An Analysis of Panel Count Data from Multiple random processes

YouSung Park and HeeYoung Kim *

Abstract

An Integer-valued autoregressive integrated (INARI) model is introduced to eliminate stochastic trend and seasonality from time series of count data. This INARI extends the previous integer-valued ARMA model. We show that it is stationary and ergodic to establish asymptotic normality for conditional least squares estimator. Optimal estimating equations are used to reflect categorical and serial correlations arising from panel count data and variations arising from three random processes for obtaining observation into estimation. Under regularity conditions for martingale sequence, we show asymptotic normality for estimators from the estimating equations. Using cancer mortality data provided by the U.S. National Center for Health Statistics (NCHS), we apply our results to estimate the probability of cells classified by 4 causes of death and 6 age groups and to forecast death count of each cell. We also investigate impact of three random processes on estimation.
Keywords : INARI(p); Non-Stationary Time Series; Three stage; Generalized estimating equations; Asymptotic Normality.

1 Introduction

The NCHS receives monthly mortality reports from Washington D.C. and the 50 states. These monthly deaths are classified by the International Classification of Diseases. These classified deaths are used to estimate classification probability with some interesting covariates and to forecast them.

We present a method to estimate classification probability, to investigate dependency of death count on covariates, and to forecast death count for each classification, accounting both serial and cross-sectional correlations.

We assume that observed death counts are obtained through three random stages. The first stage is a death process denoted by D_t for month t. In the second stage, we choose $\delta_t \times 100\%$ sample from D_t. The third stage is to classify the sample deaths according to their cause of death and demographic factors, giving observed death count denoted by y_{tij} for the ith cause of death and jth demographic factor.

We estimate the classification probability π_{tij} and forecast classified population death rate D_{tij}, simultaneously, using sample y_{tij} and D_t. We apply INARI model for D_t to keep consistency in total, that is, $\hat{D}_t = \sum_{i,j} \hat{D}_{tij}$ where $\hat{D}_t$ and $\hat{D}_{tij}$ are forecasted values of D_t and D_{tij}, respectively. To reflect categorical and serial correlations, we use optimal estimating equations (Godambe,1985). Under appropriate regularity conditions, we investigate asymptotic behavior of estimators obtained by solving the optimal estimating equations.

*YouSung Park is professor, Department of Statistics, Korea University; HeeYoung Kim is Ph.D, Korea University, Korea

2 Integer-Valued Autoregressive Integrated Processes

In this section, we briefly discuss integer-valued autoregressive process with order p (INAR(p)) (Al-Osh and Alzaid 1987, Alzaid and Al-Osh 1990, Jin-Guan and Yuan 1991, McKenzie 1986, McCormick and Park 1997, Park and Oh 1997). Let $w_{\alpha j}$ be an i.i.d. sequence of Bernoulli random variables with $P[w_{\alpha j}=1]=\alpha$. Using this $w_{\alpha j}$, INAR(p) is defined as

$$X_t = \alpha_1 \circ X_{t-1} + \alpha_2 \circ X_{t-2} + \cdots + \alpha_p \circ X_{t-p} + \epsilon_t \tag{1}$$

where $\alpha_i \circ X_{t-i} = \sum_{j=1}^{X_{t-i}} w_{\alpha_i j}$ with $P[w_{\alpha_i j}=1]=\alpha_i$ for $i=1,\dots,p$ and for all j, the $\circ$-operation is referred to binomial thinning operator and $\{\epsilon_t\}$ are i.i.d. non-negative integer-valued random variables with a finite second moment.

This INAR(p) model has the same autocorrelation function as the continuous AR(p) model. But, unlike continuous AR(p), INAR(p) has positive auto correlations because of positive α_i's. All marginal models including (1) for count data assume stationarity except McKenzie (1985) and Brännäs (1995) who include time trend into disturbance term of an INAR(1) model.

