Focused and Model Average Estimation for Regression Analysis of Panel Count Data

HAIYING WANG

Department of Mathematics and Statistics, University of New Hampshire

YANG LI

Department of Mathematics and Statistics, University of North Carolina at Charlotte

JIANGUO SUN

Department of Statistics, University of Missouri

Abstract

Panel count data arise in many fields and a number of estimation procedures have been developed along with two procedures for variable selection (Sun & Zhao, 2013). In this paper, we discuss model selection and parameter estimation together. For the former, a focused information criterion (FIC) is presented and for the latter, a frequentist model average (FMA) estimation procedure is developed. A main advantage, also the difference from the existing model selection methods, of the FIC is that it emphasizes the accuracy of the estimation of the parameters of interest, rather than all parameters. Further efficiency gain can be achieved by the FMA estimation procedure as unlike existing methods, it takes into account the variability in the stage of model selection. Asymptotic properties of the proposed estimators are established and a simulation study conducted suggests that the proposed methods work well for practical situations. An illustrative example is also provided.

Keywords: estimating function, focused information criterion, frequentist model average, model selection

1 Introduction

Panel count data arise from event history studies concerning recurrent events when study subjects are observed only at discrete time points (Kalbfleisch & Lawless, 1985; Sun & Zhao, 2013). In other words, only incomplete information is available for the sample path of the underlying recurrent event process. More specifically, with panel count data, one knows only the number of the events that occur between observation times, but not the occurrence times. In contrast, the observed data are usually referred to as recurrent event data if each subject is observed continuously or complete information is available (Cook & Lawless, 2007).

A well-known example of panel count data is the bladder cancer data discussed in Sun & Wei (2000) and Wellner & Zhang (2007) among others. The data arose from a study conducted by the Veterans Administration Cooperative Urological Research Group (Wei *et al.*, 1989; Ghosh & Lin, 2002). In the study, the patients with superficial bladder cancer were randomly assigned to one of three treatment groups and followed for the recurrences of the superficial bladder tumors. Furthermore, the patients visited the clinical centers periodically and at each visit, the bladder tumors that occurred since the previous visit were removed and the number was recorded. In other words, only panel count data are available about the tumor recurrence process. Other areas that often produce panel count data include epidemiological studies, reliability studies and social sciences.

Many statistical procedures have been developed for the analysis of panel count data. For example, Sun & Kalbfleisch (1995) and Wellner & Zhang (2000) investigated the nonparametric estimation of the mean function of recurrent event processes based on panel count data. Sun & Wei (2000), Cheng & Wei (2000), Zhang (2002) and Wellner & Zhang (2007) considered regression analysis of the data, while Thall & Lachin (1988), Sun & Kalbfleisch (1993), Sun & Fang (2003), Park *et al.* (2007), Zhang (2006), Balakrishnan & Zhao (2011) and Zhao & Sun (2011) discussed the treatment comparison based on the data. In addition, He *et al.* (2008) and Li *et al.* (2011) studied the analysis of multivariate panel count data and Tong *et al.* (2009), Wu & He (2012) and Zhang *et al.* (2013) considered the variable selection problem. In the following, we discuss the model selection and estimation problems together.

The three variable selection procedures given in the last three references mentioned above share two common features. One is that all were developed based on the overall goodness-of-fit of the data and thus they treat all variables or covariates equally. It is apparent that in practice, sometimes there may exist some variables or covariates of interest and others are simply nuisance variables. Furthermore, the potential best models may be different depending on the variables of interest. The other common feature is that all three methods focus on the underlying recurrent event process only. It is well-known that in the case of panel count data, unlike recurrent event data, there exists another process, the observation process, that characterizes observation time points. This observation process could play an important rule in the analysis and thus in the variable selection (Sun & Zhao, 2013). To address the two issues mentioned above, we will present a different model selection procedure based on the focused information criterion (FIC), which was originally proposed by Claeskens & Hjort (2003) for general parametric models and further investigated by Hjort & Claeskens (2006), Zhang & Liang (2011) and Zhang *et al.* (2012) among others.

The idea behind the FIC is to minimize the asymptotic mean square error of the estimators of pre-specified parameters and it is easy to see that it can choose different models depending on the variables of interest. On the other hand, the selected model is random by any criteria that depends on data. Corresponding to this, we will develop a frequentist model average (FMA) estimation procedure by following Hjort & Claeskens (2003), Zhang & Liang (2011), Zhang *et al.* (2012), Wang & Zou (2012) and Wang *et al.* (2012). Instead of performing estimation based on a single chosen model, the FMA procedure employs the weighted average of the estimates obtained under all possible models. It is apparent that the approach can take into account the variability in model selection and thus yield more efficient estimation.

Compared to the existing results for the FIC and the FMA in the literature, the setup for panel count data presents new challenge since the same covariates may affect both the observation process and the underlying response process. In special cases when the recurrent event process and the observation process are independent, the application of the FIC and the FMA can be simplified and some similar results on other models can be found in Claeskens & Hjort (2008), Wang *et al.* (2012), Zhang & Liang (2011) among others. In general, however, the observation process partly determines the asymptotic distributions of the resulting FMA estimators and should not be ignored from the implementation of the FIC and the FMA procedures. Technical details about this issue will be remarked in Sections 3 and 4.

The rest of the article is organized as follows. We will begin in Section 2 with introducing the notation, models and some assumptions that will be used throughout the paper. A commonly used estimation procedure for regression parameters is also reviewed. Section 3 discusses parameter estimation for a given sub-model and presents some preliminary results needed for the development of the FIC and the FMA procedures in Section 4. In addition, Section 4 gives a method for the construction of confidence intervals for the parameter of interest. Section 5 presents some results obtained from a simulation study conducted for the evaluation of the finite sample properties of the proposed estimators, and an illustrative example is discussed in Section 6. Section 7 contains some concluding remarks.

2 Models, Assumptions and Review

Consider an event history study that involves n subjects who may experience some recurrent events. For subject i, let $N_i(t)$ denote the cumulative number of the events that have occurred up to time t, $0 \le t \le \xi$, where ξ is a known time point representing the study length, $i = 1, ..., n$. Also for subject i, suppose that there exists a d-dimensional vector of covariates denoted by $Z_i = (Z_{i1}, ..., Z_{id})^\top$ and without loss of generality, assume that the expected value of Z_i is 0. To describe the effects of covariates on $N_i(t)$, we assume that given Z_i, the mean function of $N_i(t)$ has the form

$$\mu_i(t) = \mu_0(t) \exp(\beta^\top Z_i) \, , \tag{1}$$

where $\mu_0(\cdot)$ is a completely unspecified function and β denotes the vector of unknown regression parameters. That is, the $N_i(t)$'s follow the proportional mean model (Cook and Lawless, 2007).

