RANGE : Jurnal Pendidikan Matematika
Dikelolah Oleh Program Studi Pendidikan Matematika Universitas Timor
ISSN Online : 2685-2373

Volume 5, Nomor 2, Januari 2024, pp. 24-33.

Survival Model Estimator for Type II Censored Data Based on Lognormal Distribution

Eduardus Beo Seso Delvion[1]*, Cecilia Novianti Salsinha[2], Minanur Rohman[3]
University of Timor[1,2)], Ondukuz Mayis University[3)]
*ed_del@unimor.ac.id

Received: November 18th, 2023. Accepted: January 2nd, 2024. Published: January 30th, 2024.

ABSTRAK

Analisis survival merupakan teknik statistik yang berguna menganalisis variabel random positif mencakup waktu tahan hidup dan waktu kegagalan.Permasalahan yang terjadi dalam menganalisis model survival berdasarkan distribusiparametrik diperlukan nilai parameter yang diperoleh dari estimator. Penelitian ini bertujuan untuk mendapatkan model estimator dari distribusi lognormal yang memiliki parameter $\mu_{\ln t}$ dan $\sigma_{\ln t}$. Oleh karena itu, digunakan pendekatan metode maksimum *likelihood* untuk memperoleh estimator$\hat{\mu}_{\ln t}$ dan $\hat{\sigma}_{\ln t}$. Namun kendala yang ditemui bahwa model maksimum *likelihood* dari distribusi lognormal untuk data *cencoringType* IImenghasilkan persamaan yang cukup kompleks sehingga diperlukan pendekatan dengan deret Taylor hanya untuk orde pertama. Hasil yang diperoleh adalah estimator $\hat{\mu}_{\ln t} = \alpha + \beta\sigma_{\ln t}$ dan $r\hat{\sigma}_{\ln t}{}^2 - d_1\hat{\sigma}_{\ln t} - d_2 = 0$ yang dapat diaplikasikan pada data *cencoringType* II berdasarkan distribusi lognormal.

Kata Kunci:analisis survival, data *cencoringType* II, metode maksimum *likelihood*, deret Taylor, distribusi lognormal.

ABSTRACT

Survival analysis is also a statistical technique that is useful for analyzing positive random variables including survival time and failure time.The problems that occur in analyzing survival models based on parametric distributions require parameter values obtained from estimators. This study aims to obtain an estimator model of the lognormal distribution which has parameters $\mu_{ln\,t}$*and* $\sigma_{ln\,t}$*. Therefore, the maximum likelihood method approach is used to obtain the estimators*$\hat{\mu}_{ln\,t}$*and* $\hat{\sigma}_{ln\,t}$*. However, the problem encountered is that the maximum likelihood model of the lognormal distribution for Type II cencoring data produces quite complex equations, so an approach with a Taylor series is needed only for the first order. The results obtained are estimators*$\hat{\mu}_{ln\,t} = \alpha + \beta\sigma_{ln\,t}$ *and* $r\hat{\sigma}_{ln\,t}{}^2 - d_1\hat{\sigma}_{ln\,t} - d_2 = 0$*which can be applied to Type II cencoring data based on a lognormal distribution.*

Keywords:*survival analysis, Type II cencoring data, maximum likelihood method, Taylor series, lognormal distribution.*

How to Cite: Delvion, E. B. S., Salsinha, C. N.& Rohman, M. (2024). Survival Model Estimator For Type II Censored Data Based On Lognormal Distribution. *Range: Jurnal Pendidikan Matematika*, 5 (2), 24-33.

Introduction

The survival analysis is one of the statistical techniques used to analyze the survival time of an object within a certain period. The survival time used in this analysis is based on research data obtained from the studied sample (P. Wang et al., 2019). The data includes Type I and II censoring, depending on the casewhether the observed object survives until a specified time or the observed object's survival approaches zero. The number of simultaneously tested data, n, is stopped after obtaining r number of deaths within the data interval, where $1 \leq r \leq n$(Lawless, 2003).