To overcome this problem, we introduce a new operator with symbol $\odot$ which we call "signed binomial thinning" operator. Let $\{w_{\alpha,t,j}\}$ be i.i.d. Bernoulli random variables at time t with $P(w_{\alpha,t,j}=1)=|\alpha|$. That is, $\{w_{\alpha,t,j}\}$ are mutually independent when at least one of the subscripts, α, t, and j, is different. We call $\{w_{\alpha,t,j}\}$ counting series (Jin-Guan and Yuan 1991). Define $\check{D}_t = \nabla^d \nabla_s^D D_t$ for nonnegative integers d and D, and $sgn(x)=1$ if $x \geq 0$ and $sgn(x)=-1$ if $x<0$. Using this notation, we define INARI(p) in Section 2.1, using the signed binomial thinning

$$\alpha \odot \check{D}_t \equiv sgn(\alpha) sgn(\check{D}_t) \sum_{j=1}^{|\check{D}_t|} w_{\alpha,t,j}, \tag{2}$$

where the subscript t in $w_{\alpha,t,j}$ indicates the observed time of the process $\check{D}_t$. To simplify notation, we drop the subscripts α and t from $w_{\alpha,t,j}$ hereafter.

2.1 Basic Properties of INARI(p) Process

Using the signed binomial thinning operator $\odot$, we now define integer-valued autoregressive process with order p (INARI(p)) to remove stochastic time trend and seasonality.

$$\check{D}_t = \sum_{i=1}^{p} \alpha_i \odot \check{D}_{t-i} + \epsilon_t, \quad t = 0 \pm 1, \pm 2, \dots \tag{3}$$

where $\{\epsilon_t\}$ is a sequence of i.i.d. integer-valued random variables with mean μ_ϵ and variance σ_ϵ^2, $0 \leq |\alpha_i| \leq 1$ for $i=1,\dots,p$, and all counting series $\{w_{\alpha_i,t,j}\}$ are mutually independent. The $\{\epsilon_t\}$ are uncorrelated with $\check{D}_{t-i}$ for $i \geq 1$.

To show that the INARI(p) process uniquely exists and is stationary, we consider the following process which is similar to that of Jin-Guan and Yuan (1991).

$$\check{D}_{n,t} = \left(\begin{array}{cc} 0 & n<0 \\ \epsilon_t & n=0 \\ \alpha_1 \odot \check{D}_{n-1,t-1} + \cdots + \alpha_p \odot \check{D}_{n-p,t-p} + \epsilon_t & n>0 \end{array} \right. \tag{4}$$

where $Cov(\check{D}_{n,t'}, \epsilon_t) = 0$ when $t' < t$ for any n, and the signs of $\check{D}_{n,t}$ and $\check{D}_{n',t'}$ are the same when $t=t'$ for any n and n'. Adopting the same approach used in Jin-Guan and Yuan (1991) together with basic properties of signed binomial operator, we obtain

Proposition 2.1. *If all roots of the polynomial*

$$\lambda^p - \alpha_1\lambda^{p-1} - \cdots - \alpha_{p-1}\lambda - \alpha_p = 0 \tag{5}$$

are inside the unit circle, then $\check{D}_{n,t}$ converges in L_2 where $L_2 = \{\check{D}_t : E\check{D}_t^2 < \infty\}$.

We construct a INARI(p) process from $\check{D}_{n,t}$ satisfying the condition of Proposition 2.1 as shown below.

Theorem 2.2. *Let $\check{D}_t = lim_{n\to\infty}\check{D}_{n,t}$. Then, the process $\check{D}_t$ uniquely satisfies*

$$\check{D}_t = \sum_{i=1}^{p} \alpha_i \odot \check{D}_{t-i} + \epsilon_t, \quad t = 0 \pm 1, \pm 2, \ldots \tag{6}$$

where $Cov(\check{D}_{t'}, \epsilon_t) = 0$ for $t' < t$. Furthermore, this process $\check{D}_t$ is stationary.

2.2 Estimation

Let $\mathbf{w}(t)$ be all counting series in $\alpha_1 \odot \check{D}_{t-1} + \cdots + \alpha_p \odot \check{D}_{t-p}$. By following Jin-Guan and Yuan (1991), it can be also shown that the process $\check{D}_t$ given in Theorem 2.2 is ergodic because $\{\mathbf{w}(t), \epsilon_t\}$ are independent sequence. This ergodicity and the stationarity of $\check{D}_t$ ensure that, by Wang (1982) and Durrett (1991),

$$\frac{1}{n}\sum_{t=1}^{n} \check{D}_t \xrightarrow{a.s} E(\check{D}_t), \quad \frac{1}{n}\sum_{t=1}^{n} |\check{D}_t| \xrightarrow{a.s} E(|\check{D}_t|),$$
$$\text{and} \quad \frac{1}{n}\sum_{t=1}^{n} \check{D}_t\check{D}_{t-k} \xrightarrow{a.s} E(\check{D}_t\check{D}_{t-k}) \quad \text{for} \quad k = 0, 1, 2, \cdots. \tag{7}$$

