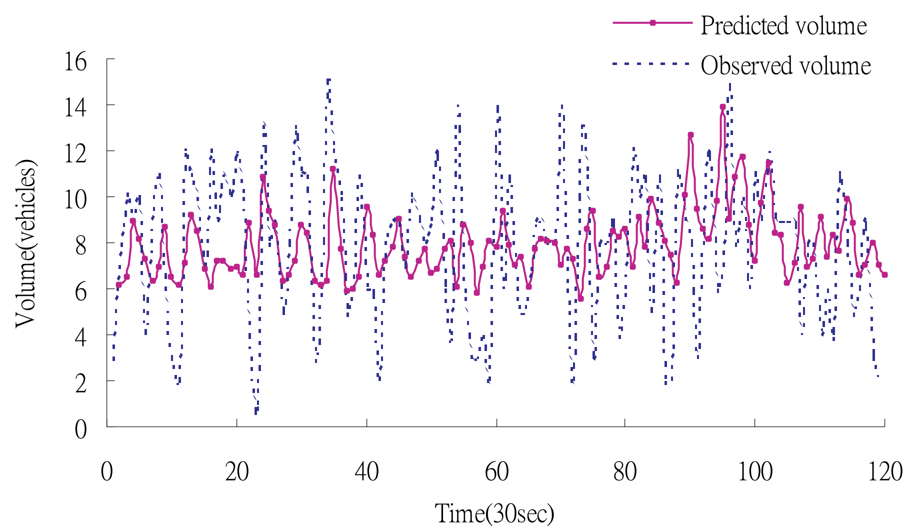

Figure 7: Time series plot of prediction compared with observation for Model 3 (ARPE = 0.5387).

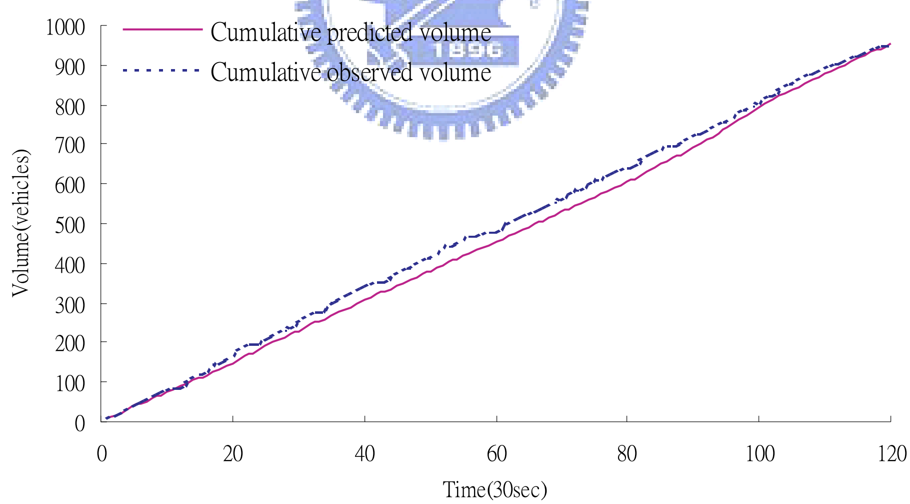

Figure 8: Time series plot of cumulative prediction compared with cumulative observation for Model 1 (ARCPE = 0.0622).

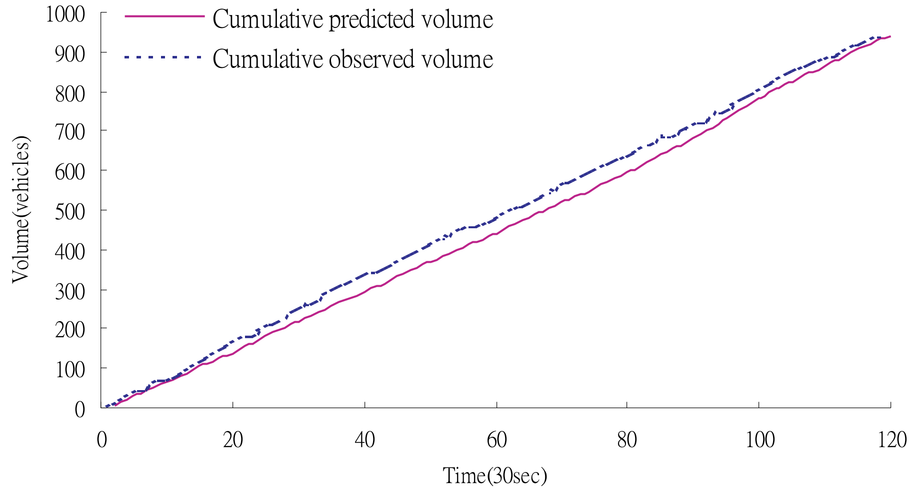

Figure 9: Time series plot of cumulative prediction compared with cumulative observation for Model 2 (ARCPE =0.0693).

