

<b>Statistica Sinica Preprint No: SS-2024-0109</b>	
<b>Title</b>	Joint Mean and Correlation Regression Models for Multivariate Data
<b>Manuscript ID</b>	SS-2024-0109
<b>URL</b>	<a href="http://www.stat.sinica.edu.tw/statistica/">http://www.stat.sinica.edu.tw/statistica/</a>
<b>DOI</b>	10.5705/ss.202024.0109
<b>Complete List of Authors</b>	Zhi Yang Tho, Francis K. C. Hui and Tao Zou
<b>Corresponding Authors</b>	Zhi Yang Tho
<b>E-mails</b>	zhiyang.tho@anu.edu.au

---

## Joint Mean and Correlation Regression Models for Multivariate Data

Zhi Yang Tho<sup>\*</sup>, Francis K. C. Hui, and Tao Zou<sup>\*</sup>

*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,

---

<sup>\*</sup>Corresponding author.

---

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).

To study how the means of a set of responses vary as a function of covariates, a popular approach in the statistical literature is generalized estimating equations (GEEs, Liang and Zeger, 1986) or variations thereof, where a mean model for each response is coupled with a working, between-response correlation matrix. The latter can be specified in different ways, leading to varying degrees of im-

---

proved statistical efficiency for inference on the mean regression coefficients (e.g., Qu et al., 2000; Ye and Pan, 2006; Warton, 2011). GEEs focus primarily on modeling and estimating the mean component; they are not designed to answer scientific questions relating to the correlation between responses themselves. On the other hand, various methods have been developed that explicitly link the covariance matrix of the response vector, or functions thereof, to a linear combination of known symmetric matrices (e.g., Anderson, 1973; Chiu et al., 1996; Zwiernik et al., 2017). Of particular note are the studies of Pourahmadi (1999), Zhang and Leng (2012), Zhang et al. (2015) and Bonat and Jørgensen (2016), who integrated mean modeling with multivariate covariance modeling using covariate information under specific model structures. These studies focused primarily on longitudinal or spatio-temporal data, so they considered common structures e.g., compound symmetry or neighborhood structure in Bonat and Jørgensen (2016), to account for known longitudinal or spatio-temporal associations in the data. More recently, Hu et al. (2024) considered to regress the elements of the generalized z-transformed correlation matrices of general correlated data on covariates that are formed by taking the difference of the predictors associated with each pair of responses, while Tang et al. (2019) studied a joint mean-correlation regression for discrete longitudinal data by modeling the correlation matrix in a Gaussian copula via a hyperspherical reparameterization.

---

In this article, we propose a new joint mean and correlation regression model that simultaneously quantifies the relationship between the mean of each response and a set of covariates, and the relationship between the response correlation matrix and a set of similarity measures of predictors. Analogous to classical GEEs, the proposed joint model requires specifying only the first two moments of each response along with the correlation matrix between responses. The first moment of each response is regressed against a set of covariates with the help of link functions, and using mean regression coefficients that are heterogeneous across responses i.e., each response has its own set of regression coefficients. The specification of the second moment uses variance functions to capture potential mean-variance relationships within each response. Turning to the between-response correlation matrix, we propose to regress this against a set of similarity matrices that are either observed directly as part of the data collection process, or induced from available predictor information associated with each response. As one example of such similarity measures, in text frequency analysis where the responses are frequencies of different words across documents, similarity matrices can be formed from various characteristics of the words such as their topical meanings (e.g., Zhu and Xing, 2011); see also Section 5 for similarity matrices formed from species traits in ecology. Unlike the aforementioned works on joint mean-covariance modeling that focused on longitudinal or

---

spatio-temporal data, we aim to provide an explicit, data-driven quantification of how the unknown dependence structure for general correlated data is informed by these similarity measures of auxiliary predictor information, through a set of correlation regression parameters that are common across responses. Furthermore, unlike the studies of Tang et al. (2019) and Hu et al. (2024) that considered hyperspherical reparameterization and generalized z-transformation of the correlation matrices, respectively, we directly model the correlation matrix, noting that the covariates constructed by Hu et al. (2024) are analogous to our idea of similarity measures.

The idea of correlation regression builds upon recent developments of covariance regression in Zou et al. (2017, 2020, 2022), although we make three clear advances on these works. First, we extend covariance regression to correlation regression, while allowing for heterogeneous variances across responses. This differs from Zou et al. (2017, 2020, 2022) who all assumed a homogeneous variance across responses. Note also that we model the correlation rather than the covariance matrix, as our method is developed within the GEE framework which requires the specification of a working correlation matrix. Second, our proposed joint mean and correlation regression model can handle a much wider variety of discrete and (semi-)continuous responses e.g., overdispersed counts, binomial, and non-negative continuous responses that arise in ecology and quan-

---

titative genetics among other fields. Third, we develop and study asymptotic inference for both the heterogeneous mean regression coefficients and correlation regression parameters, in contrast to Zou et al. (2017, 2020, 2022) who focused solely on the covariance regression parameters.

In the spirit of GEEs, we formulate a set of joint estimating equations for estimating the heterogeneous mean regression coefficients and correlation regression parameters. Note unlike the classical GEE literature which often treats the working correlation matrix either as known or effectively as a nuisance parameter (e.g., Liang and Zeger, 1986; Qu et al., 2000; Wang and Carey, 2003), here we consider both mean and correlation components to be equally important in the modeling process. That is, we are interested in performing inference on both the mean regression coefficients and correlation regression parameters, and this requires an asymptotic theory for the joint set of parameters.

Under a setting where the number of responses can diverge with an increasing number of clusters, we establish estimation consistency and asymptotic normality of the mean regression coefficient estimators and correlation regression parameter estimators, showing that they exhibit differing rates of convergence. We note that the total number of mean regression coefficients can still tend to infinity even if the number of covariates is fixed, since the mean regression coefficients are heterogeneous across responses. Hence, the involvement of both

---

the multivariate mean and correlation regression substantially increases the difficulty in deriving the theoretical properties of the proposed estimators. Indeed, to our knowledge, this article is the first to formally demonstrate consistency and asymptotic normality for the estimator of a joint model specifying heterogeneous mean regression coefficients, where the number of responses is allowed to diverge.

We develop an estimation procedure that iterates between the aforementioned estimating equations to update the mean regression coefficients (along with any dispersion parameters) and correlation regression parameters. For the latter, estimation must be done in a way to obtain an overall valid correlation matrix i.e., positive definite with ones on the diagonals and off-diagonals between -1 and 1. To overcome this challenge, we propose a novel algorithm that adapts the positive definiteness constrained algorithm of Zou et al. (2017) to attain a positive definite covariance matrix estimate under the GEE framework first, before standardizing it to produce valid correlation regression parameter estimates.

Simulation studies across a number of response types demonstrate our proposed approach provides similar performance to several existing GEE-type methods for estimating the mean regression coefficients, but outperforms them in recovering the between-response correlation matrix. We also investigate the impact of misspecified similarity measures on the proposed estimator, with results



---

demonstrating the robustness of the mean regression coefficient estimators. The proposed model is applied to a multivariate abundance dataset in ecology comprising overdispersed counts of 38 Carabidae ground beetle species sampled throughout Scotland, along with information about the environmental covariates of each site and the traits of each species. Results show the beetle species exhibit quite differing relationships with key environmental indicators such as soil pH and land management, while species traits such as beetle total length and breeding season have important effects in driving the correlations between the species. These findings are not particularly sensitive to different specifications of trait similarity measures.

The rest of this article is organized as follows. Section 2 introduces the joint mean and correlation regression model, a set of associated estimating equations, and asymptotic results for the resulting joint estimator. Section 3 presents details of the iterative estimation procedure, while Sections 4 and 5 discuss results of the simulation study and application to the ground beetle dataset, respectively. Section 6 offers some concluding remarks. All technical conditions, proofs, and additional simulation results are presented in the Appendix and supplementary material.

## 2. A Joint Mean and Correlation Regression Model

Let  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ip})^\top$  be a  $p$ -dimensional response vector collected from the  $i$ -th cluster for  $i = 1, \dots, n$ . In this article, we follow Pourahmadi (1999), Warton (2011) and Müller et al. (2013) among others and use  $i$  to index the  $i$ -th “cluster”, noting related works adopt other terminologies such as “subject” in Liang and Zeger (1986) and Qu et al. (2000), “observation” in Zou et al. (2021), and “observational unit” in Hui et al. (2023) to refer to cluster. Note also that we focus our developments on the balanced data setting i.e., all  $n$  clusters have the same number of responses  $p$ ; see Section 6 for a discussion on extensions to the case of unequal number of responses  $p_i$  for each of the  $i$ -th cluster. Let  $\mathbf{x}_i = (x_{i1}, \dots, x_{id})^\top$  denote a set of covariates associated with the  $i$ -th cluster that we will relate to the mean of the responses, where  $x_{i1} = 1$  corresponds to the intercept term for  $i = 1, \dots, n$ , and  $\{\mathbf{W}_k = (w_{j_1 j_2}^{(k)})_{p \times p} : k = 1, \dots, K\}$  denote a set of  $K$  similarity matrices of dimension  $p \times p$  that we will link to the correlation matrix of the responses. The matrices  $\mathbf{W}_k$  may be available directly as part of the data collection process itself, or constructed from auxiliary information vectors  $\mathbf{z}_j = (z_{j1}, \dots, z_{jK})^\top$  associated with the  $j$ -th response for  $j = 1, \dots, p$ . For the latter, each element  $w_{j_1 j_2}^{(k)}$  in  $\mathbf{W}_k$  measures the similarity between  $z_{j_1 k}$  and  $z_{j_2 k}$  for  $j_1 \neq j_2$ . For instance, if  $z_{jk}$  is quantitative, then the similarity  $w_{j_1 j_2}^{(k)}$  can be defined as  $w_{j_1 j_2}^{(k)} = \exp(-|z_{j_1 k} - z_{j_2 k}|^2)$ , whereas if  $z_{jk}$  is

qualitative we can set  $w_{j_1 j_2}^{(k)} = 1$  if  $z_{j_1 k}$  and  $z_{j_2 k}$  have the same categorical level, and  $w_{j_1 j_2}^{(k)} = 0$  otherwise (see also Johnson and Wichern, 2013). For reasons of parameter identifiability, as will be illustrated later on, the diagonals  $w_{jj}^{(k)}$  are set to zeros for all  $j = 1, \dots, p$  and  $k = 1, \dots, K$ .