Let $\hat{\alpha}_i$ $(i = 1, 2, \ldots, p)$ be the estimator satisfying

$$\hat{\gamma}_k = \hat{\alpha}_1\hat{\gamma}_{k-1} + \cdots + \hat{\alpha}_i\hat{\gamma}_{k-i} + \cdots + \hat{\alpha}_p\hat{\gamma}_{k-p} \tag{8}$$

where $\hat{\gamma}_k = (1/(n-k))\sum_{t=1}^{n-k}(\check{D}_t - \bar{\check{D}})(\check{D}_{t-k} - \bar{\check{D}})$ with $\hat{\gamma}_k = \hat{\gamma}_{-k}$. Using $\hat{\alpha}_i$, define $\hat{\sigma}_\epsilon^2 = (1/n)\sum_{t=1}^{n}(\hat{\epsilon}_t - \bar{\epsilon}_n)^2 + (1/n)\sum_{t=1}^{n}|\check{D}_t| \cdot \sum_{i=1}^{p}|\hat{\alpha}_i|(1 - |\hat{\alpha}_i|)$ and $\bar{\epsilon}_n = (1/n)\sum_{t=1}^{n}\hat{\epsilon}_t$ where $\hat{\epsilon}_t = \check{D}_t - \hat{\alpha}_1\check{D}_{t-1} - \cdots - \hat{\alpha}_p\check{D}_{t-p}$. Then, by (7), we have the following result.

Lemma 2.3. *Let $\sigma_\epsilon^2 = Var(\epsilon_t)$ and $\mu_\epsilon = E(\epsilon_t)$. Then, $\hat{\alpha}_i$, $\hat{\sigma}_\epsilon^2$, and $\bar{\epsilon}_n$ are strong consistent estimators of respective parameters α_i, σ_ϵ^2, and μ_ϵ.*

By the definition of signed binomial thinning, $E(\check{D}_t|\mathcal{F}_{t-1}) = \mu_\epsilon + \alpha_1\check{D}_{t-1} + \cdots + \alpha_p\check{D}_{t-p}$ where $\mathcal{F}_t = \sigma(\check{D}_t, \check{D}_{t-1}, \cdots)$. Denote $\boldsymbol{\xi} = (\mu_\epsilon, \alpha_1, \alpha_2, \cdots, \alpha_p)$. Then, the conditional least squares estimators (CLS) $\widehat{\boldsymbol{\xi}}_n^{LS}$ for $\boldsymbol{\xi}$ can be obtained by minimizing

$$Q_n(\boldsymbol{\xi}) = \sum_{t=p+1}^{n} [\check{D}_t - E(\check{D}_t|\mathcal{F}_{t-1})]^2. \tag{9}$$

Under appropriate regularity conditions which are similar to the conditions to be discussed in Section 3.2 (Klimko and Nelson, 1978), we can show that

$$\sqrt{n}(\widehat{\boldsymbol{\xi}}_n^{LS} - \boldsymbol{\xi}) \xrightarrow{d} N(\mathbf{0}, \mathbf{V^{-1}WV^{-1}}) \tag{10}$$

where

$$\mathbf{V} = \Big[E\Big(\frac{\partial E(\check{D}_{p+1}|\mathcal{F}_p)}{\partial \xi_i}\frac{\partial E(\check{D}_{p+1}|\mathcal{F}_p)}{\partial \xi_j}\Big)\Big]_{(p+1)\times(p+1)} \quad \text{and}$$

$$\mathbf{W} = E\Big[\big(\check{D}_{p+1} - E(\check{D}_{p+1}|\mathcal{F}_p)\big)^2\frac{\partial E(\check{D}_{p+1}|\mathcal{F}_p)}{\partial \xi_i}\frac{\partial E(\check{D}_{p+1}|\mathcal{F}_p)}{\partial \xi_j}\Big]_{(p+1)\times(p+1)}.$$

It is easily seen that $(\hat{\alpha}_i', \bar{\epsilon}_n)$ in Lemma 2.3 and $\widehat{\xi}_n^{LS}$ are asymptotically the same. Thus, $\hat{\alpha}_i$ and $\bar{\epsilon}_n$ are $\sqrt{n}$-consistent for α_i and μ_ϵ by (10). Consequently, it can be shown by (7) and (8) that $\hat{\sigma}_\epsilon^2$ in Lemma 2.3 is a $\sqrt{n}$-consistent estimator of σ_ϵ^2. We use $\hat{\sigma}_\epsilon^2$ in Section 3.