Survival data requires a parametric data distribution model, necessitating assumptions about the distribution of the population. Distributions that can be used to analyze survival models include normal, log-normal, logistic, exponential, gamma distributions, and others (Dutta & Kayal, 2022). In this study, the log-normal distribution is used, which has two parameters, $\mu_{\ln t}$and, determined by their respective estimators based on Type II censoring data. This is because the log-normal distribution model is effective in health-related issues, such as cancer patients, where the probability of survival decreases to zero following the survival model curve. The determination of the estimator for the log-normal distribution with Type II censoring data is approached using the likelihood method defined as follows(Monteiro, 2015).

Type II censored data is r data obtained from observations of n observed samples and the experiment will be stopped after the r-th failure occurs. In this study, there are n data and for example T is a random variable from n observed, $f(t_1), f(t_2), \cdots, f(t_r)$ is the density function The probability of a component failing from the 1st to the r-th time while the component that still survives beyond the time of the r-th component is written by $T_{r+1}, T_{r+2}, \cdots, T_n$ as $n-r$, so that we get the joint probability density function of$T_{r+1}, T_{r+2}, \cdots, T_n$from the observed data is as follows:

Let $T_1, T_2, \cdots, T_n$ be a random sample with $f(T_i;\ \theta_1, \theta_1 \cdots \theta_k)$ as the parameter (Hamilton, 1994). The likelihood function is then defined as:

$$L(T;\ \theta_1, \theta_1 \cdots \theta_k) = \prod_{i=1}^{n} f(T_i; \theta_1, \theta_1 \cdots \theta_k) \tag{1}$$

The survival function is the probability that a component can survive until time $t > 0$. So if T is a random variable that denotes the reliability time of a component, then $R(t)$is the probability that a component can last longer than time $t(T > t)$.The log-normal distribution, also referred to as Galton's distribution, represents a continuous probability distribution for a random variable whose natural logarithm follows a normal distribution(Gross & Holland, 1968). When the random variable$t$follows a log-normal distribution, denoted as, $T = \ln t$, the probability density function is expressed as follows:

$$f(t; \mu_{\ln t}, \sigma_{\ln t}) = \frac{1}{t\sigma_{\ln t}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln t - \mu_{\ln t}}{\sigma_{\ln t}}\right)^2}; t > \tag{2}$$

where t is a random variable with $\mu_{\ln t} \in$ R and $\sigma_{\ln t} > 0$ as the mean and variance (Seso Delvion, 2019).

Research Methodology

The method used in this research is a literature review. Through this method, references regarding the determination of estimators using the maximum likelihood approach based on the Lognormal distribution model for Type II censoring data will be examined.

Study Procedure

The steps taken include identifying material from various references to determine estimators using the maximum likelihood approach based on the Lognormal distribution model for Type II censoring data, reviewing the survival model for Type II censoring data based on the Lognormal distribution, and determining the estimator for the survival model for Type II censoring data based on the Lognormal distribution. The steps are as follows:

1. Determining the Likelihood function of the Lognormal distribution (2) with the formula(Chattamvelli & Shanmugam, 2023):

$$L(t; \mu_{\ln t}, \sigma_{\ln t}) = \prod_{i=1}^{n} f(t_i; \mu_{\ln t}, \sigma_{\ln t}) \tag{3}$$

2. Forming the natural logarithm function of equation (1) with the formula(Y. Wang et al., 2022):

$$\ln L(t; \mu_{\ln t}, \sigma_{\ln t}) = \ln \prod_{i=1}^{n} f(t_i; \mu_{\ln t}, \sigma_{\ln t}) \tag{4}$$

3. Lowering equation (4) to determine the estimators for$\hat{\mu}_{\ln t}$and$\hat{\sigma}_{\ln t}$thereby allowing the determination of parameter values$\mu_{\ln t}$and$\sigma_{\ln t}$for Type II censoring data based on the Lognormal distribution with the following formula(Jenatabadi et al., 2020):

$$\frac{dL}{d\mu_{\ln t}} = 0 \tag{5}$$

and

$$\frac{dL}{d\sigma_{\ln t}} = 0 \tag{6}$$

Results and Discussion

The survival model for Type II censoring data can be expressed as follows:

$$S(t) = 1 - F(t) \tag{7}$$

$$= 1 - \int_0^t f(t; \mu_{\ln t}, \sigma_{\ln t})\, dt$$

$$= 1 - \int_0^t \frac{1}{t\sigma_{\ln t}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln t - \mu_{\ln t}}{\sigma_{\ln t}}\right)^2} dt$$