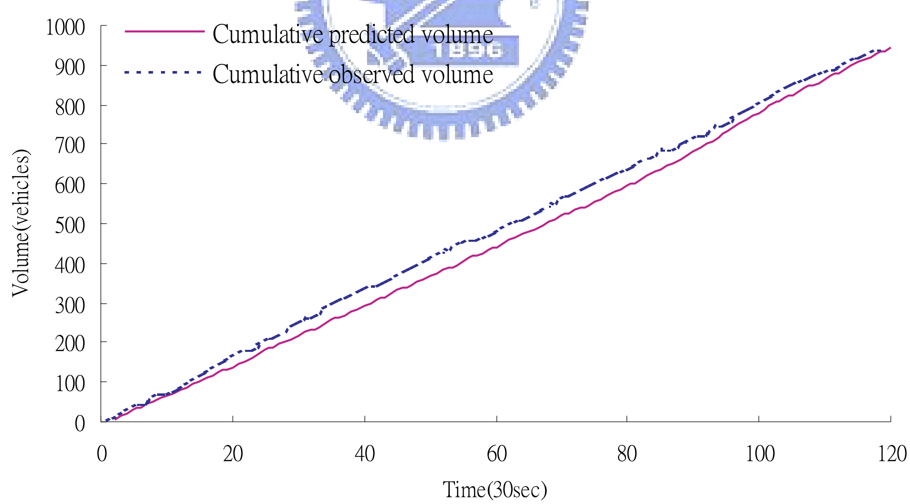

Figure 10: Time series plot of cumulative prediction compared with cumulative observation for Model 3 (ARCPE =0.0706).

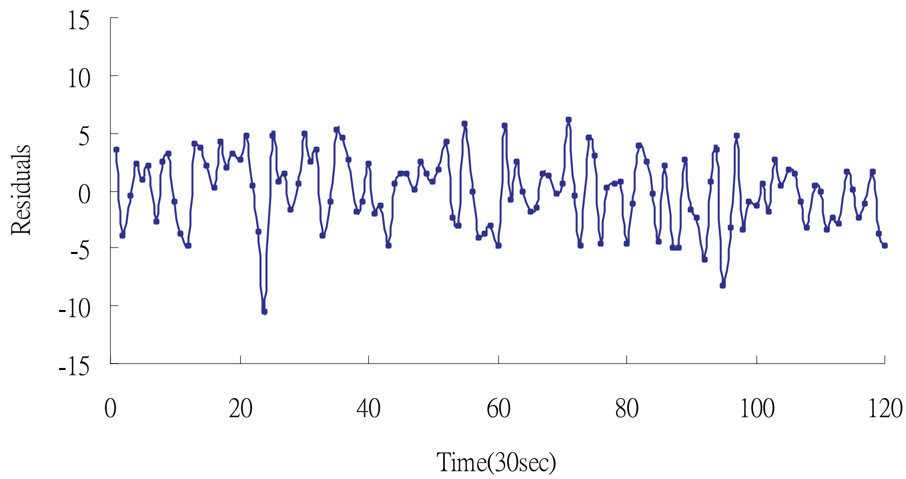

Figure 11: Time series plot of residuals for Model 1.