Observe that

$$Var(\check{D}_t|\mathcal{F}_{t-1}) = Var\big(\sum_{i=1}^{p}\alpha_i \odot \check{D}_{t-i} + \epsilon_t|\mathcal{F}_{t-1}\big) = \sum_{i=1}^{p}|\alpha_i|(1-|\alpha_i|)|\check{D}_{t-1}| + \sigma_\epsilon^2.$$

3 Three-Stage Procedures

We assume that the differenced process of death D_t (i.e., $\check{D}_t$) follows INARI(p) described by (2). In the second stage, a δ_t100% sample of deaths is selected from D_t according to a probability sampling where δ_t is given for each t. We assume that the sample is taken by simple random sample without replacement.

The third stage is the post-classification of the sampled records. We assume that the classification by the primary cause i of death and some covariate category j follows a multinomial distribution. For $i = 1, 2, \cdots, l$ and $j = 1, \cdots, g$, let $d_{tm}(i,j) = 1$ if the mth file among D_t belongs to the (i,j)th category and $d_{tm}(i,j) = 0$ otherwise. Then the conditional distribution of $(\sum_{m=1}^{D_t} d_{tm}(1,1), \sum_{m=1}^{D_t} d_{tm}(1,2), \cdots, \sum_{m=1}^{D_t} d_{tm}(l, g-1))$ given D_t is multinomial with parameters $(\pi_{t11}, \pi_{t12}, \cdots, \pi_{t,l,g-1})$ in which $\sum_{i,j}\pi_{tij} = 1$. Because we are interested in the dependency of probability π_{tij} on some useful covariate, $\mathbf{x}_t$, we express $\pi_{tij} = P[d_{tm}(i,j) = 1|\mathbf{x}_t]$. Using this indicator variable $d_{tm}(i,j)$ for classification, the observed count in the (i,j)th category is now

$$y_{tij} = \sum_{m=1}^{d_t} d_{tm}(i,j), \tag{11}$$

where $d_t(=\delta_t \cdot D_t)$ is the sample size at time t.

3.1 Moments

For the relationship among the three random processes, we consider the first two moments of D_t and y_{tij}. Let $\mathcal{F}_t$ be a σ-field generated by $\{D_{t-t'}, t' \geq 0\}$.

Since we assumed that the death process model $\nabla_s^D \nabla^d D_t$ follows the INARI(p) expressed in (3),

$$\begin{aligned} E(D_t|\mathcal{F}_t) &= D_{t-1}^* + \sum_{i=1}^{p}\alpha_i \nabla_s^D \nabla^d D_{t-i} + \mu_\epsilon \quad \text{and} \\ Var(D_t|\mathcal{F}_t) &= \sum_{i=1}^{p}|\alpha_i|(1-|\alpha_i|)|\nabla_s^D \nabla^d D_{t-i}| + \sigma_\epsilon^2 \end{aligned} \tag{12}$$

where $D_{t-1}^* = D_t - \nabla_s^D \nabla^d D_t$, $\mu_\epsilon = E(\epsilon_t)$, and $\sigma_\epsilon^2 = Var(\epsilon_t)$.

Because y_{tij} is observed through the three random processes, using (12) and variations from simple random sampling and multinomial classification, we have

Lemma 3.1. *For y_{tij} expressed in (11) which is obtained through the three random processes,*

$$
\begin{aligned}
&E(y_{tij}|\mathcal{F}_{t-1}) = \delta_t\pi_{tij}E(D_t|\mathcal{F}_{t-1}),\\
&Var(y_{tij}|\mathcal{F}_{t-1}) = \delta_t\pi_{tij}(1-\pi_{tij})E(D_t|\mathcal{F}_{t-1}) + \delta_t^2\pi_{tij}^2 Var(D_t|\mathcal{F}_{t-1}), \quad and\\
&Cov(y_{ti_1j_1}, y_{ti_2j_2}|\mathcal{F}_{t-1}) = \delta_t\pi_{ti_1j_1}\pi_{ti_2j_2}\big(\delta_t Var(D_t|\mathcal{F}_{t-1}) - E(D_t|\mathcal{F}_{t-1})\big).
\end{aligned} \tag{13}
$$

where $i, i_1, i_2 = 1, \cdots, l$, $j, j_1, j_2 = 1, \cdots, g$ and $(i_1, j_1) \neq (i_2, j_2)$.

