Lee, Shen-Ming; Lukusa, T. Martin; Li, Chin-Shang

Estimation of a zero-inflated Poisson regression model with missing covariates via nonparametric multiple imputation methods. (English) [Zbl 1482.62015]


Summary: Zero-inflated Poisson (ZIP) regression is widely applied to model effects of covariates on an outcome count with excess zeros. In some applications, covariates in a ZIP regression model are partially observed. Based on the imputed data generated by applying the multiple imputation (MI) schemes developed by D. Wang and S. X. Chen [Ann. Stat. 37, No. 1, 490–517 (2009; Zbl 1155.62021)], two methods are proposed to estimate the parameters of a ZIP regression model with covariates missing at random. One, proposed by D. B. Rubin [“Multiple imputations in sample surveys: a phenomenological Bayesian approach to nonresponse”, in: Proceedings of the survey research methods section of the American Statistical Association, Vol. 1. Boston: American Statistical Association. 20–28 (1978)], consists of obtaining a unified estimate as the average of estimates from all imputed datasets. The other, proposed by R. E. Fay [J. Am. Stat. Assoc. 91, No. 434, 490–498 (1996; Zbl 0869.62015)], consists of averaging the estimating scores from all imputed data sets to solve the imputed estimating equation. Moreover, it is shown that the two proposed estimation methods are asymptotically equivalent to the semiparametric inverse probability weighting method. A modified formula is proposed to estimate the variances of the MI estimators. An extensive simulation study is conducted to investigate the performance of the estimation methods. The practicality of the methodology is illustrated with a dataset of motorcycle survey of traffic regulations.

MSC:

62-08 Computational methods for problems pertaining to statistics
62D05 Sampling theory, sample surveys

Keywords:
zero-inflated Poisson regression; nonparametric multiple imputation; inverse probability weighting (IPW); missing at random; count data

Full Text: DOI

References:

and the Lee, SM; Li, CS; Hsieh, SH; Huang, LH, Semiparametric estimation of logistic regression model with missing covariates and
Zbl 0344.62034
Johnson, NL; Kemp, AW; Kotz, S., Univariate discrete distributions (2005), New York: Wiley, New York
Lee, SM; Gee, MJ; Hsieh, SH, Semiparametric methods in the proportional odds model for ordinal response data with missing
covariates, Biometrics, 60, 788-798 (2011) - Zbl 1226.62039
Lee, JH; Han, G.; Fulp, WJ; Giuliano, AR, Analysis of overdispersed count data: application to the human papillomavirus
infection in men (HIM) study, Epidemiol Infect, 140, 1087-1094 (2012)
Lee, SM; Li, CS; Hsieh, SH; Huang, LH, Semiparametric estimation of logistic regression model with missing covariates and
outcome, Metrika, 75, 621-653 (2012) - Zbl 1362.62089
Lee, SM; Hwang, WH; Tapsoba, JD, Estimation in closed capture-recapture models when covariates are missing at random,
Biometrics, 72, 1294-1304 (2016) - Zbl 1390.62289
Li, CS, Score test for semiparametric zero-inflated Poisson model, Int J Stat Probab, 1, 1-7 (2012)
Little, RJ; A Regression with missing X's: a review, J Am Stat Assoc, 87, 1227-1237 (1992)
Liu, H.; Powers, DA, Growth curve models for zero-inflated count data: An application to smoking behavior, Struct Equ
Model Multidiscip J, 14, 247-279 (2007)
Lu, SE; Lin, Y.; Shih, WCJ, Analyzing excessive no changes in clinical trials with clustered data, Biometrics, 60, 257-267
(2004) - Zbl 1130.62376
Lukusa, TM; Lee, SM; Li, CS, Semiparametric estimation of a zero-inflated Poisson regression model with missing covariates,
Metrika, 79, 457-483 (2016) - Zbl 1439.62137
Paelke, BT; Preisser, JS; Stearns, SC; Rozier, RG, Multiple imputation of dental caries data using a zero-inflated Poisson
regression model, J Public Health Dent, 71, 71-78 (2011)
Reilly, M.; Pepe, MS, A mean score method for missing and auxiliary covariates data in regression methods, Biometrika, 82,
Cape Town, pp 179-192
Righi, P.; Falorsi, S.; Fasulo, A., Methods for variance estimation under random hot deck imputation in business surveys,
Robins, JM; Rotnitzky, A.; Zhao, LP, Estimation of regression coefficients when some regressors are not always observed,
Rubin, DB, Inference and missing data, Biometrika, 63, 581-592 (1976) - Zbl 0344.62034
Rubin DB (1978) Multiple imputations in sample surveys: a phenomenological Bayesian approach to nonresponse. In: Proceed-
ing of the survey research methods section of the American Statistical Association, vol. 1. American Statistical Association,
Boston, pp 20-28
Rubin, DB, Multiple imputation for nonresponse in surveys (1987), New York: Wiley, New York
Rubin, DB, Multiple imputation after 18+ years, J Am Stat Assoc, 91, 473-489 (1996) - Zbl 0869.62014
Rubin, DB; Schenker, N., Multiple imputation for interval estimation from simple random samples with ignoreable nonresponse,
Samani, EB; Ganjali, M.; Amirian, Y., Zero-inflated power series joint model to analyze count data with missing responses,
Schafer, JL; Graham, JW, Missing data: our view of the state of the art, Psychol Methods, 7, 147-177 (2002)
Singh, S., A note on inflated Poisson distribution, J Indian Stat Assoc, 1, 149-144 (1963)
Wang, S.; Wang, CY, A note on kernel assisted estimators in missing covariate regression, Stat Probab Lett, 55, 439-449
(2001) - Zbl 0994.62034


[51] Wang, CY; Chen, JC; Lee, SM; Ou, ST, Joint conditional likelihood estimator in logistic regression with missing covariate data, Stat Sin, 12, 555-574 (2002) · Zbl 1137.62330


This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.