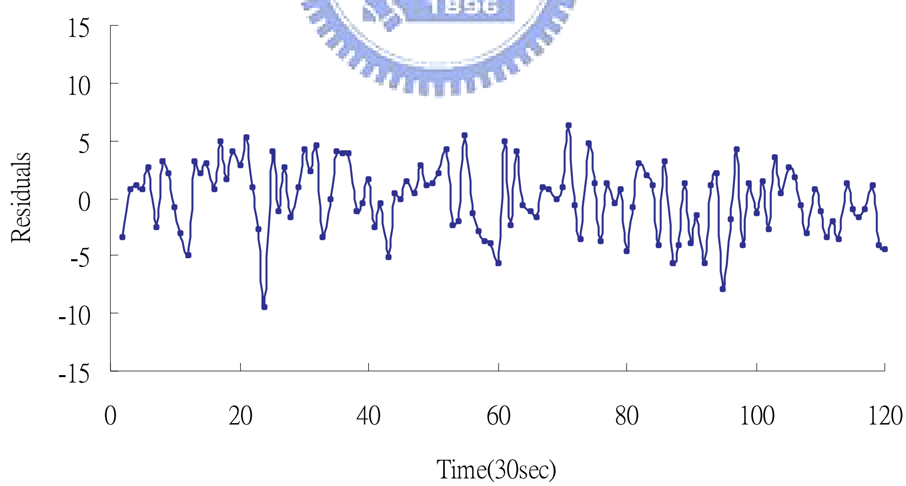

Figure 12: Time series plot of residuals for Model 2.

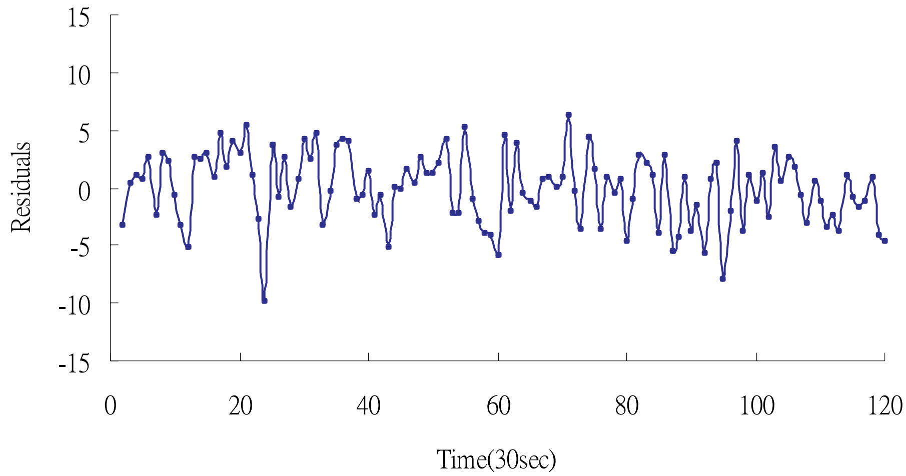

Figure 13: Time series plot of residuals for Model 3.

4.3 SML-GHK Simulator

In this section, results of the analysis of dynamic Tobit models with lagged latent dependent variables by using the SML-GHK simulator for our traffic data will be presented. Firstly, the latent volume values are simulated by generating uniform random numbers and applied to GHK simulator. The parameters of interests will then be estimated by maximizing the unbiased likelihood estimator numerically

$$\hat{L} = \frac{1}{R}\sum_{r=1}^{R}\prod_{t=1}^{T}\left(f(y_t|y_{t-1}, y_{t-1}^{*(r)})\right)^{I_t}\left(P(I_t = 0|y_{t-1}, y_{t-1}^{*(r)})\right)^{1-I_t}$$

Two models are proposed as follows.

- **Model 4 :** $y_t^* = \lambda_1 y_{t-1}^* + \beta x_t + \alpha + \epsilon_t \quad t = 1, \ldots, 120$

- **Model 5 :** $y_t^* = \lambda_1 y_{t-1}^* + \lambda_2 y_{t-2}^* + \beta x_t + \alpha + \epsilon_t \quad t = 2, \ldots, 120$

Let the number of simulations $R = 15$ in these cases. The latent volume, y_{35}^* and y_{97}^*, for two models above can then be estimated and the results have reported in Table 5, the SDs of simulated volume are calculated as well. The estimated parameters and their SDs are summarized in Table 6. As it indicates, the traffic volume y_{t-1} and y_{t-2} affect the present volume in the adverse direction.The prediction and cumulative prediction summaries are listed in Table 7 and Table 8 representatively. ARPE and ARCPE both are smaller for Model 5. The time series plots of prediction and cumulative prediction are shown in Figure 14 to Figure 17. The results of residual analysis, summarized in Table 9, indicate that both models are adequate. The times series plots of residuals have been plotted in Figure18 and Figure 19.

Table 5: Simulation Estimation with SML-GHK simulator where $R = 15$.

