The Multivariate Poisson-Generalized Inverse Gaussian Claim Count Regression Model with Varying Dispersion and Shape Parameters

George Tzougas* Despoina Makariou†

March 15, 2024

The regression analysis of multivariate count data for capturing the dependence structures between multiple count response variables based on explanatory variables is encountered across several disciplines such as biology, biometrics, genetics, medicine, marketing, ecology, sociology, econometrics, and insurance. In general, multivariate count data models can be classified into the following three classes: multivariate Poisson models, multivariate mixed Poisson (MVMP) models, and copula-based models. For more details, the interested reader can refer to the works of M'Kendrick (1925), Stein and Juritz (1987), Stein et al. (1987), Kocherlakota (1988), Aitchison and Ho (1989), Jung and Winkelmann (1993), Joe (1997), Johnson et al. (1997), Krummenauer (1998), Lakshminarayana et al. (1999), Lee (1999), Munkin and Trivedi (1999), Gurmu and Elder (2000), Chib and Winkelmann (2001), Ho and Singer (2001), Kocherlakota and Kocherlakota (2001), Cameron et al. (2004), Karlis and Meligkotsidou (2005), Zimmer and Trivedi (2006), Genest and Nešlehová (2007), Park and Lord (2007), Ma et al. (2008), Winkelmann (2008), Aguero-Valverde and Jovanis (2009), El-Basyouny and Sayed (2009), Famoye (2010), Nikoloulopoulos and Karlis (2010), Ghitany et al. (2012), Cameron and Trivedi (2013), Nikoloulopoulos (2013), Rüschendorf (2013), Zhan et al. (2015), Marra and Wyszynski (2016), Nikoloulopoulos (2016), Chen and Hanson (2017), Silva et al. (2019), and Chiquet et al. (2020).

In a non-life insurance setting, the actuary may be concerned with modelling jointly different types of claims and their associated counts. In this market segment, there are several circumstances where the interest lies in developing models which can accommodate for positively correlated claims whilst accounting for overdispersion which is a direct consequence of unobserved heterogeneity due to systematic effects in the data. Furthermore, these dependence structures between different claim types may be observed within the same insurance policy, such as property damage and bodily injury claims in motor

*Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh, EH14 4AS. Email: G.Tzougas@hw.ac.uk

†Institute of Insurance Economics, University of St. Gallen, Tannenstrasse 19, CH-9000 St. Gallen, Switzerland. Email: despoina.makariou@unisg.ch

third party liability (MTPL) insurance, or in alternative types of coverage, such as home and auto insurance, bundled together under a single policy. Regarding the latter, some of the advantages for the policyholder are multi-product premium discounts, straightforward tracking of policy renewal dates, easy claims reporting, and a more "personal" relationship between the insured and their insurer where the latter closely identify their needs and mitigate possible insurance coverage gaps. From the insurer's perspective though, bundling multiple types of insurance for the same policyholder translates into a need to develop predictive models which can efficiently capture the joint dynamics of different claims types associated with various insurance business lines. With regards to the use of alternative multivariate count models in non-life insurance, see for instance, Bermúdez and Karlis (2011), Bermúdez and Karlis (2012), Shi and Valdez (2014a), Shi and Valdez (2014b), Abdallah et al. (2016), Bermúdez and Karlis (2017), Bermúdez et al. (2018), Pechon et al. (2018), Pechon et al. (2019), Bolancé and Vernic (2019), Denuit et al. (2019), Fung et al. (2019), Bolancé et al. (2020), Pechon et al. (2021), Jeong and Dey (2021), Gómez-Déniz and Calderín-Ojeda (2021) and Tzougas and di Cerchiara (2021).

In the current study, we develop a multivariate Poisson-Generalised Inverse Gaussian (MVPGIG) regression model with varying dispersion and shape for modelling positively correlated and overdispersed claim counts from different types of coverage in a flexible manner. In particular, within the adopted modelling framework, in addition to the marginal mean parameters, which are traditionally modelled using risk factors, regressors are allowed on the dispersion and shape parameters. The proposed approach allows us to model the skewness and kurtosis of the model explicitly as a function of the explanatory variables for the mean, dispersion and shape parameters. Instead, if only the mean parameter is modelled in terms of explanatory variables then this can lead to a misclassification of policyholders with a high number of claims due to the unobserved heterogeneity changes with covariates.Furthermore, the MVPGIG, is a broad family of models including many MVMP models considered in the aforementioned literature ones as special and/or limiting cases, such as, for example, the multivariate Negative Binomial (MVNB), or multivariate Poisson-Gamma, multivariate Poisson-Inverse Gaussian (MVPIG), multivariate Poisson–Inverse Exponential, multivariate Poisson–Inverse Chi Squared, and multivariate Poisson–Scaled Inverse Chi Squared distributions, depending on the estimated values of the dispersion and shape parameters which are modelled based on covariate information, hence enabling us to account for the tail behaviour of observed data in versatile manner. The latter can be regarded as an important property for capturing overdispersion since this phenomenon is not necessarily attributed to an excess of zeros but it may be also caused by a heavy tail in the claim count data, see Shared (1980). For illustrative purposes, the bivariate Poisson-Generalised Inverse Gaussian (BPGIG) regression model with varying dispersion and shape is fitted on Motor Third Party Liability (MTPL) insurance bodily injury and property damage claim count data using a novel Expectation-Maximization (EM) type algorithm. The proposed maximum likelihood (ML) estimation scheme takes advantage of the stochastic mixture representation of the BPGIG model in order to reduce the problem of maximizing its cumbersome likelihood function which is expressed in terms of the modified Bessel function of the third kind to the simple problem of maximising the likelihood function of its mixing density.