In the following, suppose that only panel count data on the $N_i(t)$'s are available. In other words, each counting process $N_i(\cdot)$ is observed only at a sequence of discrete time points denoted by $T_{i1} < T_{i2} < \cdots < T_{iK_i}$. Let C_i denote the follow-up time for the ith subject and define $\tilde{N}_i(t) = H_i\{\min(t, C_i)\}$, where $H_i(t) = \sum_{l=1}^{K_i} I(T_{il} \le t)$, representing the underlying observation process on $N_i(\cdot)$. Note that for the $\tilde{N}_i(t)$'s, one observes recurrent event data and in reality, covariates may have effects on $\tilde{N}_i(t)$ too. For this, we will assume that $H_i(t)$ follows the Cox intensity model given by

$$\tilde{\mu}_i(t) = \tilde{\mu}_0(t) \exp(\gamma^\top Z_i) \tag{2}$$

(Andersen & Gill, 1982), where $\tilde{\mu}_0(\cdot)$ is also an unknown function as $\mu_0(t)$ and γ denotes the covariate effect. Also we will assume that given Z_i, H_i and N_i are stochastically independent of each other, C_i is independent of (H_i, N_i, Z_i), and $\{(H_i, N_i, Z_i, C_i), i = 1, ..., n\}$ are independent and identically distributed for $t \in [0, \xi]$.

As discussed above, if all variables or covariates are equally of interest, several methods have been developed for estimation of models (1) and (2). In particular, for estimation of β and γ, Sun & Wei (2000) proposed to use the following estimating equations

$$W_n(\beta, \gamma) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i e^{-(\beta+\gamma)^\top Z_i} \bar{N}_i = 0 \tag{3}$$

and

$$U_n(\gamma) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{\xi} \left\{ Z_i - \frac{\sum_{i=1}^{n} Y_j(t) Z_j e^{\gamma^\top Z_j}}{\sum_{i=1}^{n} Y_j(t) e^{\gamma^\top Z_j}} \right\} \mathrm{d}\tilde{N}_i = 0\,. \tag{4}$$

In the above, $Y_j(t) = I(C_j \geq t)$ and

$$\bar{N}_i = \sum_{l=1}^{K_i} N_i(T_{il}) I(T_{il} \leq C_i) = \int_0^{\xi} N_i(t) \mathrm{d}\tilde{N}_i(t)\,.$$

In the next section, we will generalize the method above to a more general set-up.

In the following, we will assume that there exist some variables or covariates of interest. Note that this can occur in many situations and one example is that one of the covariates is the treatment indicator and the main goal of the study is to estimate the treatment effect. For the case, it is apparent that one would like the variables or covariates to be included in all candidate models. To formulate the set-up, we will follow the local mis-specification framework discussed in Claeskens & Hjort (2003) and Hjort & Claeskens (2003) and assume that the true model is in a local neighborhood of a certain fixed model. More specifically, it is supposed that the true values of regression parameters β and γ can be represented by

$$\begin{aligned} \beta_{\text{true}} &= (\beta_1^\top, \beta_2^\top)^\top = (\beta_1^\top, \delta^\top/\sqrt{n})^\top, \\ \gamma_{\text{true}} &= (\gamma_1^\top, \gamma_2^\top)^\top = (\gamma_1^\top, \eta^\top/\sqrt{n})^\top, \end{aligned} \tag{5}$$

respectively. Here we assume that β_1 and β_2 are d_{β_1} and d_{β_2} dimensional vectors, γ_1 and γ_2 are d_{γ_1} and d_{γ_2} dimensional vectors, and $\delta/\sqrt{n}$ and $\eta/\sqrt{n}$ represent the degrees of departure from the fixed null model given by $\beta = \beta_0 = (\gamma_1^\top, 0^\top)^\top$ and $\gamma = \gamma_0^\top = (\gamma_1^\top, 0^\top)^\top$, respectively.

3 Some Preliminary Results

Before presenting the FIC and the FMA procedures, we will first generalize the estimation procedure based on the estimating equations (3) and (4) to a more general set-up in this section. Note that for both model selection and estimation, the inclusion of β_1 and γ_1 in the models is mandatory, while that of β_2 and γ_2 is optional. That is, while all the components of β_1 and γ_1 are included in the models, only some components of β_2 and γ_2 are included or none is included.

In the following, we will assume that $d_{\beta_1} > d_{\gamma_1}$, indicating that if a covariate is included for the observation process, then it is also included for modeling the recurrent event process. Hence there are $2^{d_{\beta_2}+d_{\gamma_2}}$ possible sub-models for the consideration. Let S be a subset of $\{1, 2, \ldots, d_{\beta_2}\}$ and R be a subset of $\{1, 2, \ldots, d_{\gamma_2}\}$. We will use sets S and R to refer a sub-model that includes exactly $\beta_{2,s}$ components of β_2 for $s \in S$ and $\gamma_{2,r}$ components of γ_2 for $r \in R$. Denote the regression coefficients under the sub-model $\{S, R\}$ by β_{S} and γ_{R} such that $\beta_{\mathrm{S}} = \Pi_{\mathrm{S}}\, \beta$ and $\gamma_{\mathrm{R}} = \Psi_{\mathrm{R}}\, \gamma$, where Π_{S} and Ψ_{R} denote the selection matrices that project β and γ to their sub-vectors β_{S} and γ_{R}, respectively. Note that the upper-left sub-matrices of Π_{S} and Ψ_{R} are always two identity matrices since β_1 and γ_1 are always included in the models.

For estimation of β_{S} and γ_{R} under the sub-model $\{S, R\}$, denote $Z_{i\mathrm{S}} = \Pi_{\mathrm{S}} Z_i$ and $Z_{j\mathrm{R}} = \Psi_{\mathrm{R}} Z_j$ and define

$$G_{n\mathrm{R}}^{(0)}(\gamma_{\mathrm{R}}, t) = \frac{1}{n}\sum_{i=1}^{n} Y_j(t) e^{\gamma_{\mathrm{R}}^\top Z_{j\mathrm{R}}}, \qquad G_{n\mathrm{R}}^{(1)}(\gamma_{\mathrm{R}}, t) = \frac{1}{n}\sum_{i=1}^{n} Y_j(t) Z_{j\mathrm{R}} e^{\gamma_{\mathrm{R}}^\top Z_{j\mathrm{R}}},$$

$$G_{n\mathrm{R}}^{(2)}(\gamma_{\mathrm{R}}, t) = \frac{1}{n}\sum_{i=1}^{n} Y_j(t) Z_{j\mathrm{R}}^{\otimes 2} e^{\gamma_{\mathrm{R}}^\top Z_{j\mathrm{R}}}, \qquad E_{n\mathrm{R}}(\gamma_{\mathrm{R}}, t) = \frac{G_n^{(1)}(\gamma_{\mathrm{R}}, t)}{G_n^{(0)}(\gamma_{\mathrm{R}}, t)},$$