	Model 4		Model 5	
Latent volume	Mean	SD	Mean	SD
y_{35}^*	16.8244	1.6308	17.1131	1.6767
y_{97}^*	17.1570	1.7882	17.0346	1.7135

Table 6: Estimated parameters and the corresponding standard deviations (SD) with SML-GHK simulator.

	Model 4		Model 5	
Parameter	Estimate	SD	Estimate	SD
λ_1	-0.0370	0.0048	-0.0335	0.0047
λ_2			-0.2215	0.0024
β	0.0920	0.0026	0.0850	0.0024
α	6.5605	0.0086	8.5090	0.0074
σ	3.1385	0.0071	3.0365	0.0053

Table 7: Descriptive statistics of prediction compared with observation by SML-GHK simulator.

		Model 4	Model 5
	Observation	Prediction	Prediction
Mean	7.9250	7.8574	7.9183
SD	3.3460	1.1706	1.3030
Max	15	12.8021	12.2905
Min	1	6.4969	5.4005
ARPE[a]		0.5462	0.5409

[a] Average relative prediction error $= \frac{1}{T}\sum_{t=1}^{T}\frac{|y_t-\hat{y}_t|}{y_t}$.

Table 8: Descriptive statistics of cumulative prediction compared with observation by SML-GHK simulator.

		Model 4	Model 5
	Cumulate Observation	Cumulate Prediction	Cumulate Prediction
Mean	491.4083	463.7327	462.2012
SD	277.8814	273.5527	274.0060
Max	951	942.8892	942.2777
ARCPE[a]		0.0696	0.0664

[a] Average relative cumulative prediction error $= \frac{1}{T}\sum_{t=1}^{T}\frac{|Y_t-\hat{Y}_t|}{Y_t}$.

Table 9: Residual analysis for Model 4 and 5 with SML-GHK Simulator.

	Model 4		Model 5	
Testing Method	Statistic	p-value	statistic	p-value
Ljung-Box[a]	24.397	0.2255	13.459	0.8568
McLeod-Li[a]	18.026	0.5857	19.967	0.4600
Jarque-Bera(normality)[b]	0.8467	0.6549	1.0543	0.5903

[a] The test statistic has asymptotically a χ^2 distribution with 20 degrees of freedom.

[b] The test statistic has asymptotically a χ^2 distribution with 2 degrees of freedom.

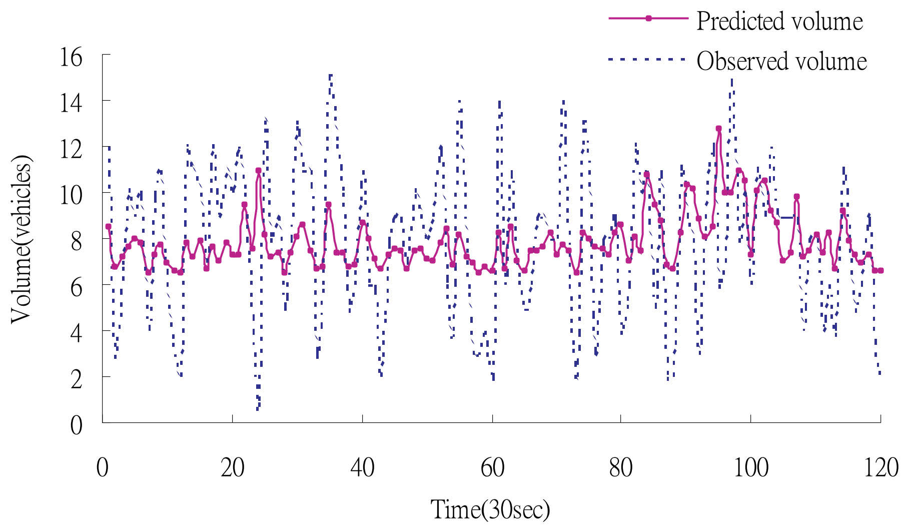

Figure 14: Time series plot of prediction compared with observation for Model 4 (ARPE = 0.5462).

