

## **A bivariate framework to jointly model count and continuous responses**

**Máira Blumer Fatoretto**

Department of Exact Sciences, University of São Paulo, Piracicaba, São Paulo, Brazil.

*email:* mairafatoretto@gmail.com

**and**

**Caroline Brophy**

School of Computer Science and Statistics, Trinity College Dublin, Dublin 2, Ireland.

*email:* caroline.brophy@tcd.ie

**and**

**Clarice Garcia Borges Demétrio**

Department of Exact Sciences, University of São Paulo, Piracicaba, São Paulo, Brazil.

*email:* clarice.demetrio@usp.br

**and**

**Rafael de Andrade Moral**

Department of Mathematics and Statistics, Maynooth University, Maynooth, Ireland.

*email:* rafael.deandrademoral@mu.ie

**SUMMARY:** Multivariate data are present in many studies in the natural sciences. Generally, the interest is in more than one response variable and in how factors and covariates affect them, simultaneously. However, when the interest

000 0000

is in one count variable and one continuous variable, we need to use different approaches to model these variables jointly. In this work, motivated by a data set from entomology, in which a continuous and a count response were observed simultaneously, a novel mean-dispersion bivariate model was developed. The proposed model is based on the bivariate normal approximations, and adds a new parameter to the variance-covariance matrix to better accommodate the main characteristics of count data, such as overdispersion or underdispersion. This modelling framework showed good results and also allowed for greater flexibility in the case of heterogeneity of variances, and allows for the correlation between the two responses to vary according to an experimental treatment.

KEY WORDS: Count data; Entomology; Joint modelling; Multivariate data; Overdispersion; Poisson model; Underdispersion.

## 1. Introduction

In many observational or experimental studies, multiple response variables are of interest, and they may be correlated. In statistical modelling techniques, multiple response variables can be analysed in isolation using univariate methods, where the dependence between them is not considered in the modelling. However, the correlation between the response variables can provide valuable insight, and their incorporation in the modelling framework can lead to more reliable tests and improved inference, especially when this correlation is of high magnitude. Multivariate regression techniques facilitate the inclusion of correlated responses and are useful for exploring data patterns that may exist in more than one dimension (Raykov and Marcoulides, 2008; Everitt and Hothorn, 2011).

For the majority of univariate analysis, many inferential results are based on the normal distribution, and since many multivariate techniques are extensions of univariate analysis, the majority of multivariate procedures have the multivariate normal distribution as their underpinning. For example, in bivariate analysis, the relationship between two continuous variables with unbounded symmetric distribution could be described by a joint probability distribution using the bivariate normal density (Jobson, 2012). However, other distributions according to characteristics of the response variables can also be considered. The analysis of non-normal univariate or multivariate data involves, mostly, the generalized linear models (GLM) formulated by Nelder and Wedderburn (1972), that represent an elegant and encompassing mathematical framework to model response variables whose distribution belongs to the exponential family (such as normal, binomial, Poisson, gamma and inverse Gaussian). A feature of the exponential family of distributions is the *mean-variance* relationship, i.e., the fact that the variance is a function of the mean.

Johnson et al. (1997) presented a range of discrete multivariate models that can jointly analyze count and binary data. Nevertheless, when it comes to joint modelling of one

discrete and one continuous variable, suitable techniques and current software implementations are still emerging (Bonat and Jørgensen, 2016). Bonat et. al. (2018) proposed a flexible framework to model response variables of different nature jointly (e.g. unbounded and bounded continuous variables and discrete variables). They created the multivariate covariance generalized linear modelling framework (McGLM), implemented as the `mcglm` package for R software (R Core Team, 2022). In their approach, marginal models are fitted through quasi-likelihood functions based on first and second-moment assumptions (Bonat et al., 2017; Bonat, 2018). The framework includes flexibility in the specification of the covariance structure and provides reliable tests for fixed effects. However, one drawback of this framework is that it does not provide an associated probability distribution. Another possibility for modelling discrete and continuous multivariate response variables jointly are copula models. The copula model allows us to incorporate the correlation between the variables and also allows flexibility when choosing the correlation structure, however, the use of different copulas generates completely different results (Nikoloulopoulos and Karlis, 2009; Krupskii and Joe, 2013). In this paper, we present another option for modelling count and continuous data jointly using an approximation based on the central limit theorem and an extension of the variance-covariance matrix.

When analyzing count data, the Poisson model is a natural first choice, however, the Poisson distribution can be approximated asymptotically by the normal distribution (Mood, 1973). A multivariate extension based on the normal distribution is advantageous in the sense that this distribution presents known and closed formulas. A drawback of the Poisson model is the assumption of equality of mean and variance, which is not met very often in practice due to extra variability and other features, such as zero-inflation (Haslett et al. (2021)). The overdispersion in count data may be caused by a deficiency of relevant covariates or heterogeneity of samples, or repeated measures, and it is necessary to model this extra

variability to obtain reliable inference about the parameters (Hinde and Demétrio, 1998; Ver Hoef and Boveng, 2007). Variance homogeneity is typically assumed for continuous data modelled by classical linear models based on the normal distribution; however, this variance assumption may not be reasonable or reflective of the real characteristics of the data. Aitkin (1987) proposed the joint modelling of mean and dispersion in the normal regression analysis using a Fisher-scoring algorithm for the simultaneous maximum likelihood estimation. This proposal brought a powerful approach to deal with the heterogeneity of variance, which was modelled rather than being transformed away. McCullagh and Nelder (1989) presented the joint modelling of mean and dispersion in the generalized linear models, using the extended quasi-likelihood function (Nelder and Pregibon, 1987).