3.2 Optimal Estimating Equations and Asymptotic Normality

In this section, we establish relation between the proportion $\pi_t = (\pi_{t11}, \pi_{t12}, \cdots, \pi_{tlg})'$ and some useful covariate $\mathbf{x}_t$ by regression model, using some link functions. Then we estimate the proportion by optimal estimating equations and obtain asymptotic normality of estimator.

Using y_{tij} and $\mathcal{F}_{t-1}$, we estimate the parameters $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_p)'$, μ_ϵ, and σ_ϵ^2 for INARI of D_t as well as $\boldsymbol{\beta}_i$ and $\boldsymbol{\eta}_j$ for classification of y_{tij}. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}_1', \cdots, \boldsymbol{\beta}_{l-1}', \boldsymbol{\eta}_1', \cdots, \boldsymbol{\eta}_{g-1}', \boldsymbol{\alpha}', \mu_\epsilon)'$. Denote $\mathcal{Y}_t = (y_{t11}, y_{t12}, \cdots, y_{tlg})'$ and $\mathcal{K}_t(\boldsymbol{\theta}) = \mathcal{Y}_t - E(\mathcal{Y}_t|\mathcal{F}_{t-1})$. Then optimal estimating functions in the sense of Godambe (1985) are defined by

$$
S_n(\boldsymbol{\theta}, \sigma_\epsilon^2) = \sum_{t=1}^{n} \frac{\partial \mathcal{K}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} V_{t-1}^{-1} \mathcal{K}_t(\boldsymbol{\theta}) \tag{14}
$$

where $V_{t-1} = Var(\mathcal{Y}_t|\mathcal{F}_{t-1})$ whose elements are given in (12) and Lemma 3.1. The nuisance parameter σ_ϵ^2 is included in V_{t-1}.

Let

$$
G_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \frac{\partial \mathcal{K}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} V_{t-1}^{-1} \frac{\partial \mathcal{K}_t(\boldsymbol{\theta})'}{\partial \boldsymbol{\theta}}.
$$

We call G_n the conditional information matrix of $S_n(\boldsymbol{\theta}, \sigma_\epsilon^2)$ and $H_n(\boldsymbol{\theta}) = E(G_n)$ the unconditional information matrix. By these two information matrices, we assume the following two regularity conditions. For true $\boldsymbol{\theta}_0$ and $\sigma_{\epsilon,0}$,

(R1) $\frac{\partial \mathcal{K}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} V_{t-1}^{-1} \frac{\partial \mathcal{K}_t(\boldsymbol{\theta}_0)'}{\partial \boldsymbol{\theta}}$ almost surely lies in a nonrandom compact subset Γ and

(R2) there is a positive definite matrix G such that

$$
\frac{1}{n} G_n \xrightarrow{p} G.
$$

Let $\widehat{\boldsymbol{\theta}}_n$ be the solution of

$$
S_n(\boldsymbol{\theta}, \hat{\sigma}_\epsilon^2) = \mathbf{0} \tag{15}
$$

where $\hat{\sigma}_\epsilon^2$ is a $\sqrt{n}$-consistent estimator for $\sigma_{\epsilon,0}^2$. Namely, we consider (14) replaced $\sigma_{\epsilon,0}^2$ by $\hat{\sigma}_\epsilon^2$. Such a consistent estimator of $\sigma_{\epsilon,0}^2$ is provided by Lemma 2.3. Because we use all three stages for $\widehat{\boldsymbol{\theta}}_n$, we call it 3 stage-GEE.