The proposed joint mean and correlation regression model is formulated as follows. First, we assume the mean of each response, denoted here as  $E(Y_{ij}) = \mu_{ij}(\beta_j)$ , is related to the covariates as given by

$$g\{\mu_{ij}(\beta_j)\} = \mathbf{x}_i^\top \beta_j, \text{ for } i = 1, \dots, n \text{ and } j = 1, \dots, p, \quad (2.1)$$

where  $\beta_j = (\beta_{j1}, \dots, \beta_{jd})^\top$  is a vector of mean regression coefficients that are heterogeneous across responses, and  $g(\cdot)$  is a known link function. Next, we specify the second moment of  $Y_{ij}$  as  $\text{Var}(Y_{ij}) = h\{\mu_{ij}(\beta_j); \phi_j\}$  for  $i = 1, \dots, n$  and  $j = 1, \dots, p$ , where  $\phi_j > 0$  are dispersion parameters that are also heterogeneous across  $j = 1, \dots, p$ , and  $h(\cdot)$  is a known function characterizing the mean-variance relationship of the responses. Common examples of link and variance functions include the logit link  $g(\mu) = \log\{\mu/(1-\mu)\}$  and the variance function  $h(\mu; \phi) = \phi\mu(1-\mu)$  for binary responses, the log link  $g(\mu) = \log(\mu)$  and the variance function  $h(\mu; \phi) = \phi\mu$  (or  $h(\mu; \phi) = \mu + \phi\mu^2$ ) for (overdispersed) counts, and the identity link  $g(\mu) = \mu$  coupled with the constant func-

tion  $h(\mu; \phi) = \phi$  for continuous responses (Fitzmaurice et al., 2011). Turning to the correlation regression component of the model, let  $\text{Corr}(Y_{ij_1}, Y_{ij_2}) = r_{j_1 j_2}(\boldsymbol{\rho})$  for  $j_1, j_2 = 1, \dots, p$ , and subsequently define  $\mathbf{R}(\boldsymbol{\rho}) = (r_{j_1 j_2}(\boldsymbol{\rho}))_{p \times p}$  as the  $p \times p$  (working) correlation matrix of  $\mathbf{Y}_i$  for  $i = 1, \dots, n$ . Building upon Zou et al. (2017, 2020, 2022), we model this correlation matrix as

$$\mathbf{R}(\boldsymbol{\rho}) = \mathbf{I}_p + \sum_{k=1}^K \rho_k \mathbf{W}_k, \quad (2.2)$$

where  $\mathbf{I}_p$  is the  $p$ -dimensional identity matrix, and  $\boldsymbol{\rho} = (\rho_1, \dots, \rho_K)^\top$  is a vector of correlation regression parameters that possess a simple, direct interpretation as quantifying the impact of the similarities on the correlation between responses. For example, in ecology where similarity matrices are constructed from species traits, a higher, positive value of  $\rho_k$  implies that, conditional on other traits, two species with more similar values in their  $k$ -th trait variable are expected to have a stronger positive correlation after accounting for differences in their mean response (which may suggest this trait is important in mediating biotic interactions between species; see the application in Section 5).

Equation (2.2) offers a parsimonious yet explicit way to model correlations between responses, as it only involves estimating a vector of  $K$  correlation regression parameters,  $\boldsymbol{\rho}$ . This formulation also includes various structured cor-

relation matrices (not driven by the data itself) as special cases, including traditional autoregressive, moving average, compound symmetry, and banded structures (see Zou et al., 2017). By denoting  $\mathbf{W}_0 = \mathbf{I}_p$  and  $\rho_0 = 1$ , the correlation regression model can be written as  $\mathbf{R}(\boldsymbol{\rho}) = \sum_{k=0}^K \rho_k \mathbf{W}_k$ . As reviewed in Section 1, similar ideas of expressing the covariance or correlation matrix as a function of a linear combination of matrices have been considered previously in the literature. Importantly, we consider  $\boldsymbol{\rho}$  as being equal in importance to  $\boldsymbol{\beta}$ . This is in contrast to previous studies and variations of GEEs, where the (parameters characterizing the) working correlation have been either treated as a nuisance or used largely to improve the efficiency of inference on the mean model. Therefore, it is imperative to develop asymptotic theory for the joint estimator of all the regression parameters  $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\rho}^\top)^\top$ , so as to provide a basis for simultaneous inference on the mean and correlation components of the proposed model.

Note we model  $\mathbf{R}(\boldsymbol{\rho})$  as a correlation instead of a covariance matrix: this is consistent with the requirement of a working correlation matrix in the GEE framework where the response variances are already modeled by the variance function, and also circumvents parameter identifiability problems between the covariance parameters and dispersion parameters when the variance function has a multiplicative relationship with the dispersion parameter e.g., if  $h(\mu; \phi) = \phi\mu$  and  $\mathbf{R}(\boldsymbol{\rho}) = \rho_0 \mathbf{I}_p + \sum_{k=1}^K \rho_k \mathbf{W}_k$  is modeled as a covariance matrix, then the

## 2.1 Estimating Equations

covariance parameters  $c\rho_0, \dots, c\rho_K$  and dispersion parameters  $\phi_1/c, \dots, \phi_p/c$  would result in the same covariance structure of  $\mathbf{Y}_i$  vectors for any value of  $c > 0$ . On the other hand, the estimation of  $\boldsymbol{\rho}$  is now more challenging compared to previous covariance regression models of Zou et al. (2017, 2020, 2022). Specifically, since equation (2.2) is a regression model for the correlation matrix,  $\boldsymbol{\rho}$  must be in the parameter space  $\mathcal{P}^+ = \{\boldsymbol{\rho} : \mathbf{R}(\boldsymbol{\rho}) \text{ is a valid correlation matrix}\}$ . That is, the diagonals of  $\mathbf{R}(\boldsymbol{\rho})$  need to be ones and off-diagonals between -1 and 1, in addition to  $\mathbf{R}(\boldsymbol{\rho})$  being positive definite. We provide two simple but insightful examples to illustrate this, in the case of a single similarity matrix  $\mathbf{W}_1$ . First, if  $\mathbf{W}_1$  is a compound symmetry matrix i.e.,  $w_{j_1 j_2}^{(1)} = c > 0$  for  $j_1 \neq j_2$ , then  $\mathcal{P}^+ = \{\rho_1 : \rho_1 \in (-(p-1)^{-1}c^{-1}, c^{-1})\}$ . Second, if  $\mathbf{W}_1$  is a tridiagonal matrix with  $w_{j_1 j_2}^{(1)} = c \neq 0$  for  $|j_1 - j_2| = 1$  and zero otherwise, then the parameter space is  $\mathcal{P}^+ = \{\rho_1 : \rho_1 \in (0.5|c|^{-1}[\cos\{p\pi/(p+1)\}]^{-1}, -0.5|c|^{-1}[\cos\{p\pi/(p+1)\}]^{-1})\}$ . Needless to say, when multiple general similarity matrices are involved the requirement for  $\boldsymbol{\rho} \in \mathcal{P}^+$  becomes even more complex; this motivates the estimating equations proposed in the following subsection, which will take this requirement into consideration.

## 2.1 Estimating Equations

We establish a set of estimating equations that will be used as the basis for estimation and inference with the proposed joint model. To this end, we write

## 2.1 Estimating Equations

the model in a vectorized form as follows. Let  $\mathbf{Y} = (\mathbf{Y}_1^\top, \dots, \mathbf{Y}_n^\top)^\top$  and  $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \dots, \boldsymbol{\beta}_p^\top)^\top$  denote the stacked  $np$ - and  $pd$ -dimensional vectors of all responses and mean regression coefficients, respectively. Subsequently, we can write  $\boldsymbol{\mu}_i(\boldsymbol{\beta}) = (\mu_{i1}(\boldsymbol{\beta}_1), \dots, \mu_{ip}(\boldsymbol{\beta}_p))^\top$  for  $i = 1, \dots, n$  and  $\boldsymbol{\mu}(\boldsymbol{\beta}) = (\boldsymbol{\mu}_1^\top(\boldsymbol{\beta}), \dots, \boldsymbol{\mu}_n^\top(\boldsymbol{\beta}))^\top$ . Next, let  $\mathbf{A}_i(\boldsymbol{\beta}) = \text{diag}[h\{\mu_{i1}(\boldsymbol{\beta}_1); \phi_1\}, \dots, h\{\mu_{ip}(\boldsymbol{\beta}_p); \phi_p\}]$  denote  $p \times p$  diagonal matrices of the variance functions at the  $i$ -th cluster for  $i = 1, \dots, n$ , where for ease of notation we have suppressed the dependence on the  $\phi_j$ 's, and let  $\mathbf{A}(\boldsymbol{\beta}) = \text{diag}\{\mathbf{A}_1(\boldsymbol{\beta}), \dots, \mathbf{A}_n(\boldsymbol{\beta})\}$  be a block diagonal matrix with the  $i$ -th block being  $\mathbf{A}_i(\boldsymbol{\beta})$ . By denoting  $\text{Cov}(\mathbf{Y})$  as the full  $np \times np$  covariance matrix of the vector  $\mathbf{Y}$ , equations (2.1) – (2.2) can be expressed as

$$E(\mathbf{Y}) = \boldsymbol{\mu}(\boldsymbol{\beta}), \quad \text{Cov}(\mathbf{Y}) = \mathbf{A}^{1/2}(\boldsymbol{\beta}) \{ \mathbf{I}_n \otimes \mathbf{R}(\boldsymbol{\rho}) \} \mathbf{A}^{1/2}(\boldsymbol{\beta}), \quad (2.3)$$

where  $\otimes$  is the Kronecker product operator and the form of the covariance in equation (2.3) is analogous to that seen in GEEs previously (e.g., Liang and Zeger, 1986; Warton, 2011).