$$= 1 - \Phi\left[\frac{\ln t - \mu_{\ln t}}{\sigma_{\ln t}}\right]$$

Where $F(t)$ is the cumulative distribution function of the lognormal distribution (Lawless, 1982). In Figure 1, let T be a random variable representing of the survival time of$n$ observed individuals., $f(t_1)$ is the probability density function of the survival time for the first individual, and similarly$f(t_r)$for the$r$-th individual who failed or died (Lebreton et al., 1992). For individuals still surviving, they can be denoted as$T_{r+1}, \cdots, T_n$totaling$n - r$individuals(Bain & Engelhardt, 1992).

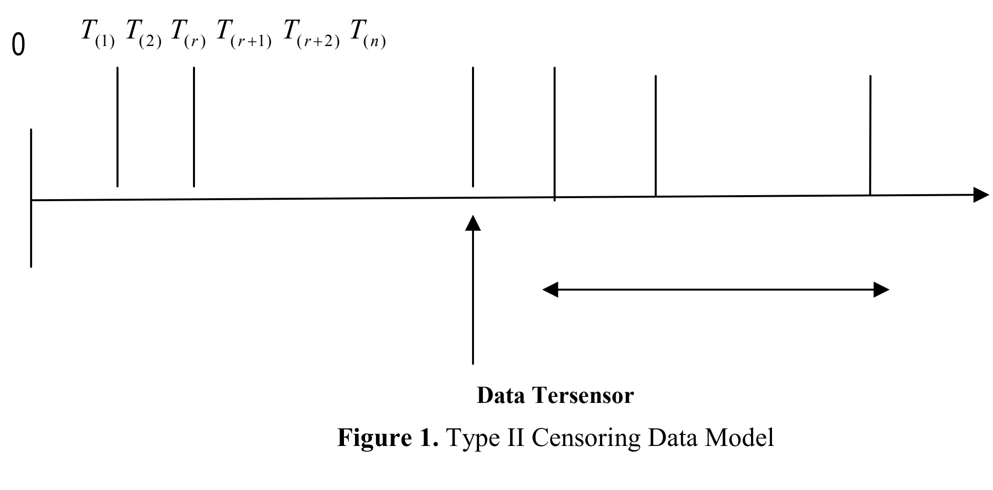

Figure 1. Type II Censoring Data Model

Due to the fact that this sample follows a multinomial distribution, the joint probability density function of $T_1, T_2, \cdots, T_n$ is as follows:

$$f(t_1, \cdots, t_r) = \frac{n!}{(n-r)!} f(t_1) \cdots f(t_r)[P(T_{r+1} \geq t_r) \cdots P(T_n \geq t_r)]$$
$$= \frac{n!}{(n-r)!}\left[\prod_{i=1}^{r} f(t_i)\right][S(t_r)]^{n-r}$$
$$= \frac{n!}{(n-r)!}\left[\prod_{i=1}^{r} \frac{1}{t\sigma_{\ln t}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln t_i - \mu_{\ln t}}{\sigma_{\ln t}}\right)^2}\right]\left[1 - \Phi\left[\frac{\ln t_r - \mu_{\ln t}}{\sigma_{\ln t}}\right]\right]^{n-r}$$

Thus, the likelihood function of the lognormal distribution for Type II censored data is obtained as follows:

$$L(t; \mu_{\ln t}, \sigma_{\ln t}) = \frac{n!}{(n-r)!}\left[\left(\frac{1}{t\sigma_{\ln t}\sqrt{2\pi}}\right)^r e^{-\frac{1}{2}\sum_{1=i}^{r}\left(\frac{\ln t_i - \mu_{\ln t}}{\sigma_{\ln t}}\right)^2}\right]$$
$$\left[1 - \Phi\left[\frac{\ln t_r - \mu_{\ln t}}{\sigma_{\ln t}}\right]\right]^{n-r} \tag{8}$$