The first-order Taylor expansion of the optimal estimating equations $S_n(\widehat{\boldsymbol{\theta}}_n, \hat{\sigma}_\epsilon^2)$ is

$$
S_n(\widehat{\boldsymbol{\theta}}_n, \hat{\sigma}_\epsilon^2) = 0 = S_n(\boldsymbol{\theta}_0, \hat{\sigma}_\epsilon^2) + \frac{\partial S_n(\boldsymbol{\theta}^*, \hat{\sigma}_\epsilon^2)}{\partial \boldsymbol{\theta}} (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \tag{16}
$$

where $\boldsymbol{\theta}^*$ is an intermediate point between $\widehat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$. In addition to (R1) and (R2), we assume that

(R3) $\frac{1}{n}\frac{\partial S_n(\boldsymbol{\theta}_0, \tilde{\sigma}_\epsilon^2)}{\partial \sigma_\epsilon^2} = o_p(1)$ for any $\sqrt{n}$-consistent estimator $\tilde{\sigma}_\epsilon^2$.

(R4) $\frac{1}{n}\frac{\partial S_n(\boldsymbol{\theta}_0, \hat{\sigma}_\epsilon^2)}{\partial \boldsymbol{\theta}} \xrightarrow{p} W$ where W is a positive definite matrix.

(R5) $\frac{1}{n}(\frac{\partial S_n(\boldsymbol{\theta}^*, \hat{\sigma}_\epsilon^2)}{\partial \boldsymbol{\theta}} - \frac{\partial S_n(\boldsymbol{\theta}_0, \hat{\sigma}_\epsilon^2)}{\partial \boldsymbol{\theta}}) = o_p(1)$.

The asymptotic normality of $\widehat{\boldsymbol{\theta}}_n$ is now obtained by

Theorem 3.2. *Under (R1) through (R5),*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \longrightarrow N(0, W^{-1}GW^{-1}).$$

4 Application

During the 116 months from Jan.1980 to Aug.1989, about 4,477,600 deaths of malignant neoplasms were reported to NCHS. Death records include demographic, medical information, and cause of death information. After careful investigation, we realized that monthly deaths reveals a clear linear time trend as well as a seasonality. Thus, we take $\check{D}_t$ to be $\nabla_{12}\nabla D_t$ to apply INARI(p) model in which D_t is the deaths by malignant neoplasms at month t. Because INARI(p) process for $\check{D}_t$ has the same autocorrelation structure as the usual continuous AR(p) process as seen in Lemma 2.3. The sample autocorrelation function shows ARMA$(1,0)\times$ $(1,0)_{12}$ process.

When we denote $\check{D}_t = \nabla_{12}\nabla D_t$, the followings are the CLS given in Section 2.

$$\begin{aligned} \check{D}_t &= -.213 \odot \check{D}_{t-1} - .232 \odot \check{D}_{t-2} - .360 \odot \check{D}_{t-3} - .145^* \odot \check{D}_{t-4} \\ &\quad -.209 \odot \check{D}_{t-6} - .412 \odot \check{D}_{t-12} - 32.55^* \end{aligned} \tag{17}$$

where $*$ indicates non-significant coefficients, -32.55 is the estimate of μ_ϵ.

NCHS selects a ten percent sample of total deaths each month, giving an aggregate of 437,767 sample deaths during 116 months from Jan.1980 to Aug.1989. This sample deaths were classified into 468 categories: 9 types of malignant neoplasms, 13 age groups, 2 sexes and 2 races.

To investigate the effect of age on the deaths of the neoplasms, we consider the 2 forms of models. The first model for the ith cancer

$$log\frac{\pi_{ti|j}}{\pi_{t4|j}} = \mathbf{x}_{tj}'\boldsymbol{\beta}_i \quad \text{for } i = 1,2,3 \text{ and } j = 1,2,\cdots,6 \tag{18}$$

where $\mathbf{x}_{tj} = (1, A_1, \cdots, A_5, POP_t)'$, $\boldsymbol{\beta}_i = (\beta_{0i}, \beta_{1i}, \cdots, \beta_{6i})'$, and POP_t indicates the U.S. population (which is divided by 10,000,000) at month t.

The second model for the jth age-sex group

$$log\frac{\pi_{tj}}{\pi_{t,6}} = (1, \ t)\boldsymbol{\eta}_j \quad \text{for } j = 1,2,3,4,5 \tag{19}$$

where $\boldsymbol{\eta}_j = (\eta_{0j}, \eta_{1j})'$. We use INARI(12) for $\nabla_{12}\nabla D_t$.