Next, consider the matrix-valued function  $\boldsymbol{\Sigma}(\boldsymbol{\alpha}) = \alpha_0 \mathbf{I}_p + \sum_{k=1}^K \alpha_k \mathbf{W}_k$ , where  $\boldsymbol{\alpha} = (\alpha_0, \dots, \alpha_K)^\top \in \mathcal{A}^+$  lies in the positive definite parameter space  $\mathcal{A}^+ = \{\boldsymbol{\alpha} : \boldsymbol{\Sigma}(\boldsymbol{\alpha}) \text{ is positive definite}\}$ . Recall the diagonals of  $\mathbf{W}_k$  are zeros for  $k = 1, \dots, K$ . Then we have the following proposition.

## 2.1 Estimating Equations

**Proposition 1.** *If there exists an  $\alpha \in \mathcal{A}^+$ , then  $(\alpha_1/\alpha_0, \dots, \alpha_K/\alpha_0)^\top \in \mathcal{P}^+$ .*

The above result follows directly from the definition of a correlation matrix.

Importantly, it implies if we are able to obtain values  $\alpha \in \mathcal{A}^+$ , then a valid (estimated) correlation matrix  $\mathbf{R}(\rho)$  can be immediately obtained by setting  $\rho_1 = \alpha_1/\alpha_0, \dots, \rho_K = \alpha_K/\alpha_0$ , with the resulting values of  $\rho$  satisfying  $\rho \in \mathcal{P}^+$ .

Motivated by this, we propose to consider a slight reparameterization of the correlation regression model in equation (2.2) as  $\Sigma(\alpha) = \alpha_0 \mathbf{I}_p + \sum_{k=1}^K \alpha_k \mathbf{W}_k = \mathbf{R}(\rho)$ , with  $\alpha_0 = 1$  and  $\alpha_k = \rho_k$  for  $k = 1, \dots, K$ . The covariance equation in (2.3) can then be rewritten as  $\text{Cov}(\mathbf{Y}) = \mathbf{A}^{1/2}(\beta) \{ \mathbf{I}_n \otimes \Sigma(\alpha) \} \mathbf{A}^{1/2}(\beta)$ , and subsequently the parameters we solve for are now  $\vartheta = (\beta^\top, \alpha^\top)^\top$ .

Let  $\mathbf{D}(\beta) = \partial \mu(\beta) / \partial \beta^\top \in \mathbb{R}^{np \times pd}$  and  $\tilde{\Sigma}(\alpha) = \mathbf{I}_n \otimes \Sigma(\alpha)$ . For estimating the mean regression coefficients  $\beta$ , we consider the estimating equation

$$\psi_\beta(\beta, \alpha) = \mathbf{D}^\top(\beta) \mathbf{A}^{-1/2}(\beta) \tilde{\Sigma}^{-1}(\alpha) \mathbf{A}^{-1/2}(\beta) \{ \mathbf{Y} - \mu(\beta) \} = \mathbf{0}_{pd}, \quad (2.4)$$

where  $\mathbf{0}_{pd}$  is a  $pd$ -dimensional vector of zeros. Equation (2.4) for the mean regression coefficients is the same as in GEEs, meaning we have  $E\{\psi_\beta(\beta, \alpha)\} = \mathbf{0}_{pd}$  when the estimating equation is evaluated at the true values of the mean regression coefficients regardless of the specification of  $\tilde{\Sigma}^{-1}(\alpha)$ . Next, let  $\epsilon_i(\beta) = \mathbf{A}_i^{-1/2}(\beta) \{ \mathbf{Y}_i - \mu_i(\beta) \}$  for  $i = 1, \dots, n$ ,  $\epsilon(\beta) = (\epsilon_1^\top(\beta), \dots, \epsilon_n^\top(\beta))^\top$ , and



## 2.1 Estimating Equations

$\tilde{\mathbf{W}}_k = \mathbf{I}_n \otimes \mathbf{W}_k$  for  $k = 0, \dots, K$ . Then for the reparameterized correlation regression parameters  $\alpha$ , we consider solving the estimating equation

$$\psi_{\alpha}(\beta, \alpha) = \left( \epsilon^{\top}(\beta) \tilde{\mathbf{W}}_k \epsilon(\beta) \right)_{(K+1) \times 1} - \left( \text{tr}(\tilde{\mathbf{W}}_{k_1} \tilde{\mathbf{W}}_{k_2}) \right)_{(K+1) \times (K+1)} \alpha = \mathbf{0}_{K+1}, \quad (2.5)$$

where  $(\text{tr}(\tilde{\mathbf{W}}_{k_1} \tilde{\mathbf{W}}_{k_2}))_{(K+1) \times (K+1)}$  denotes a  $(K+1) \times (K+1)$  matrix whose  $(k_1+1, k_2+1)$ -th element is given by  $\text{tr}(\tilde{\mathbf{W}}_{k_1} \tilde{\mathbf{W}}_{k_2})$  for  $k_1, k_2 = 0, \dots, K$ , and  $(\epsilon^{\top}(\beta) \tilde{\mathbf{W}}_k \epsilon(\beta))_{(K+1) \times 1}$  is a  $(K+1)$ -dimensional vector whose  $(k+1)$ -th element is  $\epsilon^{\top}(\beta) \tilde{\mathbf{W}}_k \epsilon(\beta)$  for  $k = 0, \dots, K$ . Equation (2.5) is related to the first order condition of the least squares optimization problem considered by Zou et al. (2017, 2020), who estimated covariance regression parameters by minimizing the least squares loss function  $Q(\beta, \alpha) = \sum_{i=1}^n \|\epsilon_i(\beta) \epsilon_i^{\top}(\beta) - \Sigma(\alpha)\|_F^2$  with respect to  $\alpha$  when  $\beta$  is given, where  $\|\mathbf{H}\|_F = \{\text{tr}(\mathbf{H}^{\top} \mathbf{H})\}^{1/2}$  is the Frobenius norm for a generic matrix  $\mathbf{H}$ . Specifically, the solution of equation (2.5) satisfies the first order condition of the optimization problem  $\min_{\alpha} Q(\beta, \alpha)$  since  $-2\psi_{\alpha}(\beta, \alpha) = \partial Q(\beta, \alpha) / \partial \alpha$ . It can also be verified that  $E\{\psi_{\alpha}(\beta, \alpha)\} = \mathbf{0}_{K+1}$  where  $\psi_{\alpha}$  is evaluated at the true values of the mean regression coefficients and the reparameterized correlation regression parameters. Other forms of estimating equation could be considered for the reparameterized correlation regression parameters e.g. see Section 3.3.1 of Lipsitz and Fitzmaurice (2008),

## 2.2 Asymptotic Theory

although we focus our developments on equation (2.5) in this article.

By solving the joint estimating equation  $\psi(\boldsymbol{\vartheta}) = (\psi_{\beta}^{\top}(\boldsymbol{\beta}, \boldsymbol{\alpha}), \psi_{\alpha}^{\top}(\boldsymbol{\beta}, \boldsymbol{\alpha}))^{\top} = \mathbf{0}_{pd+K+1}$ , we obtain the estimator  $\hat{\boldsymbol{\vartheta}} = (\hat{\boldsymbol{\beta}}^{\top}, \hat{\boldsymbol{\alpha}}^{\top})^{\top}$ . Afterward, following Proposition 1 we obtain the joint estimator of the mean regression coefficients and the correlation regression parameters as  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^{\top}, \hat{\boldsymbol{\rho}}^{\top})^{\top}$ , where  $\hat{\rho}_1 = \hat{\alpha}_1/\hat{\alpha}_0, \dots$ ,  $\hat{\rho}_K = \hat{\alpha}_K/\hat{\alpha}_0$ .