$$V_{n\mathrm{R}}(\gamma_{\mathrm{R}}, t) = \frac{G_n^{(2)}(\gamma_{\mathrm{R}}, t)}{G_n^{(0)}(\gamma_{\mathrm{R}}, t)} - E(\gamma_{\mathrm{R}}, t)^{\otimes 2},$$

where $M^{\otimes 2} = MM^\top$ for any matrix M. Also let $G_n^{(i)}(\cdot,\cdot)$, $i = 0, 1, 2$, $E_n(\cdot,\cdot)$ and $V_n(\cdot,\cdot)$ denote the quantities defined above but corresponding to the full model, respectively, and $g^{(i)}(\cdot,\cdot)$, $i = 1, 2, 3$, $\mathbf{e}(\cdot,\cdot)$ and $v(\cdot,\cdot)$ denote their corresponding limits when $n \to \infty$, which exist under the regularity conditions given in the appendix. Then motivated by the estimating equations (3) and (4), under the sub-model $\{S, R\}$, one can estimate β_{SR} and γ_{R} by the solutions to the following estimating equations

$$W_{n\mathrm{SR}}(\beta_{\mathrm{SR}}, \gamma_{\mathrm{R}}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{i\mathrm{S}} e^{-(\beta_{\mathrm{S}} + \Pi_{\mathrm{S}} \Psi_{\mathrm{R}}^{\top} \gamma_{\mathrm{R}})^{\top} Z_{i\mathrm{S}}} \bar{N}_i = 0\,, \tag{6}$$

$$U_{n\mathrm{R}}(\gamma_{\mathrm{R}}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{\xi} \{Z_{i\mathrm{R}} - E_{n\mathrm{R}}(\gamma_{\mathrm{R}}, t)\}\, \mathrm{d}\tilde{N}_i = 0\,. \tag{7}$$

Let $\hat{\beta}_{\mathrm{SR}}$ and $\hat{\gamma}_{\mathrm{R}}$ denote the estimators of β_{SR} and γ_{R} defined above, respectively. The following theorem gives the asymptotic properties of them.

Theorem 1. *Assume that the conditions in the Appendix hold. Then under the local misspecification framework* (5)*, we have that*

$$\begin{aligned}
\sqrt{n}(\hat{\beta}_{\mathrm{SR}} - \beta_{\mathrm{S}0}) \xrightarrow{d} & (\Pi_{\mathrm{S}} A \Pi_{\mathrm{S}}^{\top})^{-1} \Pi_{\mathrm{S}} W - \Pi_{\mathrm{S}} \Psi_{\mathrm{R}}^{\top} (\Psi_{\mathrm{R}} B \Psi_{\mathrm{R}}^{\top})^{-1} \Psi_{\mathrm{R}} U \\
& + (\Pi_{\mathrm{S}} A \Pi_{\mathrm{S}}^{\top})^{-1} \Pi_{\mathrm{S}} A \left\{ \begin{pmatrix} 0 \\ \delta \end{pmatrix} + \begin{pmatrix} 0 \\ \eta \end{pmatrix} \right\} - \Pi_{\mathrm{S}} \Psi_{\mathrm{R}}^{\top} (\Psi_{\mathrm{R}} B \Psi_{\mathrm{R}}^{\top})^{-1} \Psi_{\mathrm{R}} B \begin{pmatrix} 0 \\ \eta \end{pmatrix},
\end{aligned} \tag{8}$$

$$\sqrt{n}(\hat{\gamma}_{\mathrm{R}} - \gamma_{\mathrm{R}0}) \xrightarrow{d} (\Psi_{\mathrm{R}} B \Psi_{\mathrm{R}}^{\top})^{-1} \Psi_{\mathrm{R}} U + (\Psi_{\mathrm{R}} B \Psi_{\mathrm{R}}^{\top})^{-1} \Psi_{\mathrm{R}} B \begin{pmatrix} 0 \\ \eta \end{pmatrix}. \tag{9}$$

Here $A = E\big[\bar{N}_i (Z_i)^{\otimes 2} \exp\{-(\beta_0 + \gamma_0)^{\top} Z_i\}\big]$, $\Gamma = E\big[(\bar{N}_i Z_i)^{\otimes 2} \exp\{-2(\beta_0 + \gamma_0)^{\top} Z_i\}\big]$, B *denotes the limit of*

$$B_n(\gamma_0) = -\frac{1}{\sqrt{n}} \frac{\partial U_n(\gamma_0)}{\partial \gamma} = \frac{1}{n} \sum_{i=1}^{n} \int_0^{\xi} V_n(\gamma_0, t) \mathrm{d}\tilde{N}_i\,,$$

$\gamma_{\mathrm{R}0} = \Psi_{\mathrm{R}} \gamma_0$, $\beta_{\mathrm{S}0} = \Pi_{\mathrm{S}} \beta_0$, U *and* W *are normal variables with joint distribution*

$$\begin{pmatrix} W \\ U \end{pmatrix} \sim N \left\{ 0, \begin{pmatrix} \Gamma & \rho \\ \rho^{\top} & B \end{pmatrix} \right\},$$

and $\rho = E\big[N_i Z_i \exp\{-(\beta_0 + \gamma_0)^{\top} Z_i\} \int_0^{\xi} \{Z_i - \mathbf{e}(\gamma_0, t)\}^{\top} \mathrm{d}\tilde{N}_i\big]$.

Remark 1. *If the recurrent event process and the observation process are independent,* (8) *can be simplified to*

$$\sqrt{n}(\hat{\beta}_{\mathrm{SR}} - \beta_{\mathrm{S0}}) \xrightarrow{d} (\Pi_{\mathrm{S}} A \Pi_{\mathrm{S}}^{\top})^{-1} \Pi_{\mathrm{S}} W + (\Pi_{\mathrm{S}} A \Pi_{\mathrm{S}}^{\top})^{-1} \Pi_{\mathrm{S}} A \begin{pmatrix} 0 \\ \delta \end{pmatrix}.$$

This expression is similar to existing results on sub-model estimates (e.g. Zhang & Liang, 2011), and can be further simplified as in Wang et al. *(2012).*

Remark 2. *Note that the variance matrix of U is B. With some matrix algebra manipulation, it can be shown that the result in* (9) *is equivalent to that in Lemma 1 of Hjort & Claeskens (2006). However, our focus in this paper is on the recurrent event process while γ is a parameter associated with the observation process and is not our primary interest.*

It is worth to note that the parameters δ and η represent the distance between a candidate model and the full model and they will play an important role in developing the FIC and the FMA procedures. For their natural estimators $\hat{\delta} = \sqrt{n}\hat{\beta}_2$ and $\hat{\eta} = \sqrt{n}\hat{\gamma}_2$, based on the theorem above, it is easy to show that under the full model, we have that

$$\hat{\delta} \xrightarrow{d} \Delta_{\delta} = (0, \mathbf{I}_{d_{\beta_2}})(A^{-1}W - B^{-1}U) + \delta,$$
$$\hat{\eta} \xrightarrow{d} \Delta_{\eta} = (0, \mathbf{I}_{d_{\gamma_2}})B^{-1}U + \eta,$$

where $\mathbf{I}_d$ denotes the d dimensional identity matrix.