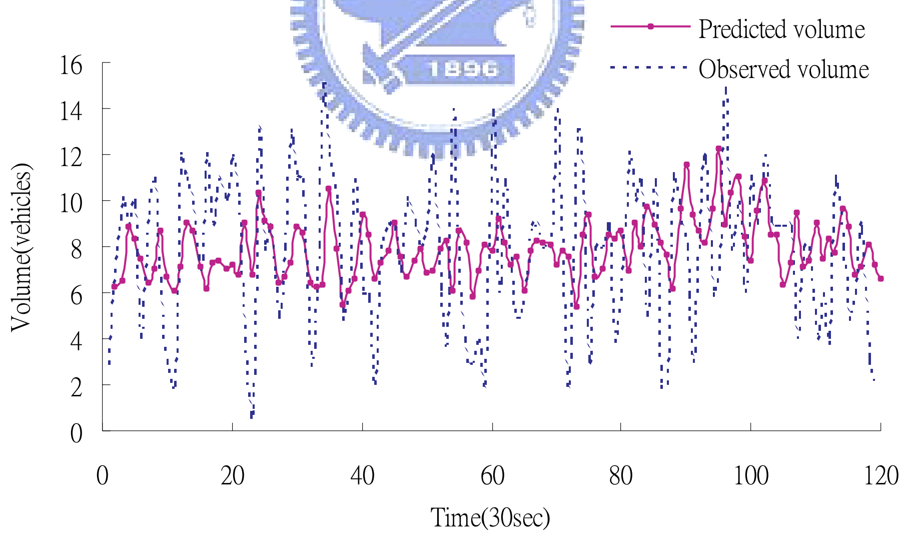

Figure 15: Time series plot of prediction compared with observation for Model 5 (ARPE = 0.5409).

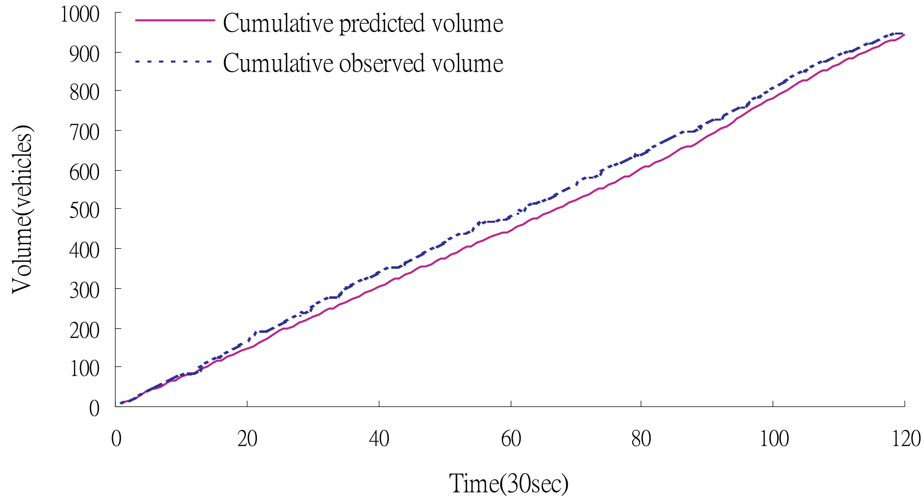

Figure 16: Time series plot of cumulative prediction compared with cumulative observation for Model 4 (ARCPE = 0.0696).

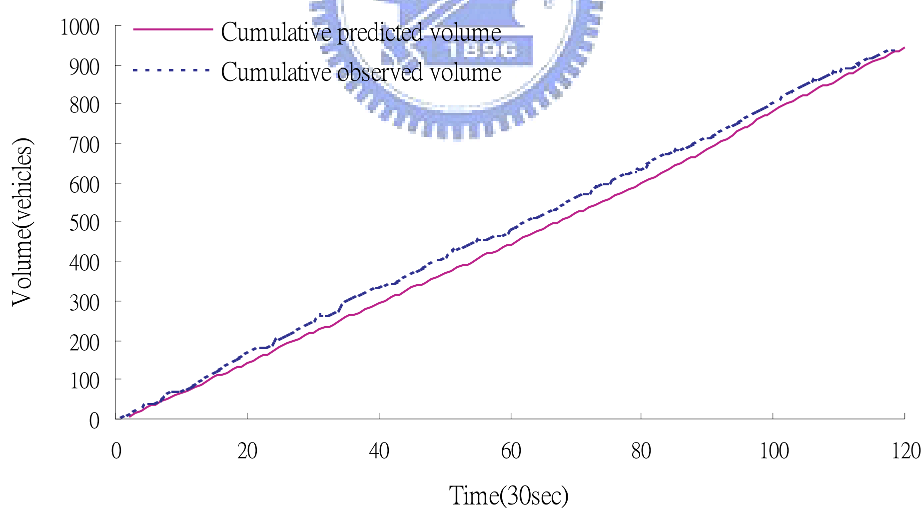

Figure 17: Time series plot of cumulative prediction compared with cumulative observation for Model 5 (ARCPE = 0.0664).