## 2.2 Asymptotic Theory

We first study the asymptotic properties of  $\hat{\boldsymbol{\vartheta}} = (\hat{\boldsymbol{\beta}}^{\top}, \hat{\boldsymbol{\alpha}}^{\top})^{\top}$  as the number of clusters  $n \rightarrow \infty$  and the number of responses  $p$  can grow with increasing  $n$ , with the dispersion parameters assumed to be known. We emphasize that although both  $d$  and  $K$  are assumed to be fixed,  $\hat{\boldsymbol{\vartheta}}$  has a growing dimension as  $p \rightarrow \infty$  since the dimension of  $\boldsymbol{\beta}$  grows with order  $p$ . To facilitate developments of the asymptotic theory, we introduce a matrix  $\Xi^{(S)} = \text{diag}(\mathbf{T}^{(S)} \otimes \mathbf{I}_d, \mathbf{I}_{K+1})$ , where  $S = \{s_1, \dots, s_q\}$  generically denotes a subset of  $\{1, \dots, p\}$  with finite size  $q$ ,  $\mathbf{T}^{(S)} = (\mathbf{t}_{s_1}, \dots, \mathbf{t}_{s_q})^{\top}$  with  $\mathbf{t}_{s_j}$  being the  $s_j$ -th column of the identity matrix  $\mathbf{I}_p$  for  $j = 1, \dots, q$ , and  $\mathbf{0}_{k_1 \times k_2}$  denotes a  $k_1 \times k_2$  matrix of zeros. It follows that  $\Xi^{(S)}\hat{\boldsymbol{\vartheta}} = (\hat{\boldsymbol{\beta}}_{s_1}^{\top}, \dots, \hat{\boldsymbol{\beta}}_{s_q}^{\top}, \hat{\boldsymbol{\alpha}}^{\top})^{\top}$  is a finite dimensional sub-vector of  $\hat{\boldsymbol{\vartheta}}$ .

Let  $\boldsymbol{\vartheta}^{(0)} = (\boldsymbol{\beta}^{(0)\top}, \boldsymbol{\alpha}^{(0)\top})^{\top}$  denote the true parameter value for  $\boldsymbol{\vartheta}$ , where  $\Sigma(\boldsymbol{\alpha}^{(0)})$  is positive definite. In the following theorem, the estimators  $(\hat{\boldsymbol{\beta}}_{s_1}^{\top}, \dots$ ,

## 2.2 Asymptotic Theory

$\hat{\beta}_{s_q}^\top)^\top$  and  $\hat{\alpha}$  have differing rates of convergence towards their respective true values. To accommodate this, we define  $\mathbf{Z} = \text{diag}(\sqrt{n}\mathbf{I}_{qd}, \sqrt{np}\mathbf{I}_{K+1})$  and proceed to derive the asymptotic distribution of the quantity  $\mathbf{Z}\Xi^{(S)}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}^{(0)})$ . Let  $\tilde{\mathbf{Z}} = \text{diag}(\sqrt{n}\mathbf{I}_{pd}, \sqrt{np}\mathbf{I}_{K+1})$  and define the matrices  $\mathbf{B} = \tilde{\mathbf{Z}}^{-1}E\{\partial\boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})/\partial\boldsymbol{\vartheta}^\top\}\tilde{\mathbf{Z}}^{-1}$ ,  $\mathbf{U} = \tilde{\mathbf{Z}}^{-1}\text{Cov}\{\boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})\}\tilde{\mathbf{Z}}^{-1}$  and  $\boldsymbol{\Omega}(\Xi^{(S)}) = \Xi^{(S)}\mathbf{B}^{-1}\mathbf{U}\mathbf{B}^{-1\top}\Xi^{(S)\top}$ , where closed-form expressions for both  $E\{\partial\boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})/\partial\boldsymbol{\vartheta}^\top\}$  and  $\text{Cov}\{\boldsymbol{\psi}(\boldsymbol{\vartheta}^{(0)})\}$  are provided in Lemma S1 of supplementary material S4. We then have the following result for the estimator  $\hat{\boldsymbol{\vartheta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\alpha}}^\top)^\top$ .

**Theorem 1.** *Under Conditions 1–5 in the Appendix, it follows that  $\boldsymbol{\Omega}^{-1/2}(\Xi^{(S)})\mathbf{Z}\Xi^{(S)}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}^{(0)}) \xrightarrow{d} N(\mathbf{0}_{qd+K+1}, \mathbf{I}_{qd+K+1})$ , as  $n \rightarrow \infty$  and  $p = o(n^{1/2})$ .*

The selection of the subset  $\mathcal{S} = \{s_1, \dots, s_q\} \subset \{1, \dots, p\}$  and its size  $q$  is arbitrary in Theorem 1, so long as  $q$  is finite. Accordingly, for any given  $\mathcal{S}$ , Theorem 1 provides the joint limiting distribution for  $(\hat{\beta}_{s_1}^\top, \dots, \hat{\beta}_{s_q}^\top, \hat{\alpha}^\top)^\top$ . Furthermore, the theorem implies  $(\hat{\beta}_{s_1}^\top, \dots, \hat{\beta}_{s_q}^\top)^\top$  and  $\hat{\alpha}$  are  $\sqrt{n}$ - and  $\sqrt{np}$ -consistent, respectively. This is not an overly surprising result: since the mean regression coefficients are heterogeneous across responses, then only  $n$  observations  $\{Y_{ij} : i = 1, \dots, n\}$  contribute information to the estimation of each  $\beta_j$ . By contrast,  $\alpha$  are parameters common across all responses, and so its estimation leverages information across both  $i = 1, \dots, n$  and  $j = 1, \dots, p$ .

Based on the finding in Theorem 1, we have the following result for  $\hat{\alpha}$ .

## 2.2 Asymptotic Theory

**Corollary 1.** *Under Conditions 1 – 5 in the Appendix, it follows that  $P(\hat{\alpha} \in \mathcal{A}^+) \rightarrow 1$ , as  $n \rightarrow \infty$  and  $p = o(n^{1/2})$ .*

Corollary 1 provides a theoretical guarantee that the estimated reparameterized correlation regression parameters  $\hat{\alpha}$  from the joint estimator  $\hat{\vartheta}$  will fall into the required parameter space  $\mathcal{A}^+$  with probability tending to one. This allows us to apply Proposition 1 to obtain the estimated correlation regression parameters  $\hat{\rho} \in \mathcal{P}^+$  via the transformation  $\hat{\rho}_1 = \hat{\alpha}_1/\hat{\alpha}_0, \dots, \hat{\rho}_K = \hat{\alpha}_K/\hat{\alpha}_0$ .

We now proceed to the estimated parameter vector of interest, namely  $\hat{\theta} = (\hat{\beta}^\top, \hat{\rho}^\top)^\top$ . Based on the transformation used to obtain  $\hat{\rho}$  from  $\hat{\alpha}$ , we can apply techniques similar to the multivariate delta method with the vector-valued function  $\mathbf{f}(\vartheta) = (\beta^\top, \alpha_1/\alpha_0, \dots, \alpha_K/\alpha_0)^\top$  as follows. Let  $\theta^{(0)} = (\beta^{(0)\top}, \rho^{(0)\top})^\top$  denote the true parameter value for  $\theta$ , where  $R(\rho^{(0)})$  is a valid correlation matrix, and define  $\bar{\Xi}^{(S)} = \text{diag}(\mathbf{T}^{(S)} \otimes \mathbf{I}_d, \mathbf{I}_K)$  such that  $\bar{\Xi}^{(S)} \hat{\theta} = (\hat{\beta}_{s_1}^\top, \dots, \hat{\beta}_{s_q}^\top, \hat{\rho}^\top)^\top$  is a finite dimensional sub-vector of  $\hat{\theta}$ . Furthermore, let  $\bar{\mathbf{Z}} = \text{diag}(\sqrt{n}\mathbf{I}_{qd}, \sqrt{np}\mathbf{I}_K)$  and  $\bar{\Omega}(\bar{\Xi}^{(S)}) = \bar{\Xi}^{(S)} \{ \partial \mathbf{f}(\vartheta^{(0)}) / \partial \vartheta^\top \} \mathbf{B}^{-1} \mathbf{U} \mathbf{B}^{-1\top} \{ \partial \mathbf{f}(\vartheta^{(0)}) / \partial \vartheta^\top \}^\top \bar{\Xi}^{(S)\top}$ . We now state our main result for the joint estimator  $\hat{\theta} = (\hat{\beta}^\top, \hat{\rho}^\top)^\top$ .

**Theorem 2.** *Under Conditions 1 – 5 in the Appendix, it follows that  $\bar{\Omega}^{-1/2}(\bar{\Xi}^{(S)}) \bar{\mathbf{Z}} \bar{\Xi}^{(S)}(\hat{\theta} - \theta^{(0)}) \xrightarrow{d} N(\mathbf{0}_{qd+K}, \mathbf{I}_{qd+K})$ , as  $n \rightarrow \infty$  and  $p = o(n^{1/2})$ .*

Theorem 2 is valid under both cases of fixed  $p$  and diverging  $p$  as long as  $p/\sqrt{n} \rightarrow 0$ . The theorem shows any finite dimensional sub-vector of the es-

timator  $\hat{\theta}$  is consistent and asymptotically normally distributed, and forms the basis by which we can perform inference on the mean regression coefficients and correlation regression parameters of the proposed model. Moreover, Theorem 2 (again) implies  $(\hat{\beta}_{s_1}^\top, \dots, \hat{\beta}_{s_q}^\top)^\top$  and  $\hat{\rho}$  are  $\sqrt{n}$ - and  $\sqrt{np}$ -consistent, respectively, and that the asymptotic covariance matrix of  $(\hat{\beta}_{s_1}^\top, \dots, \hat{\beta}_{s_q}^\top, \hat{\rho}^\top)^\top$  is  $\mathbf{G}^{(\mathcal{S})} = \bar{\mathbf{Z}}^{-1} \bar{\mathbf{\Omega}}(\bar{\mathbf{\Xi}}^{(\mathcal{S})}) \bar{\mathbf{Z}}^{-1}$ . In practice, to apply the above theorem we need a consistent estimator of  $\mathbf{G}^{(\mathcal{S})}$  whose expression (see Lemma S1 in supplementary material S4) turns out to be dependent on  $\vartheta^{(0)}$ , the third-order moment  $\mu^{(3)}$ , and the fourth-order moment  $\mu^{(4)}$  defined in Condition 1. From this, we can verify that a consistent estimator can be obtained by replacing  $\vartheta^{(0)}$ ,  $\mu^{(3)}$  and  $\mu^{(4)}$  in  $\mathbf{G}^{(\mathcal{S})}$  by  $\hat{\vartheta}$ , the third and fourth-order empirical moments, respectively; see also supplementary material S1 for details on this.