4 FIC and FMA Procedures

Now we are ready to present the FIC and the FMA procedures. For this, suppose that our main goal is to estimate a scalar parameter $\nu = \nu(\beta)$ with true value $\nu_{\mathrm{true}} = \nu(\beta_{\mathrm{true}})$. We will assume that ν depends only on β since in practice the focus is usually on the recurrent events instead of the observation process. Some comments on this will be given below. Denote the estimator of ν under the sub-model $\{S, R\}$ by $\nu_{\mathrm{SR}}(\hat{\beta}_{\mathrm{SR}})$, which can be obtained by inserting the estimator of β. First we will establish the asymptotic property of $\nu_{\mathrm{SR}} = \nu_{\mathrm{SR}}(\hat{\beta}_{\mathrm{SR}})$ in the following theorem.

Theorem 2. *Suppose that the conclusions of Theorem 1 hold and ν is continuously differentiable at β_0. Then as $n \to \infty$, we have that*

$$\sqrt{n}(\hat{\nu}_{\mathrm{SR}} - \nu_{\mathrm{true}}) \xrightarrow{d} \Lambda_{\mathrm{S}} = \nu_\beta^\top (\Omega_{\mathrm{S}}, -\Pi_{\mathrm{S}}^\top \Pi_{\mathrm{S}} \Phi_{\mathrm{R}}) \begin{pmatrix} W \\ U \end{pmatrix} + \nu_\beta^\top (\Omega_{\mathrm{S}} A - \mathbf{I}_d, \Omega_{\mathrm{S}} A - \Pi_{\mathrm{S}}^\top \Pi_{\mathrm{S}} \Phi_{\mathrm{R}} B) \begin{pmatrix} 0 \\ \delta \\ 0 \\ \eta \end{pmatrix},$$

where $\nu_\beta = \partial \nu(\beta_0)/\partial \beta$, $\Omega_{\mathrm{S}} = \Pi_{\mathrm{S}}^\top (\Pi_{\mathrm{S}} A \Pi_{\mathrm{S}}^\top)^{-1} \Pi_{\mathrm{S}}$ and $\Phi_{\mathrm{R}} = \Psi_{\mathrm{R}}^\top (\Psi_{\mathrm{R}} B \Psi_{\mathrm{R}}^\top)^{-1} \Psi_{\mathrm{R}}$.

Remark 3. *If the recurrent event process and the observation process are independent, then the result reduces to*

$$\sqrt{n}(\hat{\nu}_{\mathrm{SR}} - \nu_{\mathrm{true}}) \xrightarrow{d} \Lambda_{\mathrm{S}} = \nu_\beta^\top \Omega_{\mathrm{S}} W + \nu_\beta^\top (\Omega_{\mathrm{S}} A - \mathbf{I}_d) \begin{pmatrix} 0 \\ \delta \end{pmatrix},$$

which is similar to some existing results in the literature (e.g. Wang et al.*, 2012; Zhang & Liang, 2011). The recurrent event process is of our primary interest and we assume that ν is a function of β. For situations when the parameter of interest is a function of the observation process, we refer the reader to Lemma 3 of Hjort & Claeskens (2006) or Theorem 6.2 of Claeskens & Hjort (2008) for corresponding results.*

Note that as mentioned above, the idea behind the FIC is to minimize the asymptotic mean square error (MSE). Thus to define the FIC with respect to estimation of $\nu = \nu(\beta)$, we need to derive the asymptotic MSE of $\hat{\nu}_{\mathrm{SR}}$, which is given by

$$\begin{aligned} \mathrm{E}\Lambda_{\mathrm{S}}^2 = & \nu_\beta^\top (\Omega_{\mathrm{S}}, -\Pi_{\mathrm{S}}^\top \Pi_{\mathrm{S}} \Phi_{\mathrm{R}}) \begin{pmatrix} \Gamma & \rho \\ \rho^\top & B \end{pmatrix} (\Omega_{\mathrm{S}}, -\Pi_{\mathrm{S}}^\top \Pi_{\mathrm{S}} \Phi_{\mathrm{R}})^\top \nu_\beta \\ & + \nu_\beta^\top (\Omega_{\mathrm{S}} A - \mathbf{I}_d, \Omega_{\mathrm{S}} A - \Pi_{\mathrm{S}}^\top \Pi_{\mathrm{S}} \Phi_{\mathrm{R}} B) \begin{pmatrix} 0 \\ \delta \\ 0 \\ \eta \end{pmatrix}^{\otimes 2} (\Omega_{\mathrm{S}} A - \mathbf{I}_d, \Omega_{\mathrm{S}} A - \Pi_{\mathrm{S}}^\top \Pi_{\mathrm{S}} \Phi_{\mathrm{R}} B)^\top \nu_\beta \end{aligned}$$

based on Theorem 2 above. Therefore it is natural to define the FIC as

$$
\begin{aligned}
\mathrm{FIC}_{\mathrm{S}} =& \nu_\beta^\top (\Omega_{\mathrm{S}}, -\Pi_{\mathrm{S}}^\top \Pi_{\mathrm{S}} \Phi_{\mathrm{R}}) \begin{pmatrix} \Gamma & \rho \\ \rho^\top & B \end{pmatrix} (\Omega_{\mathrm{S}}, -\Pi_{\mathrm{S}}^\top \Pi_{\mathrm{S}} \Phi_{\mathrm{R}})^\top \nu_\beta \\
&+ \nu_\beta^\top (\Omega_{\mathrm{S}} A - \mathbf{I}_d, \Omega_{\mathrm{S}} A - \Pi_{\mathrm{S}}^\top \Pi_{\mathrm{S}} \Phi_{\mathrm{R}} B) \begin{pmatrix} 0 \\ \Delta_\delta \\ 0 \\ \Delta_\eta \end{pmatrix}^{\otimes 2} (\Omega_{\mathrm{S}} A - \mathbf{I}_d, \Omega_{\mathrm{S}} A - \Pi_{\mathrm{S}}^\top \Pi_{\mathrm{S}} \Phi_{\mathrm{R}} B)^\top \nu_\beta \\
&- \nu_\beta^\top (\Omega_{\mathrm{S}} A - \mathbf{I}_d, \Omega_{\mathrm{S}} A - \Pi_{\mathrm{S}}^\top \Pi_{\mathrm{S}} \Phi_{\mathrm{R}} B) \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \mathbf{I}_{d_{\beta_2}} & 0 & -\mathbf{I}_{d_{\beta_2}} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{I}_{d_{\gamma_2}} \end{pmatrix} \begin{pmatrix} A^{-1}\Gamma A^{-1} & A^{-1}\rho B^{-1} \\ B^{-1}\rho^\top A^{-1} & B^{-1} \end{pmatrix} \\
&\times \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \mathbf{I}_{d_{\beta_2}} & 0 & -\mathbf{I}_{d_{\beta_2}} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{I}_{d_{\gamma_2}} \end{pmatrix}^\top (\Omega_{\mathrm{S}} A - \mathbf{I}_d, \Omega_{\mathrm{S}} A - \Pi_{\mathrm{S}}^\top \Pi_{\mathrm{S}} \Phi_{\mathrm{R}} B)^\top \nu_\beta .
\end{aligned}
$$

