Statistica Sinica **37** (2027), 1-21
doi:https://doi.org/10.5705/ss.202024.0109

JOINT MEAN AND CORRELATION REGRESSION MODELS FOR MULTIVARIATE DATA

Zhi Yang Tho*, Francis K. C. Hui, and Tao Zou*

The Australian National University

Abstract: We propose a joint mean and correlation regression model for multivariate discrete and (semi-)continuous response data, that simultaneously regresses the mean of each response against a set of covariates, and the correlations between responses against a set of similarity/distance measures. A set of joint estimating equations are formulated to construct an estimator of both the mean regression coefficients and the correlation regression parameters. Under a general setting where the number of responses can tend to infinity, the joint estimator is demonstrated to be consistent and asymptotically normally distributed, with differing rates of convergence due to the mean regression coefficients being heterogeneous across responses. An iterative estimation procedure is developed to obtain parameter estimates in the required (constrained) parameter space. Simulations demonstrate the strong finite sample performance of the proposed estimator in terms of point estimation and inference. We apply the proposed model to a count dataset of 38 Carabidae ground beetle species sampled throughout Scotland, along with information about the environmental conditions of each site and the traits of each species. Results show the relationship between mean abundance and environmental covariates differs across the beetle species, and that beetle total length is important in driving the correlations between species.

Key words and phrases: Covariance regression, generalized estimating equation, joint mean-covariance modeling, multivariate discrete data.

1. Introduction

The analysis of correlated multivariate or multi-response data is becoming increasingly important nowadays, as it provides enhanced opportunities to answer more diverse and deeper scientific questions relative to studying a univariate response. A prime example is the ability to simultaneously study how responses vary together as a function of covariates, as well as how correlations between responses are related to similarity (or distance) measures of predictors. For instance, in ecology there is a growing interest in jointly modeling how species' distributions are associated with environmental covariates along with how the covariation between species varies with phylogenetic and trait distances (Warton et al., 2015; Tikhonov et al., 2017; Ovaskainen and Abrego, 2020).

*Corresponding author. E-mail: (Tho) zhiyang.tho@anu.edu.au; (Zou) tao.zou@anu.edu.au