### 3. Estimation Procedure

We develop an iterative estimation procedure to compute the joint estimator of  $\beta$  and  $\rho$  based on the two sets of estimating equations in Section 2.1, while allowing for the estimation of (potentially unknown) dispersion parameters and ensuring the estimators of  $\alpha$  (and  $\rho$ ) are in the required parameter space.

The estimation procedure consists of iterating between the following two steps. Given  $\alpha$ , we employ a Fisher scoring method, with a tuning param-

ter  $\gamma > 0$  to adjust the step size if appropriate, to solve  $\psi_\beta(\beta, \alpha) = \mathbf{0}_{pd}$ . This leads to an update of the form  $\beta \leftarrow \beta - \gamma \mathbf{J}^{-1}(\beta, \alpha) \psi_\beta(\beta, \alpha)$  where  $\mathbf{J}(\beta, \alpha) = E\{\partial \psi_\beta(\beta, \alpha) / \partial \beta^\top\} = -\mathbf{D}^\top(\beta) \mathbf{A}^{-1/2}(\beta) \tilde{\Sigma}^{-1}(\alpha) \mathbf{A}^{-1/2}(\beta) \mathbf{D}(\beta)$ . Next, given  $\beta$ , directly solving  $\psi_\alpha(\beta, \alpha) = \mathbf{0}_{K+1}$  leads to the closed-form solution  $\alpha = (\text{tr}(\tilde{\mathbf{W}}_{k_1} \tilde{\mathbf{W}}_{k_2}))_{(K+1) \times (K+1)}^{-1} ((\epsilon^\top(\beta) \tilde{\mathbf{W}}_k \epsilon(\beta))_{(K+1) \times 1})$ . This is equivalent to the solution of the unconstrained least squares optimization problem  $\min_\alpha Q(\beta, \alpha)$  defined below equation (2.5). However, while simple to compute, this unconstrained estimator is not guaranteed to be in the required parameter space  $\mathcal{A}^+$  in finite samples as Corollary 1 is an asymptotic result. As such, given  $\beta$ , we instead consider solving the constrained least squares optimization problem

$$\min_{\alpha \in \mathcal{A}^+} Q(\beta, \alpha), \quad (3.1)$$

which ensures the resulting estimator of  $\alpha$  is in the parameter space  $\mathcal{A}^+$ . This constrained optimization can be solved, for instance, using the alternating direction method of multipliers (ADMM) algorithm, which essentially converts (3.1) into an equivalent optimization problem  $\min_{\alpha, \Delta} \{Q(\beta, \alpha) : \Sigma(\alpha) = \Delta, \Delta - \nu \mathbf{I}_p \text{ is positive definite}\}$  by introducing a  $p \times p$  augmented parameter matrix  $\Delta$ , where  $\nu$  is an arbitrarily small positive constant to guarantee positive definiteness. The solution of (3.1) is then obtained by minimizing the augmented La-

grangian function of the latter optimization problem through an iterative way; see supplementary material S2 for details of the iterative steps in the ADMM algorithm, as well as Xue et al. (2012) and Zou et al. (2017) for more general discussion of the algorithm. If the unconstrained estimator is in  $\mathcal{A}^+$ , the ADMM algorithm will end up producing the same estimator as the unconstrained estimator. Therefore, in practice it is computationally efficient to first compute the unconstrained estimator and check if it is in  $\mathcal{A}^+$ . Only when this is not satisfied do we proceed to solve (3.1). Finally, if required we can estimate each element  $\phi_j$  of the dispersion parameter vector  $\phi = (\phi_1, \dots, \phi_p)^\top$  by solving the estimating equation based on the moment condition  $E[\{Y_{ij} - \mu_{ij}(\beta_j)\}^2 / h\{\mu_{ij}(\beta_j); \phi_j\}] = 1$ , where the means  $\mu_{ij}$  are evaluated at the true mean regression coefficients.

A formal algorithm detailing the estimation procedure is provided in supplementary material S2. To summarize, it begins by estimating  $\beta$  using the above Fisher scoring method given some initial values of  $\beta$  and  $\alpha$ , before computing the unconstrained estimator of  $\alpha$  by solving  $\min_{\alpha} Q(\beta, \alpha)$  given the updated  $\beta$ . If the unconstrained estimator is not in the parameter space  $\mathcal{A}^+$ , then the ADMM algorithm is used to solve the constrained optimization problem (3.1). Afterward, the estimated  $\alpha$  is transformed into  $\rho$  by following Proposition 1, and the iterative updates of  $\beta$  and  $\rho$  proceed until convergence.

#### 4. Simulation Study

To assess the finite sample performance of the proposed estimator, we conducted a simulation study by generating correlated multi-response data from Bernoulli, Poisson, negative binomial, and Gaussian distributions, and considering different sample sizes  $n \in \{50, 100, 200, 400\}$  and number of responses  $p \in \{10, 25, 50\}$  with  $d = 4$  covariates and  $K = 5$  similarity matrices. A total of 1000 datasets are simulated for each combination of responses distribution,  $n$  and  $p$ . Full details of the simulation setup are provided in supplementary material S7.1. To assess performance, we computed the mean square error (MSE) of the mean regression coefficient estimators  $\text{MSE}(\beta) = \sum_{j=1}^p \sum_{l=1}^d (\hat{\beta}_{jl} - \beta_{jl}^{(0)})^2 / pd$ , and the correlation regression parameter estimators  $\text{MSE}(\rho) = \sum_{k=1}^K (\hat{\rho}_k - \rho_k^{(0)})^2 / K$ , noting the MSE was scaled by number of associated parameters, where  $\hat{\beta}_{jl}$  and  $\hat{\rho}_k$  denote the corresponding elements of  $\hat{\beta}$  and  $\hat{\rho}$ , respectively, from the estimation procedure in Section 3, while  $\beta_{jl}^{(0)}$  and  $\rho_{jl}^{(0)}$  denote the true values.

For brevity, we present results for the case of Bernoulli and Poisson responses here; the results for negative binomial and Gaussian responses are provided in supplementary material S7.2 and present very similar conclusions. Figure 1 shows the averaged MSE (over 1000 replications) for both  $\beta$  and  $\rho$  decreased when  $n$  increased. Furthermore, the averaged MSE of  $\rho$  decreased when  $p$  increased, but the averaged MSE of  $\beta$  was relatively unaffected by  $p$ .



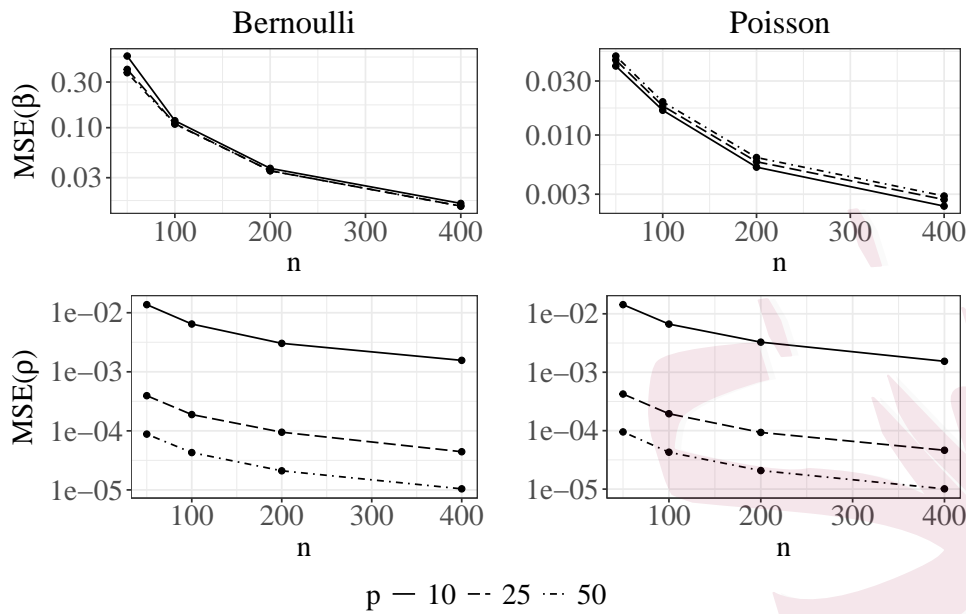


Figure 1: Averaged MSE of the mean regression coefficient estimators (top row) and correlation regression parameter estimators (bottom row) for Bernoulli (left column) and Poisson (right column) responses.