For practical use, unknown quantities A, B, Γ, ρ and ν_β should be replaced by

$$
\begin{aligned}
A_n &= n^{-1} \sum_{i=1}^n \left[\bar{N}_i (Z_i)^{\otimes 2} \exp\{-(\hat{\beta}_n + \hat{\gamma}_n)^\top Z_i\} \right], \\
\Gamma_n &= n^{-1} \sum_{i=1}^n \left[(\bar{N}_i Z_i)^{\otimes 2} \exp\{-2(\hat{\beta}_n + \hat{\gamma}_n)^\top Z_i\} \right], \\
B_n(\hat{\gamma}_n) &= \int_0^\xi V_n(\hat{\gamma}_n, t) \mathrm{d}\tilde{N}, \\
\rho_n &= n^{-1} \sum_{i=1}^n \left[N_i Z_i \exp\{-(\hat{\beta}_n + \hat{\gamma}_n)^\top Z_i\} \int_0^\xi \{Z_i - \mathbf{e}(\hat{\gamma}_n, t)\}^\top \mathrm{d}\tilde{N}_i \right],
\end{aligned}
$$

and $\hat{\nu}_\beta$, respectively. Note that the $\mathrm{FIC}_{\mathrm{S}}$ defined above is an unbiased estimate of $\mathrm{E}\Lambda_{\mathrm{S}}^2$. For model selection, the FIC procedure will select the model that gives the smallest $\mathrm{FIC}_{\mathrm{S}}$ among all sub-models. If the observation process can be ignored, then the FIC reduce to

$$\begin{aligned}\mathrm{FIC_S} =&\nu_\beta^\top(\Omega_\mathrm{S})\Gamma(\Omega_\mathrm{S})^\top\nu_\beta + \nu_\beta^\top(\Omega_\mathrm{S}A - \mathbf{I}_d)\begin{pmatrix}0\\ \Delta_\delta\end{pmatrix}^{\otimes 2}(\Omega_\mathrm{S}A - \mathbf{I}_d)^\top\nu_\beta\\ &- \nu_\beta^\top(\Omega_\mathrm{S}A - \mathbf{I}_d)\begin{pmatrix}0 & 0\\ 0 & \mathbf{I}_{d_{\beta_2}}\end{pmatrix}A^{-1}\Gamma A^{-1}\begin{pmatrix}0 & 0\\ 0 & \mathbf{I}_{d_{\beta_2}}\end{pmatrix}^\top(\Omega_\mathrm{S}A - \mathbf{I}_d)^\top\nu_\beta.\end{aligned}$$

With respect to estimation of covariate effects and their inference, it is apparent that a simple method is to focus or rely on the selected model. However, it is well-known that the selected model is random as the data are random and if a poor model is chosen, the subsequent estimate may have large bias or under estimated variance (Hjort & Claeskens, 2003). Corresponding to this and following Hjort & Claeskens (2003), we propose to employ the model averaging technique to define the FMA estimator of $\nu = \nu(\beta)$ as

$$\hat{\nu} = \sum_\mathrm{S}\sum_\mathrm{R} c_\mathrm{SR}\hat{\nu}_\mathrm{SR}\,. \tag{10}$$

In the above, the $\{\, c_\mathrm{SR} = c(S, R \mid \hat{\delta}, \hat{\eta})\,\}$ are weight functions that satisfy $\sum_\mathrm{S}\sum_\mathrm{R} c_\mathrm{SR} = 1$ and are assumed to depend on the observed data through $\hat{\delta}$ and $\hat{\eta}$ only.

It is easy to see that one advantage of the FMA estimator given above is that it can take into account the variation in the model selection process. For the selection of the weight functions, it is apparent that a simple or naive approach is to set them to be either 0 or 1 depending on if a sub-model is selected. Some general choices are given in the next section. The theorem below gives the asymptotic distribution of $\hat{\nu}$.

Theorem 3. *Suppose that the conclusions of Theorem 1 hold. If ν is differentiable at β_0 and $c(S, R \mid \cdot, \cdot)$ is continuous almost everywhere, then as $n \to \infty$, we have that*

$$\sqrt{n}(\hat{\nu} - \nu_\mathrm{true}) \xrightarrow{d} \Lambda = \nu_\beta^\top(A^{-1}W - B^{-1}U) + \nu_\beta^\top(P - \mathbf{I}_d)\begin{pmatrix}0\\ \Delta_\delta\end{pmatrix} + \nu_\beta^\top(P - Q)\begin{pmatrix}0\\ \Delta_\eta\end{pmatrix},$$

where $P = \sum_\mathrm{S,R} c(S, R \mid \Delta_\delta, \Delta_\eta)\Omega_\mathrm{S}A$ and $Q = \sum_\mathrm{S,R} c(S, R \mid \Delta_\delta, \Delta_\eta)\Pi_\mathrm{S}^\top\Pi_\mathrm{S}\Phi_\mathrm{R}B$.

Remark 4. *If the recurrent event process and the observation process are independent, then* *the result reduce to*

$$\sqrt{n}(\hat{\nu} - \nu_{\text{true}}) \xrightarrow{d} \Lambda = \nu_\beta^\top A^{-1} W + \nu_\beta^\top (P - \mathbf{I}_d) \begin{pmatrix} 0 \\ \Delta_\delta \end{pmatrix} ,$$

which is similar to existing results in the literature on some other models (e.g. Zhang & Liang, 2011).