From equations (5) and (6), the partial derivatives of the logarithm of the function (8) with respect to $\mu_{\ln t}$ and $\sigma_{\ln t}$ are obtained as follows:

$$\frac{d(\ln L(t; \mu_{\ln t}, \sigma_{\ln t}))}{d\mu_{\ln t}} = 0$$

$$\frac{d\left(\ln \frac{n!}{(n-r)!} + \ln \prod_{i=1}^{r} t_i - \frac{r}{2} \ln 2\pi(\sigma_{\ln t})^2 - \frac{1}{2}\sum_{1=i}^{r}\left(\frac{\ln t_i - \mu_{\ln t}}{\sigma_{\ln t}}\right)^2 + (n-r)\ln\left[1 - \Phi\left[\frac{\ln t_r - \mu_{\ln t}}{\sigma_{\ln t}}\right]\right]\right)}{d\mu_{\ln t}}$$
$$= 0 \tag{9}$$

$$\frac{d\left(\ln\frac{n!}{(n-r)!}+\ln\prod_{i=1}^{r}t_i-\frac{r}{2}\ln 2\pi(\sigma_{\ln t})^2-\frac{1}{2}\sum_{1=i}^{r}\left(\frac{\ln t_i-\mu_{\ln t}}{\sigma_{\ln t}}\right)^2+(n-r)\ln\left[1-\Phi\left[\frac{\ln t_r-\mu_{\ln t}}{\sigma_{\ln t}}\right]\right]\right)}{\sigma_{\ln t}}$$

$$= 0 \tag{10}$$

From equations (9) and (10), we have nonlinear equations, thus requiring the Taylor series expansion method for censored data from the lognormal distribution(Kundu, 2007). Let the first derivative for the function$h(z_r) = \frac{\phi(z_r)}{1-\Phi(z_r)}$be denoted as follows, neglecting second and higher-order derivatives:

$$h(z_r) \approx h(s) + (z_r - s) = j + kz_r \tag{11}$$

where$j = h(s) - sh'(v)$and$k = h'(v) \geq 0$. From equation (9), we obtain:

$$\frac{1}{\sigma_{\ln t}}\sum_{1=i}^{r}\left(\frac{\ln t_i - \mu_{\ln t}}{\sigma_{\ln t}}\right) + (n-r)\left(\frac{j + kz_r}{\sigma_{\ln t}}\right) = 0 \tag{12}$$

From equation (12), we obtain:

$$\left(\sum_{1=i}^{r}\ln t_i\right) - r\mu_{\ln t} + (n-r)\big(j\sigma_{\ln t} + k(\ln t_r - \mu_{\ln t})\big) = 0$$

$$\left(\sum_{1=i}^{r}\ln t_i\right) - r\mu_{\ln t} + (n-r)k\mu_{\ln t} + (n-r)(j\sigma_{\ln t}) + (n-r)k\ln t_r = 0$$

$$\left(\sum_{1=i}^{r}\ln t_i\right) - \mu_{\ln t}(r + (n-r)k) + (n-r)(j\sigma_{\ln t}) + (n-r)k\ln t_r = 0$$

So,

$$\mu_{\ln t}(r + (n-r)k) = \left(\sum_{1=i}^{r}\ln t_i\right) + (n-r)(j\sigma_{\ln t}) + (n-r)k\ln t_r$$

$$\mu_{\ln t} = \frac{(\sum_{1=i}^{r}\ln t_i) + (n-r)(j\sigma_{\ln t}) + (n-r)k\ln t_r}{(r + (n-r)k)}$$

$$= \frac{(\sum_{1=i}^{r}\ln t_i) + (n-r)k\ln t_r}{(r + (n-r)k)} + \frac{(n-r)j}{(r + (n-r)k)}\sigma_{\ln t}$$

$$\mu_{\ln t} = \alpha + \beta\sigma_{\ln t} \tag{13}$$

Equation (13) serves as an estimator in determining the value of the parameter$\mu_{\ln t}$for Type II censored data. Where

$$\alpha = \frac{(\sum_{1=i}^{r} \ln t_i) + (n-r)k \ln t_r}{(r + (n-r)k)}$$

and

$$\beta = \frac{(n-r)j}{(r+(n-r)k)}.$$

From equation (10), we obtain:

$$-\frac{r}{\sigma_{\ln t}} + \frac{1}{\sigma_{\ln t}{}^3}\left(\sum_{1=i}^{r} (\ln t_i - \mu_{\ln t})^2\right) + (n-r)z_r\left(\frac{j + kz_r}{\sigma_{\ln t}}\right) = 0$$

$$-r\sigma_{\ln t}{}^2 + \left(\sum_{1=i}^{r} (\ln t_i - \mu_{\ln t})^2\right) + (n-r)(\ln t_r - \mu_{\ln t})\big(j\sigma_{\ln t} + k(\ln t_r - \mu_{\ln t})\big) = 0$$

$$-r\sigma_{\ln t}{}^2 + \left(\sum_{1=i}^{r} (\ln t_i - \mu_{\ln t})^2\right) + (n-r)(\ln t_r - \mu_{\ln t})\big(j\sigma_{\ln t} + k(\ln t_r - \mu_{\ln t})\big) = 0 \tag{14}$$