This is not surprising given  $\beta_j$ 's were heterogeneous across responses in the proposed model. By contrast, additional responses provided more information about the correlation structure, leading to improved estimation of  $\rho$ . The differing behaviour of the averaged MSE for  $\beta$  and  $\rho$  supports our theoretical findings in Theorem 2 that the mean regression coefficient estimators and correlation regression parameter estimators have different convergence rates i.e.,  $\sqrt{n}$  and  $\sqrt{np}$ , respectively. Finally, one interesting finding from our simulation study is that across all 1000 replications for various settings of responses distribution,  $n$  and  $p$ , the unconstrained estimator discussed in Section 3 was always in the parameter space  $\mathcal{A}^+$ . This aligns with Corollary 1 and reinforces the idea of

first computing the unconstrained estimator and checking whether it produces a positive definite covariance matrix. Turning to computational efficiency, Table S4 in supplementary material S7.2 provides details of the mean runtime for our proposed estimation algorithm, from which we see convergence was exceptionally quick (taking less than three seconds on average) using an AMD Ryzen 7 CPU @ 3.60 GHz machine.

Next, we studied the inferential performance of the joint estimator. Recall the definition of  $\mathbf{G}^{(S)}$  given below Theorem 2, which we further represent as a block matrix below,

$$\mathbf{G}^{(S)} = \begin{pmatrix} \mathbf{V}_{\beta_S} & \mathbf{V}_{\beta_S \rho} \\ \mathbf{V}_{\rho \beta_S} & \mathbf{V}_{\rho} \end{pmatrix}.$$

Denoting  $\hat{\beta}_S = (\hat{\beta}_{s_1}^\top, \dots, \hat{\beta}_{s_q}^\top)^\top$  and  $\beta_S^{(0)} = (\beta_{s_1}^{(0)\top}, \dots, \beta_{s_q}^{(0)\top})^\top$ , Theorem 2 implies the quantities  $(\hat{\beta}_S - \beta_S^{(0)})^\top \mathbf{V}_{\beta_S}^{-1} (\hat{\beta}_S - \beta_S^{(0)})$  and  $(\hat{\rho} - \rho^{(0)})^\top \mathbf{V}_{\rho}^{-1} (\hat{\rho} - \rho^{(0)})$  asymptotically follow  $\chi^2$  distributions with degrees of freedom  $qd$  and  $K$ , respectively, as  $n \rightarrow \infty$  and  $p = o(n^{1/2})$ . Based on this, we investigated the empirical coverage probability of 95% confidence regions for  $\beta_S^{(0)}$  and  $\rho^{(0)}$  by considering the mean regression coefficients for the first five responses i.e.,  $\mathcal{S} = \{1, \dots, 5\}$ , and computing the proportion of simulated datasets in which  $(\hat{\beta}_S - \beta_S^{(0)})^\top \hat{\mathbf{V}}_{\beta_S}^{-1} (\hat{\beta}_S - \beta_S^{(0)}) \leq \chi_{qd,0.95}^2$  and  $(\hat{\rho} - \rho^{(0)})^\top \hat{\mathbf{V}}_{\rho}^{-1} (\hat{\rho} - \rho^{(0)}) \leq \chi_{K,0.95}^2$ . Here,  $\chi_{l,0.95}^2$  denotes the 95% quantile of the  $\chi^2$  distribution with degrees of free-

Table 1: Empirical coverage probability for 95% confidence regions of  $\beta_S^{(0)}$  and  $\rho^{(0)}$  for Bernoulli and Poisson responses, where  $\mathcal{S} = \{1, \dots, 5\}$ .

		Bernoulli			Poisson		
		$p = 10$	$p = 25$	$p = 50$	$p = 10$	$p = 25$	$p = 50$
$\beta_S^{(0)}$	$n = 50$	0.957	0.955	0.958	0.851	0.856	0.857
	$n = 100$	0.966	0.968	0.967	0.927	0.929	0.928
	$n = 200$	0.969	0.969	0.970	0.933	0.936	0.934
	$n = 400$	0.968	0.968	0.968	0.941	0.943	0.943
$\rho^{(0)}$	$n = 50$	0.852	0.874	0.886	0.857	0.858	0.880
	$n = 100$	0.901	0.920	0.918	0.917	0.909	0.912
	$n = 200$	0.932	0.930	0.933	0.924	0.925	0.940
	$n = 400$	0.930	0.945	0.933	0.924	0.941	0.950

dom  $l$ , and  $\hat{V}_{\beta_S}$  and  $\hat{V}_{\rho}$  are the estimators of  $V_{\beta_S}$  and  $V_{\rho}$ , respectively, obtained from the corresponding blocks of the consistent estimator  $\hat{G}^{(S)}$  of  $G^{(S)}$ ; see supplementary material S1 for further details on computing  $\hat{G}^{(S)}$ . Table 1 demonstrates that the resulting coverage of 95% confidence regions for  $\beta_S^{(0)}$  tended to the nominal level of 95% as  $n$  increased, but was relatively unaffected by  $p$ . There was slight overcoverage for the case of Bernoulli responses. The coverage of 95% confidence regions for  $\rho^{(0)}$  tended to 95% as  $n$  and/or  $p$  increased. In supplementary material S7.2, we present additional results for the empirical coverage probability of 95% confidence intervals for the mean regression coefficients and correlation regression parameters individually, and obtain similar conclusions.

In supplementary material S7.3, we present results comparing our proposed estimator to existing GEE methods assuming either an independence working correlation matrix, or an unstructured working correlation matrix. Overall, the

---

proposed estimator was shown to have comparable performance to these existing methods in terms of estimating  $\beta$  (this is not surprising given the robustness of the mean regression coefficient estimators to the specification of working correlation structure), but clearly outperformed GEE with unstructured working correlation matrix in recovering the true correlation matrix given it leverages the additional information available from the similarity measures. Finally, we carried out numerical studies in supplementary material S7.4 to investigate the impact of misspecifying similarity measures on the estimation performance of the proposed estimator. Results show the estimation of  $\beta$  was again largely unaffected by this misspecification, while the estimation performance for  $\rho$  and hence recovery of the true correlation matrix exhibited a small deterioration in performance.

## 5. Application to Scotland Carabidae Ground Beetle Dataset

We applied the proposed joint mean and correlation regression model to a multivariate abundance dataset from ecology comprising overdispersed counts of Carabidae ground beetle species in Scotland. The data was sampled from a total of  $n = 87$  sites spread across nine main areas in Scotland using pitfall traps (Ribera et al., 2001), with the aim of the study being to jointly quantify the effects of environmental processes and trait mediation on the ground beetle assemblages;

---

see Ribera et al. (2001) for more details of the study. For illustrative purposes, we considered a subset of  $p = 38$  carabid ground beetle species that were detected in at least 15 sites. Along with species abundance observations, we considered three environmental covariates in soil pH, elevation above sea level, and land management intensity score (Downie et al., 1999), all of which were centered and scaled to have zero mean and unit variance prior to analysis. Together with the intercept, this leads to  $d = 4$  predictors in  $\mathbf{x}_i$ . For each species, we also have records for  $K = 5$  trait predictors, with one being quantitative (total length) and four being qualitative (color of the legs, wing development, overwintering and breeding season). We converted each of the trait predictors into a similarity matrix based on the procedure discussed above equation (2.1), depending on whether it is a quantitative or qualitative trait.

From an exploratory analysis (see Figure S11 in supplementary material S8), we observed evidence of overdispersion and a quadratic mean-variance relationship for the  $p = 38$  species. Along with the fact that different carabid beetle species are known to exhibit diverse responses to the environment, we thus proceeded to fit the joint mean and correlation regression model assuming a log link function for the mean, and a quadratic variance function  $h(\mu; \phi) = \mu + \phi\mu^2$  where the species-specific overdispersion parameters  $\phi_j > 0$  were estimated using the estimation procedure in Section 3. Based on Theorem 2, we also con-

Table 2: Point estimates and 95% confidence intervals (in parentheses) for mean regression coefficients of the first ten species, and correlation regression parameters for all five trait variables, based on analysis of the Carabidae ground beetle dataset. Estimates whose confidence interval excludes zero are bolded.

Estimation of $\beta_j$				
Species	Intercept	Soil pH	Elevation	Land Management
<i>A.Muelleri</i>	<b>2.444</b> ( <b>2.015, 2.872</b> )	<b>1.382</b> ( <b>0.631, 2.132</b> )	<b>-0.730</b> ( <b>-1.396, -0.064</b> )	-0.022 (-0.725, 0.682)
<i>A.Apricaria</i>	-0.697 (-1.528, 0.135)	-0.151 (-1.186, 0.883)	<b>-1.649</b> ( <b>-2.904, -0.393</b> )	<b>1.615</b> ( <b>0.692, 2.538</b> )
<i>A.Bifrons</i>	<b>-1.349</b> ( <b>-2.462, -0.236</b> )	<b>1.415</b> ( <b>0.134, 2.696</b> )	<b>-2.145</b> ( <b>-3.720, -0.570</b> )	0.314 (-0.775, 1.404)
<i>A.Communis</i>	<b>0.604</b> ( <b>0.145, 1.062</b> )	-0.080 (-0.876, 0.715)	-0.439 (-1.058, 0.181)	<b>-0.970</b> ( <b>-1.741, -0.198</b> )
<i>A.Familiaris</i>	0.221 (-0.382, 0.823)	0.454 (-0.593, 1.500)	-0.371 (-1.287, 0.544)	0.474 (-0.509, 1.458)
<i>A.Lunicollis</i>	-0.016 (-0.533, 0.502)	-0.714 (-1.644, 0.217)	<b>-0.841</b> ( <b>-1.600, -0.082</b> )	-0.542 (-1.401, 0.317)
<i>A.Plebeja</i>	<b>3.074</b> ( <b>2.667, 3.480</b> )	<b>1.576</b> ( <b>0.860, 2.291</b> )	<b>-0.933</b> ( <b>-1.567, -0.298</b> )	-0.390 (-1.063, 0.283)
<i>A.Dorsalis</i>	0.106 (-0.513, 0.725)	0.305 (-0.544, 1.153)	-0.144 (-1.064, 0.776)	<b>2.557</b> ( <b>1.720, 3.394</b> )
<i>B.Aeneum</i>	<b>2.010</b> ( <b>1.265, 2.756</b> )	<b>1.467</b> ( <b>0.262, 2.673</b> )	<b>-2.993</b> ( <b>-4.314, -1.672</b> )	-0.236 (-1.345, 0.873)
<i>B.Guttula</i>	<b>1.948</b> ( <b>1.560, 2.336</b> )	<b>0.921</b> ( <b>0.264, 1.577</b> )	<b>-1.024</b> ( <b>-1.672, -0.376</b> )	0.448 (-0.159, 1.055)
Estimation of $\rho_k$				
Total Length	Leg Color	Wing Development	Overwintering	Breeding Season
<b>0.061</b> ( <b>0.023, 0.098</b> )	-0.002 (-0.022, 0.018)	0.021 (-0.001, 0.044)	0.004 (-0.018, 0.026)	<b>0.036</b> ( <b>0.008, 0.064</b> )

struced 95% confidence intervals for each of the mean regression coefficients and correlation regression parameters; see supplementary material S1 for further details on this construction.

