Computational Statistics and Data Analysis 57 (2013) 33–40

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Robust estimation for panel count data with informative observation times

Xingqiu Zhao[a,*], Xingwei Tong[b], Jianguo Sun[c]

[a] Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong
[b] Department of Statistics, Beijing Normal University, Beijing, 100875, China
[c] Department of Statistics, University of Missouri, Columbia, MO, 65211, USA

ARTICLE INFO

Article history:
Received 22 September 2011
Received in revised form 14 May 2012
Accepted 17 May 2012
Available online 30 May 2012

Keywords:
Informative observation process
Panel count data
Recurrent event process
Robust estimation

ABSTRACT

Panel count data usually occur in longitudinal follow-up studies that concern occurrence rates of certain recurrent events and their analysis involves two processes. One is the underlying recurrent event process of interest and the other is the observation process that controls observation times. In some situations, the two processes may be correlated and, for this, several estimation procedures have recently been developed (He et al., 2009; Huang et al., 2006; Sun et al., 2007b; Zhao and Tong, 2011). These methods, however, rely on some restrictive models or assumptions such as the Poisson assumption. In this work, a more general and robust estimation approach is proposed for regression analysis of panel count data with related observation times. The asymptotic properties of the resulting estimates are established and the numerical studies conducted indicate that the approach works well for practical situations.

1. Introduction

This work considers regression analysis of panel count data when the observation times or process may be related to the underlying recurrent event process of interest governing the panel count data. By panel count data, we mean the data that concern occurrence rates of certain recurrent events and give only the numbers of the events that occur between the observation times, but not their occurrence times. Such data naturally occur in longitudinal follow-up studies on recurrent events in which study subjects can be observed only at discrete time points rather than continuously (Cai and Schaubel, 2004; Cook and Lawless, 2007; Sun, 2006).

Many authors have discussed the analysis of panel count data when the recurrent event process of interest and the observation process are independent completely or conditional on covariates. In this case, the inference can be made conditional on the observation process. For example, Sun and Kalbfleisch (1995), Wellner and Zhang (2000) and Hu et al. (2009) studied nonparametric estimation of the mean function of the underlying counting process yielding panel counts. The same problem was also considered by Zhang and Jamshidian (2003) and Lu et al. (2007). The former employed a gamma frailty variable to account for the correlation among panel counts and developed a maximum pseudo-likelihood approach, while the latter also gave some likelihood-based estimators of the mean functions by using monotone polynomial splines. In addition, Sun and Fang (2003), Zhang (2006) and Balakrishnan and Zhao (2009) constructed some nonparametric tests for nonparametric comparison of the mean functions of counting processes. Cheng and Wei (2000), Sun and Wei (2000) and Hu et al. (2003) proposed some semiparametric models for regression analysis of panel count data and developed some

* Corresponding author. Tel.: +852 27666921; fax: +852 23629045.
E-mail addresses: xingqiu.zhao@polyu.edu.hk (X. Zhao), xweitong@bnu.edu.cn (X. Tong), sunj@missouri.edu (J. Sun).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2012.05.015