From equation (13) and (14), we obtain:

$$\left(-r\sigma_{\ln t}{}^2 + (n-r)(\ln t_r - \alpha)j\sigma_{\ln t} + (n-r)(\ln t_r - \alpha)^2 + \left(\sum_{1=i}^{r} (\ln t_i - \alpha)^2\right)\right)$$
$$+ \left(-2\beta\sigma_{\ln t}\left(\sum_{1=i}^{r} \ln t_i\right) + 2r\alpha\beta\sigma_{\ln t} + r(\alpha\beta\sigma_{\ln t})^2\right) - (n-r)\beta k\sigma_{\ln t}{}^2$$
$$- 2\beta\sigma_{\ln t}k(n-r) + (\beta\sigma_{\ln t})^2 k(n-r) = 0 \tag{15}$$

From equation $\alpha = \frac{(\sum_{1=i}^{r} \ln t_i)+(n-r)k \ln t_r}{(r+(n-r)k)}$and equation (15), it is obtained that:

$$\sum_{1=i}^{r} \ln t_i = \alpha r - (n-r)k \ln t_r$$

thus

$$\left(-r{\sigma_{\ln t}}^2 + (n-r)(\ln t_r - \alpha)j\sigma_{\ln t} + (n-r)(\ln t_r - \alpha)^2 + \left(\sum_{1=i}^{r}(\ln t_i - \alpha)^2\right)\right)$$
$$+ (-2\beta\sigma_{\ln t}(\alpha r - (n-r)k\ln t_r) + 2r\alpha\beta\sigma_{\ln t} + r(\alpha\beta\sigma_{\ln t})^2) - (n-r)\beta k{\sigma_{\ln t}}^2$$
$$- 2\beta\sigma_{\ln t}k(n-r) + (\beta\sigma_{\ln t})^2 k(n-r) = 0$$

So,

$$\left(\frac{-r{\sigma_{\ln t}}^2 + (n-r)(\ln t_r - \alpha)j\sigma_{\ln t} + (n-r)(\ln t_r - \alpha)^2 + (\sum_{1=i}^{r}(\ln t_i - \alpha)^2)}{\beta}\right) + r\beta{\sigma_{\ln t}}^2$$
$$- (n-r)k{\sigma_{\ln t}}^2 + \beta k(n-r){\sigma_{\ln t}}^2 = 0 \qquad (16)$$

From Equation $\beta = \frac{(n-r)j}{(r+(n-r)k)}$and equation (16), it is obtained that:

$$\beta k(n-r) = (n-r)k - r\beta$$

$$\left(\frac{-r{\sigma_{\ln t}}^2 + (n-r)(\ln t_r - \alpha)j\sigma_{\ln t} + (n-r)(\ln t_r - \alpha)^2 + (\sum_{1=i}^{r}(\ln t_i - \alpha)^2)}{\beta}\right) + r\beta{\sigma_{\ln t}}^2$$
$$- (n-r)k{\sigma_{\ln t}}^2 + \big((n-r)k - r\beta\big){\sigma_{\ln t}}^2 = 0$$

thus

$$r{\sigma_{\ln t}}^2 - (n-r)(\ln t_r - \alpha)j\sigma_{\ln t} - (n-r)(\ln t_r - \alpha)^2 - \left(\sum_{1=i}^{r}(\ln t_i - \alpha)^2\right) = 0$$

Simplified to:

$$r{\sigma_{\ln t}}^2 - d_1\sigma_{\ln t} - d_2 = 0 \qquad (17)$$

where

$$d_1 = j(n-r)(\ln t_r - \alpha) \text{and} d_2 = (n-r)(\ln t_r - \alpha)^2 + (\textstyle\sum_{1=i}^{r}(\ln t_i - \alpha)^2)$$