Table 2 reports the estimated mean regression coefficients for the first ten carabid ground beetle species, along with the estimated correlation regression parameters for all five trait variables; see supplementary material S8 for full estimation results of all  $p = 38$  species. As expected, carabid ground beetle species possessed considerable heterogeneity in their responses to the environment. For

---

instance, in comparing *A.Muelleri* and *A.Communis*, results show the former preferred higher levels of soil pH, while the latter presented no clear evidence of being influenced by this habitat factor. Although elevation was negatively associated with both species i.e., both tended to be recorded at lower altitude, it was only statistically significant for *A.Muelleri*. Conversely, only *A.Communis* exhibited statistically clear evidence of being negatively impacted by increased land management intensity. Elevation was negatively associated with all the abundances of the first ten carabid species in Table 2, although we note that 11 other carabid species exhibited positive estimated mean regression coefficients for the elevation covariate (see Table S7 in supplementary material S8).

Turning to the estimated correlation regression parameters, all trait variables except leg color exhibited a positive association with the correlation between the abundances of different carabid species. The correlation regression parameter estimate related to total length displayed the strongest magnitude, and its confidence interval excluded zero. This suggested conditional on other trait values, the abundances of carabid species with similar total lengths were more positively associated after accounting for differences in their mean abundances due to environmental filtering. There was also statistically clear evidence that breeding season was important in driving residual covariations between carabid species.

In supplementary material S8, we performed some additional analyses to

---

investigate the sensitivity of the proposed model to alternative specifications of the similarity measures. Results show the estimated mean regression coefficients of all beetle species, the correlation regression parameters of the trait similarity matrices, and the between-species correlation matrix remained similar across different specifications of similarity measures in this application. This consistency in conclusions further substantiates the inferences drawn from Table 2.

## 6. Conclusion

We have introduced a joint mean and correlation regression model for correlated multi-response data, which allows simultaneous analysis of the relationship between the mean components with observed covariates, and the association between the correlation components with similarity measures of additional predictor information. The proposed model can be applied to a wide variety of discrete and (semi-)continuous responses. We developed a joint estimator for the mean regression coefficients and correlation regression parameters, which is demonstrated to be consistent and asymptotically normal as the sample size  $n \rightarrow \infty$  and the number of responses  $p$  grows with increasing  $n$ . Similar to standard GEEs, the consistency of the proposed mean regression coefficient estimators is robust to misspecification of the correlation regression component of the joint model. Simulation studies demonstrate the strong empirical performance of the



---

proposed joint estimator, especially compared with existing methods such as GEE assuming an unstructured working correlation matrix in estimating the true correlation matrix. The application of our proposed model to the ground beetle dataset revealed heterogeneous relationships between environmental factors and carabid beetle abundances, while also identifying important functional traits which drive the (residual) correlation between species. A sensitivity analysis also yielded similar conclusions for the same dataset under alternative specifications of similarity measures.

Note the findings from fitting our proposed model to the ground beetle dataset are novel and differ from those obtained from the fourth corner analysis by Ribera et al. (2001), and indeed more generally fourth corner analyses in community ecology (e.g., Niku et al., 2021). In particular, while fourth corner models aim at identifying where traits mediate species responses to the environment, our approach quantifies how different traits influence the residual correlation between species after accounting for species-environmental responses. These are fundamentally different scientific questions, and while statistical methods have been developed for the former, our method is one of the first which specifically addresses the latter in statistical ecology.

A logical next step would be to extend the proposed model to handle mixed response types. This could be achieved by considering link functions  $g_j(\cdot)$  and

---

variance functions  $h_j(\cdot)$  that are allowed to differ based on the  $j$ -th response type, and Condition 4 would be altered so that its requirements apply to all link and variance functions. This could be useful in ecology, say, when we observe count records for some species and presence-absence records for other species, in which case  $g_j(\mu) = \log(\mu)$  and  $h_j(\mu; \phi) = \phi\mu$  for the count responses while  $g_j(\mu) = \log\{\mu/(1 - \mu)\}$  and  $h_j(\mu; \phi) = \phi\mu(1 - \mu)$  for the binary responses, noting that there might be complications in the interpretation of the response-specific mean regression coefficients because of the use of different link functions resulting in differing effect sizes (see also Section 6 of Hui et al., 2024). It would also be interesting to consider similarity matrices  $\mathbf{W}_k^{(i)}$  that are heterogeneous across  $i = 1, \dots, n$ ; this would give rise to heterogeneous correlation matrices  $\mathbf{R}^{(i)}$  for different response vectors  $\mathbf{Y}_i$  (e.g., Zou et al., 2022). Additionally, while this article focuses on a balanced setting by assuming each of the  $i$ -th cluster consists of  $p$  number of responses, the proposed method could be generalized to accommodate unbalanced data where we observe  $p_i$  rather than  $p$  responses at the  $i$ -th cluster (analogous to the study of clustered data with unequal cluster sizes e.g., Section 3.5 of Xue et al., 2010). In this instance, similar idea of correlation regression could be employed to model the  $i$ -th cluster correlation matrix as  $\mathbf{R}^{(i)}(\boldsymbol{\rho}) = \mathbf{I}_{p_i} + \sum_{k=1}^K \rho_k \mathbf{W}_k^{(i)}$  involving  $p_i \times p_i$  matrices, which could be considered as one special case of the heterogeneous similarity matrices

---

discussed above. The proposed method could then be adapted by appropriately adjusting the dimensions of the matrices involved in the estimation procedures to reflect the varying cluster sizes e.g.,  $\tilde{\mathbf{W}}_k$  would be changed from  $\mathbf{I}_n \otimes \mathbf{W}_k$  under the balanced data setting to  $\text{diag}(\mathbf{W}_k^{(1)}, \dots, \mathbf{W}_k^{(n)})$ .

Finally, the proposed model can be extended to allow for the clusters to be spatially and/or temporally correlated across  $i = 1, \dots, n$  e.g., by considering the generalized Kronecker product structure of Bonat and Jørgensen (2016) that involves separate modeling of the between-response and between-cluster dependence structures. In this case, the correlation regression model would still be used to model the between-response dependence structure, while similar idea could be employed to model the between-cluster dependence structure as a linear combination of similarity matrices constructed from spatio-temporal distances between different clusters (see also Hui, 2022, for a similar idea). In such cases, it may be particularly useful to consider joint variable selection on the mean regression coefficients and correlation regression parameters as the number of covariates  $d$  and similarity matrices  $K$  could be large in real practice. This could be accomplished by leveraging existing work on penalized GEEs or fast information criterion (Wang et al., 2012; Hui et al., 2023); see also the recent work of Tho et al. (2024) who considered variable selection to identify the subset of relevant similarity matrices in the context of Ising models.

## Appendix

We introduce five regularity conditions under which all theoretical results are established. Let  $\|\cdot\|_t$  denote the vector  $t$ -norm or matrix  $t$ -norm for  $1 \leq t \leq \infty$ ,  $|\mathbf{H}|_\infty = \|\text{vec}(\mathbf{H})\|_\infty$  denote the element-wise  $\ell_\infty$  norm for a generic matrix  $\mathbf{H}$  with  $\text{vec}(\mathbf{H})$  being the vectorization of matrix  $\mathbf{H}$ , and  $\lambda_{\min}(\mathbf{H})$  denote the smallest eigenvalue of a generic square matrix  $\mathbf{H}$ . The discussions of the following conditions are presented in supplementary material S3.