To solve $S_n(\boldsymbol{\theta}, \hat{\sigma}_\epsilon^2) = \mathbf{0}$ defined in (15), we use Fisher scoring iterations defined by

$$\hat{\boldsymbol{\theta}}^{(r+1)} = \hat{\boldsymbol{\theta}}^{(r)} + [\sum_{t=1}^{n} \frac{\partial \mathcal{K}_t(\hat{\boldsymbol{\theta}}^{(r)})}{\partial \boldsymbol{\theta}} V_{t-1}^{-1}(\hat{\boldsymbol{\theta}^{(r)}}, \hat{\sigma}_\epsilon)(\frac{\partial \mathcal{K}_t(\hat{\boldsymbol{\theta}}^{(r)})}{\partial \boldsymbol{\theta}})']^{-1} S_n(\hat{\boldsymbol{\theta}}^{(r)}, \hat{\sigma}_\epsilon). \tag{20}$$

Table 1: Estimates for two logistic models and INARI(12) under three random processes.

i	$\log \frac{\pi_{ti\mid j}}{\pi_{t4\mid j}} = \mathbf{x}'_{tj}\ i$			$\log(\pi_{tj}/\pi_{t,12}) = (1,\ t)(\eta_{0j},\ \eta_{1j})'$				D_t	
	RI($i=1$)	DP($i=2$)	GO($i=3$)	η_{0j}	estimate	η_{1j}	estimate		estimate
β_{0i}	−2.386 (.368)	1.126 (.369)	−1.184 (.426)	η_{01}	−.523 (.018)	η_{11}	−.00094 (.00032)	α_1	−.215 (.098)
A_1	−.601 (.029)	−1.413 (.025)	−1.264 (.031)	η_{02}	.025 (.015)	η_{12}	−.0024 (.0003)	α_2	−.232 (.093)
A_2	.785 (.023)	−.620 (.022)	−.982 (.029)	η_{03}	.960 (.012)	η_{13}	−.0026 (.0002)	α_3	−.360 (.097)
A_3	1.059 (.021)	−.337 (.018)	−.727 (.023)	η_{04}	1.255 (.011)	η_{14}	−.0012 (.0002)	α_4	−.146* (.094)
A_4	1.052 (.020)	−.141 (.017)	−.297 (.021)	η_{05}	.964 (.012)	η_{15}	−.0001* (.0002)	α_5	−.209 (.090)
A_5	.675 (.021)	−.063 (.017)	−.067 (.021)					α_6	−.411 (.079)
POP	.055 (.015)	−.052 (.015)	.016* (.018)					μ_ϵ	−32.10* (61.22)

Table 2: Three statistics for four estimation methods (N/A: not available)

Estimates	D_t			D_{tij}			π_{tij}		
	ME	MAE	MAPE	ME	MAE	MAPE	ME	MAE	MAPE
3-stage GEE	−.476	423.79	.011	.019	66.68	.052	−.0000	.0016	.051
2-stage GEE	−13.42	419.78	.011	.559	53.26	.044	−.0000	.0013	.043
DBE	1817.8	2236.6	.056	75.743	157.10	.123	.0000	.0029	.098
CLS	.0000	423.88	.011	N/A	N/A	N/A	N/A	N/A	N/A

For the nuisance parameter σ_ϵ^2 involved in S_n, we used the conditional least squares estimate $\hat{\sigma}_\epsilon^2 = 336457.5$. Using the convergence criterion

$$\max_i \left| \frac{\hat{\theta}_i^{(r+1)} - \hat{\theta}_i^{(r)}}{\hat{\theta}_i^{(r)}} \right| \le 10^{-3},$$

we estimate the 38 parameters after 5 iterations for the two logistic models in (18) and (19) and for INARI(12) of $\nabla_{12}\nabla D_t$. These estimates are obtained by considering all three random processes and thus they are 3-stage GEE. Table 1 shows these 38 estimates and their standard errors in parentheses. Note that 34 of 38 estimates are significant ("*" means not significant).

NCHS used sample death y_{tij} for analysis of death until 1995. The estimate of D_{tij} was calculated by $10 \times y_{tij}$ after slight adjustment. We call these estimates design based estimate (DBE).