Therefore, the two estimators for both$\mu_{\ln t}$and$\sigma_{\ln t}$for Type II censored data are obtained as:

$$\hat{\mu}_{\ln t} = \alpha + \beta\sigma_{\ln t}$$

and

$$r{\hat{\sigma}_{\ln t}}^2 - d_1\hat{\sigma}_{\ln t} - d_2 = 0$$

Conclusion

From the obtained results, it can be concluded that determining the parameter values of the lognormal distribution model related to survival models requires careful consideration of the use of data based on the existing problems. In this case, for Type II censored data, the estimators for$\mu_{\ln t}$ and $\sigma_{\ln t}$are obtained as follows: $\hat{\mu}_{\ln t} = \alpha + \beta\sigma_{\ln t}$ and $r\hat{\sigma}_{\ln t}{}^2 - d_1\hat{\sigma}_{\ln t} - d_2 = 0$. For further research, researchers can apply these findings to data related to the durability or survival of an object, such as data on the survival time of a cancer patient, and so on. Additionally, the application can be extended to other fields, such as economics, health, and more.

References

Bain, L. ., & Engelhardt. (1992). *Introduction to Probability and Mathematical Statistics* (D. Press (ed.); 2nd ed.).

Chattamvelli, R., & Shanmugam, R. (2023). Skewness. In *Synthesis Lectures on Mathematics and Statistics: Vol. Part F5*. https://doi.org/10.1007/978-3-031-32330-0_4

Dutta, S., & Kayal, S. (2022). Estimation of parameters of the logistic exponential distribution under progressive type-I hybrid censored sample. *Quality Technology and Quantitative Management*, *19*(2). https://doi.org/10.1080/16843703.2022.2027601

Gross, S., & Holland, P. W. (1968). The Distribution of Galton's Statistic. *The Annals of Mathematical Statistics*, *39*(6). https://doi.org/10.1214/aoms/1177698040

Hamilton, J. (1994). *Time Series Analysis*. Princeton University Press.

Isniah, S., Hardi Purba, H., & Debora, F. (2020). Plan do check action (PDCA) method: literature review and research issues. *Jurnal Sistem Dan Manajemen Industri*, *4*(1). https://doi.org/10.30656/jsmi.v4i1.2186

Jenatabadi, H. S., Radzi, C. W. J. B. W. M., & Samsudin, N. (2020). Associations of body mass index with demographics, lifestyle, food intake, and mental health among postpartum women: A structural equation approach. *International Journal of Environmental Research and Public Health*, *17*(14). https://doi.org/10.3390/ijerph17145201

Kundu, D. (2007). On hybrid censored Weibull distribution. *Journal of Statistical Planning and Inference*, *137*(7). https://doi.org/10.1016/j.jspi.2006.06.043

Lawless, J. F. (1982). *Statistical Model and Methods for Lifetime Data* (1st ed.). John Wiley and Sons, Inc.

Lawless, J. F. (2003). *Statistical Model and Methods for Lifetime Data* (2nd ed.). John Wiley and Sons, Inc.

Lebreton, J. D., Burnham, K. P., Clobert, J., & Anderson, D. R. (1992). Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological Monographs*, *62*(1). https://doi.org/10.2307/2937171

Monteiro, M. J. (2015). Fitting molecular weight distributions using a log-normal distribution model. *European Polymer Journal*, *65*. https://doi.org/10.1016/j.eurpolymj.2015.01.009

Seso Delvion, E. B. (2019). Model Hazard untuk Laju Kerusakan Pipa Air berdasarkan Data Tersensor Tipe II yang Berdistribusi Lognormal. *RANGE: Jurnal Pendidikan Matematika*, *1*(1). https://doi.org/10.32938/jpm.v1i1.195

Wang, P., Li, Y., & Reddy, C. K. (2019). Machine learning for survival analysis: A survey. *ACM Computing Surveys*, *51*(6). https://doi.org/10.1145/3214306

Wang, Y., Wang, W., & Zhao, X. (2022). Local logarithm partial likelihood estimation of panel count data model with an unknown link function. *Computational Statistics and Data Analysis*, *166*. https://doi.org/10.1016/j.csda.2021.107346