**Condition 1.** *The elements  $\varepsilon_\ell$  of the  $np$ -dimensional random vector  $\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}^{(0)}) = \tilde{\mathbf{L}}_0^{-1} \mathbf{A}^{-1/2}(\boldsymbol{\beta}^{(0)})\{\mathbf{Y} - \boldsymbol{\mu}(\boldsymbol{\beta}^{(0)})\}$  are independent and identically distributed with mean zero, variance one, third-order moment  $\mu^{(3)}$ , and fourth-order moment  $\mu^{(4)}$ , where  $\tilde{\mathbf{L}}_0 = \mathbf{I}_n \otimes \mathbf{L}_0$ , and  $\mathbf{L}_0$  is obtained through the Cholesky decomposition of  $\boldsymbol{\Sigma}(\boldsymbol{\alpha}^{(0)}) = \mathbf{L}_0 \mathbf{L}_0^\top$ . Furthermore, there exists some  $\eta > 0$  such that  $E(|\varepsilon_\ell|^{4+\eta}) < \infty$  for  $\ell = 1, \dots, np$ .*

**Condition 2.** *There exist finite positive constants  $C_L, C_X$  and  $C_W$  such that  $\max\{\|\mathbf{L}_0\|_1, \|\mathbf{L}_0^{-1}\|_1, \|\mathbf{L}_0\|_\infty, \|\mathbf{L}_0^{-1}\|_\infty\} \leq C_L$ ,  $|\mathbf{X}|_\infty \leq C_X$ , and  $\|\mathbf{W}_k\|_1 \leq C_W$  for any  $n \geq 1$ ,  $p \geq 1$  and  $k = 1, \dots, K$ , where  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$ .*

**Condition 3.** *There exists some finite positive constant  $C_\beta$  such that  $\|\boldsymbol{\beta}^{(0)}\|_\infty \leq C_\beta$  for any  $p \geq 1$ .*

**Condition 4.** *The inverse link function  $g^{-1}(\cdot)$  is twice continuously differen-*

---

tible, the variance function  $h(\cdot)$  is continuously differentiable, and the composition function  $(h \circ g^{-1})(\cdot) \geq C_h$  for some finite positive constant  $C_h$ .

**Condition 5.** *There exist finite positive constants  $C_B$  and  $C_\Omega$  such that*

$\max\{\|\mathbf{B}^{-1}\|_1, \|\mathbf{B}^{-1}\|_\infty\} \leq C_B$  and  $\lambda_{\min}(\boldsymbol{\Omega}) \geq C_\Omega$  for any  $n \geq 1$  and  $p \geq 1$ , where  $\boldsymbol{\Omega} = \mathbf{B}^{-1}\mathbf{U}\mathbf{B}^{-1\top}$ .

### Supplementary Material

The Supplementary Material includes all proofs and algorithms, as well as additional results for the simulation study and real data application.

### Acknowledgments

Zhi Yang Tho was supported by an Australian Government Research Training Program scholarship. Francis KC Hui was supported by an Australian Research Council Discovery Project DP240100143. Tao Zou's research was supported by computational resources provided by the Australian Government through the National Computational Infrastructure (NCI), under the ANU Startup Allocation Scheme. The authors thank Alan Welsh for his helpful comments.

## REFERENCES

---

### References

Anderson, T. W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure.

*Ann. Statist. 1*, 135–141.

Bonat, W. H. and B. Jørgensen (2016). Multivariate covariance generalized linear models. *J. R. Statist.*

*Soc. C 65*, 649–675.

Chiu, T. Y. M., T. Leonard, and K.-W. Tsui (1996). The matrix-logarithmic covariance model. *J. Am.*

*Statist. Assoc. 91*, 198–210.

Downie, I. S., W. L. Wilson, V. J. Abernethy, D. I. McCracken, G. N. Foster, I. Ribera, K. J. Murphy, and

A. Waterhouse (1999). The impact of different agricultural land-uses on epigeal spider diversity in

Scotland. *J. Insect Conserv. 3*, 273–286.

Fitzmaurice, G., N. Laird, and J. Ware (2011). *Applied Longitudinal Analysis*. Hoboken, New Jersey:

Wiley.

Hu, J., Y. Chen, C. Leng, and C. Y. Tang (2024). Applied regression analysis of correlations for correlated

data. *Ann. Appl. Stat. 18*, 184–198.

Hui, F. K. C. (2022). GEE-assisted forward regression for spatial latent variable models. *Journal of*

*Computational and Graphical Statistics 31*, 1013–1024.

Hui, F. K. C., K.-D. Dang, and L. Maestrini (2024). Simultaneous coefficient clustering and sparsity for

multivariate mixed models. *Journal of Computational and Graphical Statistics 34*, 1–12.

Hui, F. K. C., S. Müller, and A. Welsh (2023). GEE-assisted variable selection for latent variable models

## REFERENCES

---

- with multivariate binary data. *J. Am. Statist. Assoc.* 118, 1252–1263.
- Johnson, R. and D. Wichern (2013). *Applied Multivariate Statistical Analysis: Pearson New International Edition*. London, U.K.: Pearson Education.
- Liang, K.-Y. and S. L. Zeger (1986). Longitudinal data analysis using generalized linear models. *Biometrika* 73, 13–22.
- Lipsitz, S. and G. Fitzmaurice (2008). Generalized estimation equations for longitudinal data analysis. In G. Fitzmaurice, M. Davidian, G. Verbeke, and G. Molenberghs (Eds.), *Longitudinal Data Analysis*, Chapter 3, pp. 43–78. Boca Raton, Florida: Chapman & Hall/CRC.
- Müller, S., J. L. Scaely, and A. H. Welsh (2013). Model selection in linear mixed models. *Stat. Sci.* 28, 135–167.
- Niku, J., F. K. C. Hui, S. Taskinen, and D. I. Warton (2021). Analyzing environmental-trait interactions in ecological communities with fourth-corner latent variable models. *Environmetrics* 32, e2683.
- Ovaskainen, O. and N. Abrego (2020). *Joint Species Distribution Modelling: With Applications in R*. Cambridge, U.K.: Cambridge University Press.
- Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika* 86, 677–690.
- Qu, A., B. G. Lindsay, and B. Li (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika* 87, 823–836.
- Ribera, I., S. Dolédec, I. S. Downie, and G. N. Foster (2001). Effect of land disturbance and stress on

## REFERENCES

---

- species traits of ground beetle assemblages. *Ecology* 82, 1112–1129.
- Tang, C., W. Zhang, and C. Leng (2019). Discrete longitudinal data modeling with a mean-correlation regression approach. *Stat. Sin.* 29, 853–876.
- Tho, Z. Y., F. K. C. Hui, and T. Zou (2024). An Ising similarity regression model for modeling multivariate binary data. *Stat. Sin.*, <http://doi.org/10.5705/ss.202024.0021>.
- Tikhonov, G., N. Abrego, D. Dunson, and O. Ovaskainen (2017). Using joint species distribution models for evaluating how species-to-species associations depend on the environmental context. *Methods Ecol. Evol.* 8, 443–452.
- Wang, L., J. Zhou, and A. Qu (2012). Penalized generalized estimating equations for high-dimensional longitudinal data analysis. *Biometrics* 68, 353–360.
- Wang, Y.-G. and V. Carey (2003). Working correlation structure misspecification, estimation and covariate design: Implications for generalised estimating equations performance. *Biometrika* 90, 29–41.
- Warton, D. I. (2011). Regularized sandwich estimators for analysis of high-dimensional data using generalized estimating equations. *Biometrics* 67, 116–123.
- Warton, D. I., F. G. Blanchet, R. B. O’Hara, O. Ovaskainen, S. Taskinen, S. C. Walker, and F. K. C. Hui (2015). So many variables: Joint modeling in community ecology. *Trends Ecol. Evol.* 30, 766–779.
- Xue, L., S. Ma, and H. Zou (2012). Positive-definite  $\ell_1$ -penalized estimation of large covariance matrices. *J. Am. Statist. Assoc.* 107, 1480–1491.
- Xue, L., A. Qu, and J. Zhou (2010). Consistent model selection for marginal generalized additive model



## REFERENCES

---

- for correlated data. *J. Am. Statist. Assoc.* 105, 1518–1530.
- Ye, H. and J. Pan (2006). Modelling of covariance structures in generalised estimating equations for longitudinal data. *Biometrika* 93, 927–941.
- Zhang, W. and C. Leng (2012). A moving average Cholesky factor model in covariance modelling for longitudinal data. *Biometrika* 99, 141–150.
- Zhang, W., C. Leng, and C. Y. Tang (2015). A joint modelling approach for longitudinal studies. *J. R. Statist. Soc. B* 77, 219–238.
- Zhu, J. and E. P. Xing (2011). Sparse topical coding. In *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence*, UAI’11, Barcelona, Spain, pp. 831–838. AUAI Press.
- Zou, T., W. Lan, R. Li, and C.-L. Tsai (2022). Inference on covariance-mean regression. *J. Econometrics* 230, 318–338.
- Zou, T., W. Lan, H. Wang, and C.-L. Tsai (2017). Covariance regression analysis. *J. Am. Statist. Assoc.* 112, 266–281.
- Zou, T., R. Luo, W. Lan, and C.-L. Tsai (2020). Covariance regression model for non-normal data. In C. F. Lee and J. C. Lee (Eds.), *Handbook of Financial Econometrics, Mathematics, Statistics, and Machine Learning*, Volume 4, Chapter 113, pp. 3933–3945. Singapore: World Scientific.
- Zou, T., R. Luo, W. Lan, and C.-L. Tsai (2021). Network influence analysis. *Stat. Sin.* 31, 1727–1748.
- Zwiernik, P., C. Uhler, and D. Richards (2017). Maximum likelihood estimation for linear Gaussian covariance models. *J. R. Statist. Soc. B* 79, 1269–1292.

## REFERENCES

---

Zhi Yang Tho

The Australian National University, Canberra, ACT 2600, Australia.

E-mail: zhiyang.tho@anu.edu.au

Francis K.C. Hui

The Australian National University, Canberra, ACT 2600, Australia.

E-mail: francis.hui@anu.edu.au

Tao Zou

The Australian National University, Canberra, ACT 2600, Australia.

E-mail: tao.zou@anu.edu.au