References

Abdallah, A., Boucher, J.-P., and Cossette, H. (2016). Sarmanov family of multivariate distributions for bivariate dynamic claim counts model. *Insurance: Mathematics and Economics*, 68:120–133.

Aguero-Valverde, J. and Jovanis, P. P. (2009). Bayesian multivariate Poisson lognormal models for crash severity modeling and site ranking. *Transportation research record*, 2136(1):82–91.

Aitchison, J. and Ho, C. (1989). The multivariate Poisson-log normal distribution. *Biometrika*, 76(4):643–653.

Bermúdez, L., Guillén, M., and Karlis, D. (2018). Allowing for time and cross dependence assumptions between claim counts in ratemaking models. *Insurance: Mathematics and Economics*, 83:161–169.

Bermúdez, L. and Karlis, D. (2011). Bayesian multivariate Poisson models for insurance ratemaking. *Insurance: Mathematics and Economics*, 48(2):226–236.

Bermúdez, L. and Karlis, D. (2012). A finite mixture of bivariate Poisson regression models with an application to insurance ratemaking. *Computational Statistics & Data Analysis*, 56(12):3988–3999.

Bermúdez, L. and Karlis, D. (2017). A posteriori ratemaking using bivariate poisson models. *Scandinavian Actuarial Journal*, 2017(2):148–158.

Bolancé, C., Guillen, M., and Pitarque, A. (2020). A Sarmanov distribution with beta marginals: An application to motor insurance pricing. *Mathematics*, 8(11).

Bolancé, C. and Vernic, R. (2019). Multivariate count data generalized linear models: Three approaches based on the Sarmanov distribution. *Insurance: Mathematics and Economics*, 85:89–103.

Cameron, A. C., Li, T., Trivedi, P. K., and Zimmer, D. M. (2004). Modelling the differences in counted outcomes using bivariate copula models with application to mismeasured counts. *The Econometrics Journal*, 7(2):566–584.

Cameron, A. C. and Trivedi, P. K. (2013). *Regression analysis of count data*, volume 53. Cambridge university press.

Chen, Y. and Hanson, T. (2017). Copula regression models for discrete and mixed bivariate responses. *Journal of Statistical Theory and Practice*, 11(4):515–530.

Chib, S. and Winkelmann, R. (2001). Markov chain Monte Carlo analysis of correlated count data. *Journal of Business & Economic Statistics*, 19(4):428–435.

Chiquet, J., Mariadassou, M., and Robin, S. (2020). The Poisson-lognormal model as a versatile framework for the joint analysis of species abundances. *bioRxiv*.

Denuit, M., Guillen, M., Trufin, J., et al. (2019). Multivariate credibility modelling for usage-based motor insurance pricing with behavioural data. *Annals of Actuarial Science*, 13(2):378–399.

El-Basyouny, K. and Sayed, T. (2009). Accident prediction models with random corridor parameters. *Accident Analysis & Prevention*, 41(5):1118–1123.

Famoye, F. (2010). On the bivariate negative binomial regression model. *Journal of Applied Statistics*, 37(6):969–981.

Fung, T. C., Badescu, A. L., and Lin, X. S. (2019). A class of mixture of experts models for general insurance: Application to correlated claim frequencies. *ASTIN Bulletin: The Journal of the IAA*, 49(3):647–688.

Genest, C. and Nešlehová, J. (2007). A primer on copulas for count data. *ASTIN Bulletin: The Journal of the IAA*, 37(2):475–515.

Ghitany, M., Karlis, D., Al-Mutairi, D., and Al-Awadhi, F. (2012). An EM algorithm for multivariate mixed poisson regression models and its application. *Applied Mathematical Sciences*, 6(137):6843–6856.

Gómez-Déniz, E. and Calderín-Ojeda, E. (2021). A priori ratemaking selection using multivariate regression models allowing different coverages in auto insurance. *Risks*, 9(7):137.

Gurmu, S. and Elder, J. (2000). Generalized bivariate count data regression models. *Economics Letters*, 68(1):31–36.

Ho, L. L. and Singer, J. d. M. (2001). Generalized least squares methods for bivariate Poisson regression. *Communications in Statistics-Theory and Methods*, 30(2):263–277.