Note that the limiting distribution given above is not normal and thus a confidence interval cannot be constructed in the standard way. Nevertheless, one can show that the confidence interval for $\hat{\nu}$ can be asymptotically approximated by

$$\hat{\nu} - \frac{\hat{\nu}_\beta^\top (P_n - \mathbf{I}_d)(0^\top, \hat{\Delta}_\delta^\top)^\top + \hat{\nu}_\beta^\top (P_n - Q_n)(0^\top, \hat{\Delta}_\eta^\top)^\top}{\sqrt{n}} \pm \frac{z_\alpha \sqrt{\hat{\nu}_\beta^\top F_n \hat{\nu}_\beta}}{\sqrt{n}} .$$

In the above, z_α denotes a normal quantile at the significance level α,

$$P_n = \sum_{\text{S,R}} c(S, R \mid \hat{\delta}, \hat{\eta}) \hat{\Omega}_{\text{S}} A_n ,$$

$$Q_n = \sum_{\text{S,R}} c(S, R \mid \hat{\delta}, \hat{\eta}) \Pi_{\text{S}}^\top \Pi_{\text{S}} \hat{\Phi}_{\text{R}} B_n$$

and

$$F_n = A_n^{-1} \Gamma_n A_n^{-1} - B_n^{-1} \rho_n^\top A_n^{-1} - A_n^{-1} \rho_n B_n^{-1} + B_n^{-1} .$$

5 A Simulation Study

A simulation study was conducted to assess the performance of the FIC and the FMA procedures proposed in the previous sections. For comparison, we also considered the penalized procedure given in Tong *et al.* (2009) with the use of the Smoothly Clipped Absolute Deviation penalty function along with the Akaike information criterion (AIC) (Akaike, 1973) and the Bayesian information criterion (BIC) (Schwarz, 1978). For the latter two methods, following Hjort & Claeskens (2006) and Tong *et al.* (2009), we define them as

$$\begin{aligned}\mathrm{AIC}_{\mathrm{SR}} &= \mathcal{L}_{\mathrm{SR}} + 2(|S| + |R|),\\ \mathrm{BIC}_{\mathrm{SR}} &= \mathcal{L}_{\mathrm{SR}} + \log(n)(|S| + |R|)\end{aligned} \tag{11}$$

corresponding to the sub-model $\{S, R\}$. In the above, $|S|$ and $|R|$ denote the numbers of components in β_{S} and β_{R}, respectively, and

$$\mathcal{L}_{\mathrm{SR}} = n \log\left[\sum_{i=1}^{n} \exp\left\{-\hat{\beta}_{\mathrm{S}}^{\top} Z_{i\mathrm{S}} - \hat{\gamma}_{\mathrm{R}} Z_{i\mathrm{R}}\right\} \bar{N}_i/n\right] - 2\sum_{i=1}^{n}\int_0^{\xi} \left[\hat{\gamma}_{\mathrm{R}}^{\top} Z_{i\mathrm{R}} - \log\left\{nG_{n\mathrm{R}}^{(0)}(\hat{\gamma}_{\mathrm{R}}, t)\right\}\right] \mathrm{d}\tilde{N}_i\,.$$

As the FIC procedure, the AIC and the BIC procedures will select the models with the smallest $\mathrm{AIC}_{\mathrm{SR}}$ and $\mathrm{BIC}_{\mathrm{SR}}$, respectively. In the Appendix, we prove that AIC and BIC criteria defined in (11) depend on the data through $\hat{\delta}$ and $\hat{\eta}$ asymptotically.

To generate the observed data, we suppose that $N_i(t)$ is a Poisson process with the mean function $\mu_i(t) = 0.5t^2 \exp(\beta_{\mathrm{true}}^{\top} Z_i)$ and $H_i(t)$ follows the model in (2) with $\tilde{\mu}_0(t) = t$. Here we assume that $\beta_{\mathrm{true}} = \{1, -2, 1, c(0, 0.5, 0.5)/\sqrt{n}\}^{\top}$, $\gamma_{\mathrm{true}} = \{0.5, -0.5, 0.5, c(0, -0.1, 0.1)/\sqrt{n}\}^{\top}$ and Z_i follows the multiple normal distribution with mean 0 and the covariance matrix $\{0.5^{I(i \neq j)}\}_{ij}$, where c is a constant. For the follow-up time $C_i = \min(C_i^*, \tau)$, it is assumed that $C_i^* \sim \mathrm{Uniform}(1, 10)$ and $\tau = 8$. Also it is supposed that our interest is on four parameters $\nu_1 = \beta_1$, $\nu_2 = \beta_2$, $\nu_3 = \beta_3$ and $\nu_4 = \sum_{i=1}^{6} \beta_i$. For model selection, as mentioned above, we assume that β_1, β_2, β_3, γ_1, γ_2 and γ_3 are always included and hence there are $2^{3+3} = 64$ possible models to choose from or to average. To carry out the FMA procedure, of course, we need to choose the weight functions c_{SR}'s. For this, we considered the following three choices

$$\frac{\exp(-\frac{1}{2}\mathrm{AIC}_{\mathrm{SR}})}{\sum_{\mathrm{SR}} \exp(-\frac{1}{2}\mathrm{AIC}_{\mathrm{SR}})}, \quad \frac{\exp(-\frac{1}{2}\mathrm{BIC}_{\mathrm{SR}})}{\sum_{\mathrm{SR}} \exp(-\frac{1}{2}\mathrm{BIC}_{\mathrm{SR}})}, \quad \frac{\exp(-\frac{1}{2}\mathrm{FIC}_{\mathrm{SR}})}{\sum_{\mathrm{SR}} \exp(-\frac{1}{2}\mathrm{FIC}_{\mathrm{SR}})}\,. \tag{12}$$

All simulation results given below are based on 1000 replications with $n = 100$ or 200.

Table 1 presents the estimated mean squared errors of the estimated four parameters given by seven methods with $c = 0, 3$ and 5. In the table, p-AIC, p-BIC and p-FIC denote the estimators defined in Section 4 based on the models chosen by the AIC, the BIC and the FIC procedures, while s-AIC, s-BIC and s-FIC are the FMA estimators based on the three weight functions described above, respectively. It is easy to see that the p-FIC and s-FIC have similar MSE, but both gave smaller MSE than the other five estimators. In other words, the results suggest that the FIC procedure yields more accurate estimates of the covariate effects of interest than the AIC, the BIC and Tong's procedures. The results also indicate that although not clear based on the FIC, the FMA procedure is clearly better and more accurate than those based on the AIC and the BIC. We also considered different set-ups and obtained similar results.

6 An Illustrative Example

In this section, we provide an illustrative example by applying the methodology proposed in the previous sections to the bladder cancer data discussed above (Sun & Zhao, 2013). In this study, the patients were randomly allocated to one of three treatments: placebo, thiotepa and pyridoxine. In addition, for each patient, the information is available on two potentially important covariates, the number of initial tumors and the size of the largest initial tumor. One main goal of the study is to evaluate the effects of the treatments and the two covariates on the recurrence rates of the bladder tumors. As mentioned before, only panel count data were observed for the recurrences of the bladder tumors. For the analysis below, following Tong *et al.* (2009), we will confine ourselves to the patients in the placebo (47) and thiotepa (38) groups.