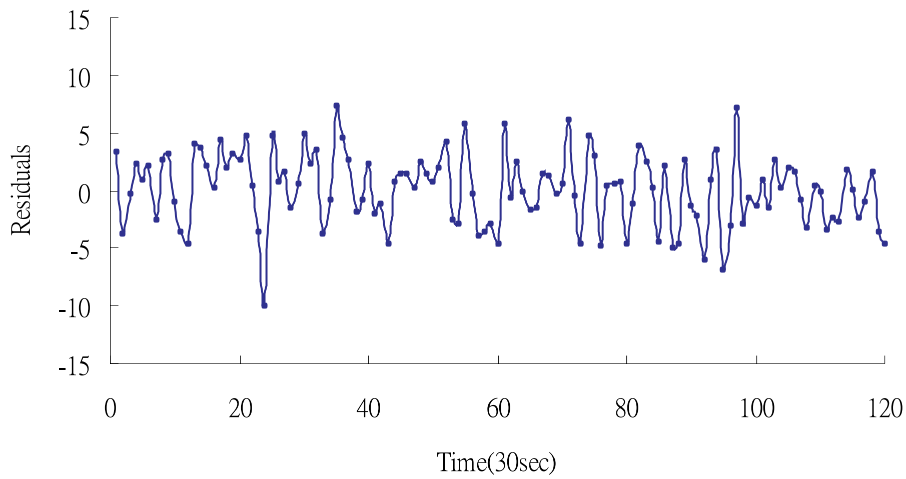

Figure 18: Time series plot of residuals for Model 4.

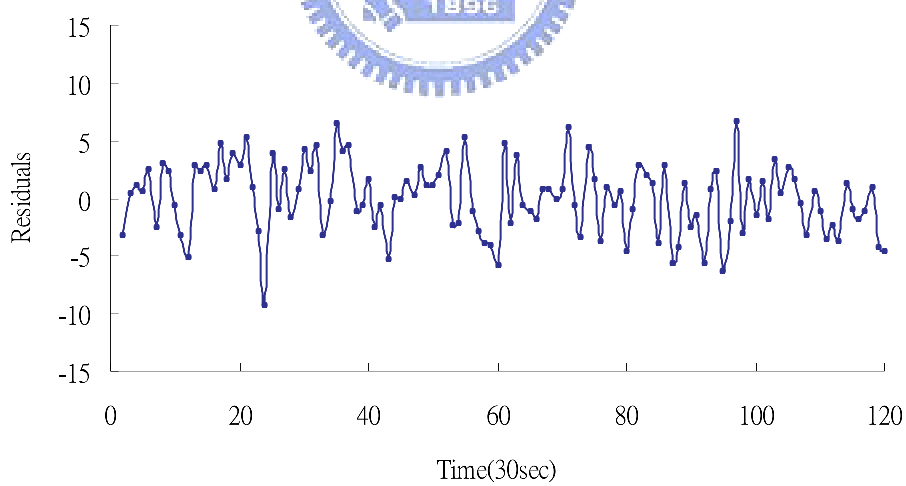

Figure 19: Time series plot of residuals for Model 5.

5. Conclusion and Discussion

This study proposed two methodologies to deal with the traffic censored data and our goal is to find out the cumulative traffic flow predictions. The Newton-Raphson optimization algorithm and SML-GHK simulator have been adopted to overcome our problem under different assumptions. We introduce the Poisson regression dynamic Tobit models and solve the MLE of parameters by using NR algorithm. The Tobit models with lagged latent dependent variables are also used to obtain the SML estimator via SML-GHK simulator. It can be found that similar results of predicted volume and cumulative predicted volume are obtained by different techniques under different models. The following Table 10 shows ARPEs and ARCPEs under five models.

Table 10: The ARPE and ARCPE from NR algorithm and SML-GHK simulator.

	ARPE[a]	ARCPE[b]
Model 1	0.5576	0.0622
Model 2	0.5370	0.0693
Model 3	0.5387	0.0706
Model 4	0.5462	0.0696
Model 5	0.5409	0.0664

[a] Average relative prediction error.
[b] Average relative cumulative prediction error.