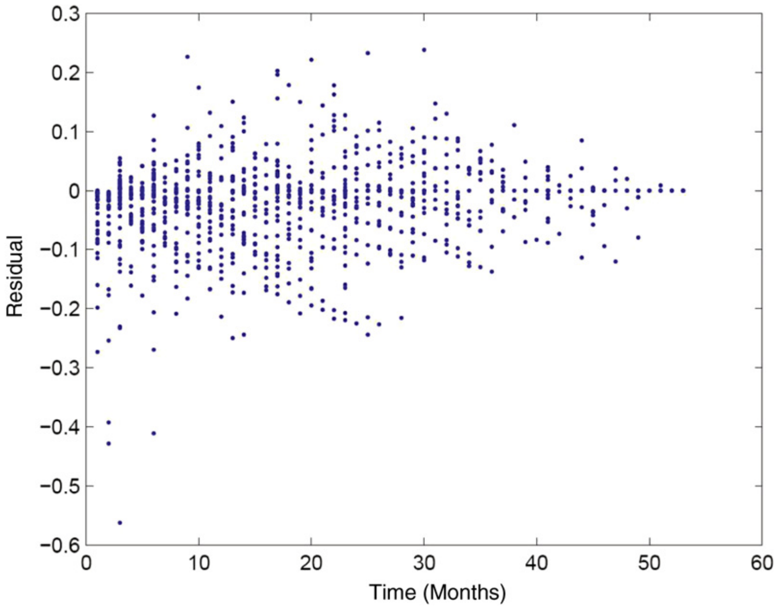

Fig. 1. Plot of the residuals for fitted model (2) with the bladder tumor data.

estimating equation-based approaches. Zhang (2002), Wellner and Zhang (2007) and Lu et al. (2009) discussed the same regression problem and gave some semiparametric likelihood-based approaches.

Sometimes the recurrent event process of interest and the observation process may be related. A well-known example of such panel count data is the bladder cancer data discussed in He et al. (2009), Huang et al. (2006), Wellner and Zhang (2007), Liang et al. (2009), Lu et al. (2009) and Sun et al. (2007b), among others. The data concern the occurrence rate of bladder tumors and during the study giving rise to the data, study patients were observed periodically and at different time points. Some patients were observed more often than others and several authors have showed that the occurrence process of bladder tumors seems to be correlated with the observation process. More details about the example are given in Section 5. For the analysis of such panel count data, several methods have been developed (He et al., 2009; Huang et al., 2006; Sun et al., 2007b; Zhao and Tong, 2011). A common and key assumption behind these methods is that the observation process is a Poisson process.

To be more specific about this, consider a recurrent event study and suppose that only panel count data are available. Let $N(t)$ and $O(t)$ denote the process of interest and the observation process, respectively, and X a vector of the covariates of interest. Then $N(t)$ is observed only at the times where $O(t)$ jumps. Suppose that $N(t)$ and $O(t)$ may be related even given X. For inference, Sun et al. (2007b) assumed that given X and a latent variable Z, the mean function of $N(t)$ has the form

$$E\{N(t)|X,Z\} = Z^{\alpha}\,\mu_0(t)\exp(X'\beta) \tag{1}$$

and $O(t)$ is a non-homogeneous Poisson process with the intensity function

$$\lambda(t|X,Z) = Z\,\lambda_0(t)\exp(X'\gamma). \tag{2}$$

In the above, β, α and γ are unknown parameters, and $\mu_0(t)$ and $\lambda_0(t)$ are unknown baseline mean and intensity functions, respectively. To examine the Poisson process assumption for the bladder cancer data, we fitted the data on the observation times to model (2) and present the residuals in Fig. 1. Also we developed a simple Kolomogorov–Smirnov test statistic procedure (Gibbons and Chakraborti, 2011) and obtained the p-value of 0.07 for testing the Poisson process assumption. Both the figure and the test suggest that the Poisson process assumption with model (2) may be questionable. In the following, we will relax this and other assumptions and develop a general and robust inference approach.

Note that a number of methods have been developed for regression analysis of longitudinal data, mostly under the assumption that the longitudinal response and the observation process are independent completely or given covariates. For example, Diggle et al. (1994) provided a comprehensive summary about the commonly used methods such as estimating equation and random effect model approaches, and Lin and Ying (2001) and Welsh et al. (2002) discussed general semiparametric regression analysis of longitudinal data. In contrast, limited research exists for regression analysis of longitudinal data where measurement times may be informative or still related to the underlying longitudinal process even given covariates, a problem similar to that discussed here (Sun et al., 2007a; Liu et al., 2008; Liang et al., 2009). Although panel count data can be regarded as a special type of longitudinal data, the use of the methods developed for longitudinal data may not be valid or efficient as they do not take into account the special structure of panel count data.