When population D_{tij} is available as NCHS has used it after 1995, sampling stage is no longer necessary. That is, we now have two-stage processes for obtaining classified deaths and hence need to consider 2-stage GEE. 2-stage GEE is obtained by letting $\delta_t = 1$ for all t and replacing y_{tij} by D_{tij} in defining the score function $S_n(\boldsymbol{\theta}, \hat{\sigma}_\epsilon^2) = \mathbf{0}$ given by (15). To compare DBE and 2-stage GEE, with the 3-stage GEE given in Table 1, we use three statistics which are the mean error (ME), the mean absolute error (MAE), and the mean absolute percentage error (MAPE).

Table 2 shows 2-stage GEE and 3-stage GEE have almost same performance to estimate (or forecast) D_t, D_{tij}, and π_{tij} but DBE is much worse than the two estimates.

We assume that D_t is a given number to see the effect of modelling of D_t on 2-stage and 3-stage GEEs. Since no serial correlation exists in this case, the 2-stage and 3-stage GEEs are reduced to the usual quasi-likelihood estimator (QLE). We call them 2-stage QLE and 3-stage QLE. If we assume that D_t is constant when it is actually random, Table 3 implies that overestimation of variance is much more serious when sampling stage is involved than when sampling stage is not involved.

Also, we forecast two months future values of D_t and D_{tij} for $t =$ Sep. 1989 and Oct. 1989 by 3-stage GEE and 2-stage GEE which obtained 91 months from Feb. 1982 to Aug. 1989. These forecasted values are also close to their observation for both estimation methods. In fact, it takes

Table 3: Ratio of variance of GEE to variance of QLE.

	2-stage			3-stage				2-stage		3-stage	
	$i=1$	$i=2$	$i=3$	$i=1$	$i=2$	$i=3$		$i=0$	$i=1$	$i=0$	$i=1$
β_{0i}	1.00	1.00	1.00	1.02	1.01	1.01	η_{i1}	1.00	1.00	1.50	1.01
β_{1i}	1.01	1.01	1.01	3.78	4.58	3.38	η_{i2}	1.00	1.00	1.69	1.01
β_{2i}	1.02	1.02	1.01	6.16	8.27	4.56	η_{i3}	1.00	1.00	2.03	1.00
β_{3i}	1.03	1.04	1.02	7.11	10.50	6.64	η_{i4}	1.00	1.00	2.13	1.00
β_{4i}	1.03	1.05	1.03	7.45	10.48	7.63	η_{i5}	1.06	1.04	2.10	.91
β_{5i}	1.17	1.27	1.17	7.52	10.63	7.94					
β_{6i}	1.00	1.00	1.00	1.00	1.00	1.00					

a few months until mortality reports are classified after NCHS receives the reports. Thus, the forecasted values by 2-stage GEE may fill this time gap.

References

Alzaid, A.A. and Al-Osh, M. (1990). An integer-valued pth-order autoregressive structure (INAR(p)) process. *Journal of Applied Probability.* 27, 314-324.

Brännäs, K. (1995). Prediction and control for time series count data model. *International Journal of Forecasting.* 11, 263-270.

Durrett, R. (1991). *Probability.* Wadsworth, Betmont, CA.

Fokianos, K. and Kedem, B. (1998). Prediction and classification of non-stationary categorical time series. *Journal of Multivariate Analysis.* 67, 277-296.

Godambe, V.P. (1985). The foundations of finite sample estimation in stochastic processes. *Biometrika.* 72, 419-428.

Jin-Guan, D. and Yuan, L. (1991). The integer-valued autoregressive (INAR(p)) model. *Journal of Time Series Analysis.* 12, 129-142.

Klimko, L.A. and Nelson, P.I. (1978). On conditional least squares estimation for stochastic processes. *The Annals of Statistics.* Vol. 6, No. 3, 629-642.

McCormick, W.P. and Park, Y.S. (1997). An Analysis of Poisson moving-average processes. *Probability in the Engineering and Informational Science.* 11, 487-507.

McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Resources Bulletin.* Vol. 21, No. 4, 645-650.

McKenzie, E. (1986). Autoregressive moving-average processes with negative binomial and geometric marginal distributions. *Advances in Applied Probability.* 18, 679-705.

Park, Y.S. and Oh, C.H. (1997). Some asymptotic properties in INAR(1) processes with Poisson marginals. *Statistical Papers.* 38, 287-302.

Biometrics. 75, 4, 621-629.

Wang, Z.K. (1982). *Stochastic Process.* Beijing: Scientific Press.