Jeong, H. and Dey, D. K. (2021). Multi-peril frequency credibility premium via shared random effects. *Available at SSRN 3825435*.

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. CRC Press.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1997). *Discrete multivariate distributions*, volume 165. Wiley New York.

Jung, R. C. and Winkelmann, R. (1993). Two aspects of labor mobility: a bivariate Poisson regression approach. *Empirical economics*, 18(3):543–556.

Karlis, D. and Meligkotsidou, L. (2005). Multivariate Poisson regression with covariance structure. *Statistics and Computing*, 15(4):255–265.

Kocherlakota, S. (1988). On the compounded bivariate Poisson distribution: A unified treatment. *Annals of the institute of Statistical Mathematics*, 40(1):61–76.

Kocherlakota, S. and Kocherlakota, K. (2001). Regression in the bivariate Poisson distribution.

Krummenauer, F. (1998). Representation of multivariate discrete distributions by probability generating functions. *Statistics & probability letters*, 39(4):327–331.

Lakshminarayana, J., Pandit, S. N., and Srinivasa Rao, K. (1999). On a bivariate Poisson distribution. *Communications in Statistics-Theory and Methods*, 28(2):267–276.

Lee, A. (1999). Applications: Modelling rugby league data via bivariate negative binomial regression. *Australian & New Zealand Journal of Statistics*, 41(2):141–152.

Ma, J., Kockelman, K. M., and Damien, P. (2008). A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using bayesian methods. *Accident Analysis & Prevention*, 40(3):964–975.

Marra, G. and Wyszynski, K. (2016). Semi-parametric copula sample selection models for count responses. *Computational Statistics & Data Analysis*, 104:110–129.

M'Kendrick, A. (1925). Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society*, 44:98–130.

Munkin, M. K. and Trivedi, P. K. (1999). Simulated maximum likelihood estimation of multivariate mixed-poisson regression models, with application. *The Econometrics Journal*, 2(1):29–48.

Nikoloulopoulos, A. K. (2013). Copula-based models for multivariate discrete response data. In *Copulae in Mathematical and Quantitative Finance*, pages 231–249. Springer.

Nikoloulopoulos, A. K. (2016). Efficient estimation of high-dimensional multivariate normal copula models with discrete spatial responses. *Stochastic environmental research and risk assessment*, 30(2):493–505.

Nikoloulopoulos, A. K. and Karlis, D. (2010). Regression in a copula model for bivariate count data. *Journal of Applied Statistics*, 37(9):1555–1568.

Park, E. S. and Lord, D. (2007). Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity. *Transportation Research Record*, 2019(1):1–6.

Pechon, F., Denuit, M., and Trufin, J. (2019). Multivariate modelling of multiple guarantees in motor insurance of a household. *European Actuarial Journal*, 9(2):575–602.

Pechon, F., Denuit, M., and Trufin, J. (2021). Home and motor insurance joined at a household level using multivariate credibility. *Annals of Actuarial Science*, 15(1):82–114.

Pechon, F., Trufin, J., Denuit, M., et al. (2018). Multivariate modelling of household claim frequencies in motor third-party liability insurance. *Astin Bulletin*, 48(3):969–993.

Rüschendorf, L. (2013). Copulas, Sklar's theorem, and distributional transform. In *Mathematical risk analysis*, pages 3–34. Springer.

Shared, M. (1980). On mixtures from exponential families. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(2):192–198.

Shi, P. and Valdez, E. A. (2014a). Longitudinal modeling of insurance claim counts using jitters. *Scandinavian Actuarial Journal*, 2014(2):159–179.

Shi, P. and Valdez, E. A. (2014b). Multivariate negative binomial models for insurance claim counts. *Insurance: Mathematics and Economics*, 55:18–29.

Silva, A., Rothstein, S. J., McNicholas, P. D., and Subedi, S. (2019). A multivariate Poisson-log normal mixture model for clustering transcriptome sequencing data. *BMC bioinformatics*, 20(1):1–11.

Stein, G. Z. and Juritz, J. M. (1987). Bivariate compound Poisson distributions. *Communications in Statistics-Theory and Methods*, 16(12):3591–3607.

Stein, G. Z., Zucchini, W., and Juritz, J. M. (1987). Parameter estimation for the Sichel distribution and its multivariate extension. *Journal of the American Statistical Association*, 82(399):938–944.

Tzougas, G. and di Cerchiara, A. P. (2021). Bivariate mixed poisson regression models with varying dispersion. *North American Actuarial Journal*, pages 1–31.

Winkelmann, R. (2008). *Econometric analysis of count data*. Springer Science & Business Media.

Zhan, X., Aziz, H. A., and Ukkusuri, S. V. (2015). An efficient parallel sampling technique for multivariate Poisson-lognormal model: Analysis with two crash count datasets. *Analytic methods in accident research*, 8:45–60.

Zimmer, D. M. and Trivedi, P. K. (2006). Using trivariate copulas to model sample selection and treatment effects: application to family health care demand. *Journal of Business & Economic Statistics*, 24(1):63–76.