To apply the proposed approach, for patient i, let $N_i(t)$ denote the cumulative number of bladder tumours that have occurred up to time t. In addition, define $Z_{i1} = 1$ if the patient was assigned to the thiotepa group and 0 otherwise, and Z_{i2} Z_{i3} to be the number of initial tumors and the size of the largest initial tumor, respectively. Also following Tong *et al.* (2009), we define the covariate vector to be $Z_i = (Z_{i1}, Z_{i2}, Z_{i3}, Z_{i2}^2, Z_{i3}^2, Z_{i2}Z_{i3})^\top$. For the analysis, we will focus on the three main effects represented by the three parameters $\nu_1 = \beta_1$, $\nu_2 = \beta_2$ and $\nu_3 = \beta_3$ and include them in all sub-models. Table 2 presents the estimated effects given by the AIC, the BIC and the FIC procedures and the 95% confidence intervals obtained by using the FMA procedure. For comparison, the results based on the full model are also obtained and included in the table. Note that in the table, the columns under Selection and FMA correspond, respectively, to the estimates obtained without and with the model average.

One can see from Table 2 that all methods indicate that the thiotepa treatment is effective in decreasing the recurrence rate of bladder tumors and the number of initial tumors is positively related to the tumor recurrence rate. However, the size of the largest initial tumor do not have a significant effect on the tumor recurrence. Note that there are some disagreements between the CI based on the full model and CIs based on FMA, especially for parameters ν_2 and ν_3. Theoretically, all CIs should have the right coverage for large sample if the full model is correct. The CIs based on FMA are asymptotically equivalent to the CI from the full model, i.e., the difference between each confidence limits is $o_P(n^{-1/2})$ (Kabaila & Leeb, 2006; Wang & Zhou, 2013). In spite of the difference in the CIs, they all give consistent conclusions for the example. To give more details about the model selection, Table 3 gives the selected covariates by each method. They show that both the AIC and the BIC chose the smallest model, while the FIC gave different models based on the parameter of interest as expected. The conclusions here on the three parameters of interest are similar to those given in Tong *et al.* (2009).

7 Concluding Remarks

This paper discussed regression analysis of panel count data with the focus on both model selection and parameter estimation. For the problem, we generalized the FIC and the FMA procedures developed in Claeskens & Hjort (2003) and Hjort & Claeskens (2003) among others. In particular, we established the asymptotic distribution of the proposed FMA estimators and provided a method for interval estimation. Also a simulation study was performed for the assessment of the proposed methods and indicates that they can yield more accurate estimates than the existing procedures.

Although the methodology proposed is based on the estimation procedure given in Sun & Wei (2000), it is easy to see that the idea can be applied to other similar estimation procedures or problems. In other words, it is straightforward to generalize the FIC and the FMA procedures presented here to other methods for panel count data or more general situations. In addition, note that the definition of the FIC does not depend on the likelihood. Thus instead of the proportional mean model (1), one could apply the idea to other models such as the additive mean model or the transformation model (Sun & Zhao, 2013). The same is true on the assumption for the observation process. It is not difficult to generalize the proposed method to the situation where the observation process follows the proportional mean or rate model (Cook & Lawless, 2007) instead of the intensity model (2).

Note that in the proposed method, it has been assumed that the parameter of interest ν depends only on β. It is apparent that sometimes it may also depend on the unknown baseline mean function $\mu_0(t)$ and thus it would be useful to generalize the presented method to this more general case. However, this would not be easy as it involves the estimation of $\mu_0(t)$. In addition, the estimated baseline mean function may have a different convergence rate than the estimate of β. Another direction for future research is that in this paper, we have assumed that the follow-up time is independent of the underlying recurrent event process of interest. Sometimes this may not be true and in this situation, a common approach is to develop some joint modeling methods (Sun & Zhao, 2013). The weight function for s-FIC in (12) is different from Hjort & Claeskens (2003) and Zhang *et al.* (2012). In Hjort & Claeskens (2003), the FIC value is scaled by a scale parameter corresponding to the full model and there is another algorithmic parameter κ which is often set to 1. We tried various values of κ and found that the s-FIC performs better when κ is equal to the scale parameter. However, there is no theoretically justification for our choice of the weight function. The selection of the weight functions for the FMA estimator, especially the optimal weights, is another topic for future research. The existing literature on this is mainly on linear models. As pointed out by Hjort & Claeskens (2003), existing bootstrapping methods do not work in the local mis-specification framework for FMA. Therefore, it would be interesting and useful to investigate how to construct valid bootstrapped CIs for the FMA approaches.

Acknowledgements

The authors thank the Associate Editor and two referees for their insightful comments and suggestions, which led to significant improvements in the paper.

Supporting information: Proofs and technical details are given in the appendix in the supplementary material.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *2nd International Symposium on Information Theory* , (B. N. Petrov and F. Cźaki, eds), pp. 267–281. Akademiai Kaidó, Budapest.

Andersen, P. K. & Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10**, 1100–1120.

Balakrishnan, N. & Zhao, X. (2011). A class of multi-sample nonparametric tests for panel count data. *Ann. Inst. Statist. Math.* **63**, 135–156.

Cheng, S. C. & Wei, L. J. (2000). Inferences for a semiparametric model with panel data. *Biometrika* **87**, 89–97.

Claeskens, G. & Hjort, N. L. (2003). The focused information criterion. *J. Amer. Statist. Assoc.* **98**, 900–916.

Claeskens, G. & Hjort, N. L. (2008). *Model selection and model averaging.* Cambridge University Press, New York.

Cook, R. J. & Lawless, J. F. (2007). *The statistical analysis of recurrent events.* Springer, New York.

Ghosh, D. & Lin, D. Y. (2002). Marginal regression models for recurrent and terminal events. *Statist. Sinica* **12**, 663–688.

He, X., Tong, X., Sun, J. & Cook, R. J. (2008). Regression analysis of multivariate panel count data. *Biostat.* **9**, 234–248.

Hjort, N. L. & Claeskens, G. (2003). Frequestist model average estimators. *J. Amer. Statist. Assoc.* **98**, 879–899.

Hjort, N. L. & Claeskens, G. (2006). Focused information criteria and model averaging for the Cox hazard regression model. *J. Amer. Statist. Assoc.* **101**, 1449–1464.

Kabaila, P. & Leeb, H. (2006). On the large-sample minimal coverage probability of confidence intervals after model selection. *J. Amer. Statist. Assoc.* **101**, 619–629.

Kalbfleisch, J. D. & Lawless, J. F. (1985). The analysis of panel data under a markov assumption. *J. Amer. Statist. Assoc.* **80**, 863–871.

Li, N., Park, D.-H., Sun, J. & Kim, K. (2011). Semiparametric transformation models for multivariate panel count data with dependent observation process. *Canad. J. Statist.* **39**, 458–474.

Park, D.-H., Sun, J. & Zhao, X. (2007). A class of two-sample nonparametric tests for panel count data. *Comm. Statist. Theory Methods* **36**, 1611–1625.

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6**, 461–464.

Sun, J. & Fang, H.-B. (2003). A nonparametric test for panel count data. *Biometrika* **90**, 199–208.

Sun, J. & Kalbfleisch, J. D. (1993). The analysis of current status data on point processes. *J. Amer. Statist. Assoc.* **88**, 1449–1454.