This is no strong evidence to discern between models, but it is recognized that both NR algorithm and SML-GHK simulator result in satisfactory cumulative predicted traffic flow. The advantage of using Poisson regression is that the discreteness nature of traffic flow has been taken into consideration, while through SML-GHK simulator the potential censored data can be recovered. Hopefully these two choices will be helpful to design the signal timing plan and ease traffic congestion problem more or less.

References

Amemiya, T. (1973), "Regression Analysis When The Dependent Variable is Truncated Normal", *Econometrica* **41**, 997-1016.

Brännäs, K. (1992), "Limited Dependent Poisson Regression", *The Statistican* **41**, 413-423.

Chang, S. (2002), "Simulation Estimation of Dynamic Panel Tobit Models with an Application to The Convergence of The Black-White Earnings Gap and Income Dynamics", Manuscript, Wayne State University.

Contoyannis, P., Jones, A. and Roberto L. (2004), "Using Simulation-Based Inference with Panel Data in Health Economics", *Health Economics* **13**, 101-122.

Geweke, J. (1992), "Efficient Simulation from the Multivariate Normal and Student-*t* Distributions Subject to Linear Constraints", *Computing Science and Statistics: Proceedings of the Twenty-Third Symposium*, 571-578.

Ghassan, A. and Benekohal, R. (2003), "Design and Evaluation of Dynamic Traffic Management Strategies for Congested Conditions", *Transportation Research Part A* **37**, 109-127.

Gourieroux, C. and Monfort, A. (1993), "Simulation-Based Inference: A Survey with Special Reference to Panel Data Models", *Journal of Econometrics* **59**, 5-33.

Gourieroux, C, Monfort, A. and Trognon, A. (1984), "Pseudo Maximum Likelihood Methods: Applications to Poisson Model", *Econometrica* **52**, 701-720.

Greene, W. (1997), Econometric Analysis, 3^{nd} Edition, New York: Prentice Hall.

Hajivassiliou, V. and McFadden, D. (1990), "The Method of Simulated Scores, with Application to Models of External Debt Crises", Manuscript, Cowles Foundation Discussion Paper No. 967.

Hajivassiliou, V. and McFadden, D. (1998), "The Method of Simulated Scores for the Estimation of LDV Model", *Econometrica* **66**, 863-896.

Hajivassiliou, V., McFadden, D. and Ruud, P. (1996), "Simulation of Multivariate Normal Rectangle Probabilities and Their Derivatives: Theoretical and Computational results", *Journal of Econometrics* **75**, 85-134.

Hendry, D and Richard, J. (1992), Likelihood Evaluation for Dynamic Latent Variables Models, In: Amman, H., Belsley, D. and Pau, L. (eds.), Computational Economics and Econometrics, Amsterdam: Kluwer.

Keane, M. (1990), "A Computationally Efficient Practical Simulation Estimator for Panel Data, with Applications to Estimating Temporal Dependence in Employment and Wages", Manuscript, University of Minnesota.

Lee, L. (1995), "Asymptotic Bias in Simulated Maximum Likelihood Estimation of Discrete Choice Models", *Econometric Theory* **11**, 437-483.

Lee, L. (1999), "Estimation of Dynamic and ARCH Tobit Models", *Journal of Econometrics* **92**, 335-390.

Lerman, S. and Manski, C. (1981), On the Use of Simulated Frequencies to Approximate Choice Probabilities. In: Manski, C. and McFadden, D. (eds.), Structural Analysis of Discrete Data with Econometric Applications. Cambridge: MIT Press.

McCullagh, P. and Nelder, J. (1999), Generalized Linear Models, 2^{nd} Edition, London: Chapman and Hall.

McFadden, D. (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration", *Econometrica* **57**, 995-1026.

Olsen, R. (1978), "Note on the Uniqueness of The Maximum Likelihood Estimator for The Tobit Model", *Econometrica* **46**, 1211-1215.

Peter, J. and Richard, A. (2002), Introduction to Time Series and Forecasting, 2^{nd} Edition, New York: Springer.

Terza, J. (1985), "A Tobit-Type Estimator for The Censored Poisson Regression Model", *Economics Letters* **18**, 361-365.

Tobin, J. (1958), "Estimation of Relationships for Limited Dependent Variable", *Econometrica* **26**, 24-36.