Sun, J. & Kalbfleisch, J. D. (1995). Estimation of the mean function of point processes based on panel count data. *Statist. Sinica* **5**, 293–302.

Sun, J. & Wei, L. J. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **62**, 293–302.

Sun, J. & Zhao, X. (2013). *The statistical analysis of panel count data*. Springer, New York.

Thall, P. F. & Lachin, J. M. (1988). Analysis of recurrent events: Nonparametric methods for random-interval count data. *J. Amer. Statist. Assoc.* **83**, 339–347.

Tong, X., Sun, L., He, X. & Sun, J. (2009). Variable selection for panel count data via non-concave penalized estimating function. *Scand. J. Stat.* **36**, 620–635.

Wang, H. & Zhou, S. Z. (2013). Interval estimation by frequentist model averaging. *Comm. Statist. Theory Methods* **42**, 4342–4356.

Wang, H. & Zou, G. (2012). Frequentist model averaging estimation for linear errors-in-variables model. *Journal of System Science and Mathematical Science, Chinese Series* **32**, 1–14.

Wang, H., Zou, G. & Wan, A. T. K. (2012). Model averaging for varying-coefficient partially linear measurement error models. *Electron. J. Stat.* **6**, 1017–1039.

Wei, L. J., Lin, D. Y. & Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Amer. Statist. Assoc.* **84**, 1065–1073.

Wellner, J. A. & Zhang, Y. (2000). Two estimators of the mean of a counting process with panel count data. *Ann. Statist.* **28**, 779–814.

Wellner, J. A. & Zhang, Y. (2007). Two likelihood-based semiparametric estimation methods for panel count data with covariates. *Ann. Statist.* **35**, 2106–2142.

Wu, T. T. & He, X. (2012). Coordinate ascent for penalized semiparametric regression on high-dimensional panel count data. *Comput. Statist. Data Anal.* **56**, 25 – 33.

Zhang, H., Sun, J. & Wang, D. (2013). Variable selection and estimation for multivariate panel count data via the seamless-l_0 penalty. *Canad. J. Statist.* **41**, 368–385.

Zhang, X. & Liang, H. (2011). Focused information criterion and model averaging for generalized additive partial linear models. *Ann. Statist.* **39**, 174–200.

Zhang, X., Wan, A. T. K. & Zhou, S. Z. (2012). Focused information criteria, model selection and model averaging in a tobit model with a non-zero threshold. *J. Bus. Econom. Statist.* **30**, 132–142.

Zhang, Y. (2002). A semiparametric pseudolikelihood estimation method for panel count data. *Biometrika* **89**, 39–48.

Zhang, Y. (2006). Nonparametric k-sample tests with panel count data. *Biometrika* **93**, 777–790.

Zhao, X. & Sun, J. (2011). Nonparametric comparison for panel count data with unequal observation processes. *Biometrics* **67**, 770–779.

HaiYing Wang, Department of Mathematics and Statistics, University of New Hampshire, Durham, NH 03824, USA

haiying.wang@unh.edu

Table 1: Estimated Mean Square Errors based on simulated data

	c	p-AIC	p-BIC	p-FIC	s-AIC	s-BIC	s-FIC	Tong
				$n = 100$				
ν_1	0	0.352	0.326	**0.201**	0.337	0.319	0.206	0.348
	3	0.453	0.439	0.204	0.445	0.433	**0.200**	0.458
	5	0.311	0.308	0.258	0.306	0.298	**0.246**	0.319
ν_2	0	0.536	0.498	**0.267**	0.520	0.491	0.273	0.497
	3	0.613	0.584	**0.216**	0.598	0.572	0.225	0.572
	5	0.443	0.415	**0.218**	0.429	0.407	0.225	0.407
ν_3	0	0.372	0.351	**0.185**	0.359	0.343	0.192	0.365
	3	0.389	0.375	0.232	0.379	0.366	**0.229**	0.387
	5	0.326	0.315	0.268	0.320	0.311	**0.255**	0.332
ν_4	0	0.257	0.239	**0.154**	0.246	0.230	0.164	0.232
	3	0.326	0.306	**0.143**	0.314	0.298	0.150	0.297
	5	0.247	0.233	0.134	0.236	0.227	**0.133**	0.227
				$n = 200$				
ν_1	0	0.086	0.080	**0.065**	0.083	0.078	0.067	0.082
	3	0.093	0.092	0.086	0.092	0.091	**0.082**	0.091
	5	0.108	0.109	0.106	0.107	0.107	**0.100**	0.110
ν_2	0	0.134	0.126	**0.100**	0.131	0.124	0.104	0.123
	3	0.137	0.129	**0.091**	0.134	0.126	0.095	0.127
	5	0.136	0.129	**0.086**	0.132	0.126	0.090	0.125
ν_3	0	0.101	0.095	**0.074**	0.098	0.094	0.077	0.096
	3	0.098	0.094	0.083	0.096	0.093	**0.081**	0.095
	5	0.094	0.094	0.097	0.093	0.092	**0.089**	0.094
ν_4	0	0.069	0.066	**0.053**	0.067	0.064	0.056	0.064
	3	0.080	0.074	**0.052**	0.076	0.072	0.057	0.072
	5	0.080	0.076	**0.056**	0.076	0.073	0.057	0.072

p-AIC, p-BIC and p-FIC are estimators obtained from the model chosen by the AIC, the BIC and the FIC, respectively; s-AIC, s-BIC and s-FIC are frequentist model averaging estimators given in (10) with weight functions defined in (12); Tong refers to the method given in Tong et al. (2009).

Table 2: Estimated covariate effects for the bladder tumour study

		Point estimates		95% CI based on FMA
		Selection	FMA	and the full model
ν_1	AIC	-0.986	-0.967	(-1.080, -0.853)
	BIC	-0.986	-0.978	(-1.092, -0.865)
	FIC	-0.963	-0.958	(-1.072, -0.845)
	FULL	-0.926		(-1.040, -0.813)
ν_2	AIC	0.660	0.832	(0.448, 1.166)
	BIC	0.660	0.726	(0.334, 1.052)
	FIC	0.640	0.654	(0.257, 0.975)
	FULL	1.228		(0.870, 1.587)
ν_3	AIC	-0.123	-0.011	(-0.491, 0.446)
	BIC	-0.123	-0.080	(-0.562, 0.374)
	FIC	-0.256	-0.189	(-0.674, 0.263)
	FULL	0.245		(-0.223, 0.714)

Table 3: Selected models for the bladder tumour study

Method	Parameter	Components Selected	
		β	γ
AIC	ν_1, ν_2, ν_3	1,2,3	1,2,3
BIC	ν_1, ν_2, ν_3	1,2,3	1,2,3
FIC	ν_1	1,2,3,5	1,2,3,5,6
	ν_2	1,2,3,5	1,2,3,4,5,6
	ν_3	1,2,3,4	1,2,3,4,5,6