In this paper, we present a joint modelling of bivariate data, which also could be extended to multivariate data of more than two responses, to accommodate two counts, or two continuous response variables, or a combination of one count and one continuous variable using a bivariate normal distribution approximation. To obtain a versatile framework to deal with count and continuous data, we included an extra parameter in the variance-covariance matrix to capture different characteristics present in many situations. The normal bivariate model proposed allows us to model the mean and dispersion jointly, while dealing with over- and underdispersion, as well as variance heterogeneity. Furthermore, it is possible to fit the model considering different correlation structures that can change depending on a covariate of interest. To exemplify the method developed, we present a case study arising from the area of entomology.

This paper is organized as follows. Section 2 presents the case-study that is a motivation for this paper. Section 3 presents the bivariate framework to jointly model count and continuous responses. Section 4 presents estimation and inference for the novel regression model based on the maximum likelihood paradigm. The results from the analysis of the case study are

presented in Section 5. The properties of the maximum likelihood and profile likelihood estimators are assessed in Section 6 through simulation studies. Finally, we present a discussion in Section 7.

## 2. Case study - motivating example

The data comes from an experiment examining *Podisus nigrispinus* (Hemiptera: Pentatomidae), a predatory stinkbug found in agricultural and forest systems in several countries of Central and South America, when fed two different prey (*Anticarsia gemmatalis* and *Diatraea saccharalis*, two agricultural pests). This stinkbug has an important role as a biological control agent for different crops, e.g., in forestry the predator can be useful to control defoliating caterpillars in *Eucalyptus* plantations. The response variables are the weight and the fertility of female predators. The weight is one way of measuring the development of an insect over its life cycle, and the fertility measures how well the insect can establish itself in an ecosystem. Here, it is interesting to model these two variables together because an adequate diet can help to guarantee a high level of fertility. This information is essential for the choice of biological control strategies in pest management programmes (Parra, 2002).

The experiment was carried out at the Laboratory of Forest Ecology and Entomology, Department of Entomology and Acarology, ESALQ - USP, Piracicaba, Brazil. The insects were placed in individualized incubator chambers, where they were fed and monitored daily. The two diets or treatments consisted of offering three caterpillars of either *A. gemmatalis* or *D. saccharalis* daily. The experiment started with 50 stinkbug insects for each diet, however, as there was a high mortality rate, the final measurements were made only with 9 replicates of the *A. gemmatalis* treatment and 18 of the *D. saccharalis* treatment. After 18 days of receiving the diet, the weight of females was measured. Next, the females were allocated with a male that also received the same diet. The couples were kept until the death of the females. All eggs laid by the females were placed in Petri dishes and counted.

Looking at the scatter plot between female weight and number of eggs (see Figure 1), we see that stinkbugs being fed the *A. gemmatalis* caterpillar diet performed more poorly overall than those fed the *D. saccharalis* caterpillar diet. A strong correlation between female weight and number of eggs is observed for the *D. saccharalis* diet; this indicates that a modelling approach that allows for varying correlation strength according to an experimental treatment could be suitable here.

[Figure 1 about here.]

### 3. Joint Modelling of a Count and a Continuous Response

We propose the same marginal distribution structure for either a count or a continuous variable. Both assume a normal distribution, with mean  $\boldsymbol{\mu}_i, i = 1, 2$  and standard deviation  $\phi_i \boldsymbol{\mu}_i, i = 1, 2$ , which depends on the mean and a dispersion parameter  $\phi_i$ . It is also possible to capture variance heterogeneity, overdispersion, or underdispersion by incorporating covariates in the dispersion parameters. We model the mean  $\boldsymbol{\mu}$  through a monotonic and differentiable link function  $\mathbf{g}(\boldsymbol{\mu}) = \boldsymbol{\eta}$ , such that the linear predictor is  $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ , with  $\mathbf{X}$  a design matrix and  $\boldsymbol{\beta}$  a vector of regression parameters.

Let  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  be vectors of count or continuous random variables. We then assume that, marginally,  $\mathbf{Y}_1 \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$  and  $\mathbf{Y}_2 \sim N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ , with the joint distribution of  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  taken to be multivariate normal, i.e.,

$$\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} \sim N \left( \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_2 \end{bmatrix} \right), \quad (1)$$

where  $\boldsymbol{\Sigma}_1$  denotes the variance-covariance matrix for the first response;  $\boldsymbol{\Sigma}_2$  denotes the variance-covariance matrix for the second response and  $\boldsymbol{\Sigma}_{12}$  denotes the covariance matrix between the two responses, where  $\boldsymbol{\Sigma}_{12}$  depends on a correlation parameter  $\rho$ ,  $\boldsymbol{\Sigma}_1$  and  $\boldsymbol{\Sigma}_2$ .

We write

$$\boldsymbol{\Sigma}_i = \mathbf{D}_\phi \mathbf{D}_\mu,$$

where  $\mathbf{D}_\phi = \text{diag}(\mathbf{X}_\phi \boldsymbol{\phi})$ ,  $\mathbf{D}_\mu = \text{diag}(g^{-1}(\mathbf{X}_\beta \boldsymbol{\beta}))$ , and the function  $\text{diag}(\mathbf{x})$  is defined as a square matrix with the elements of the vector  $\mathbf{x}$  in its main diagonal;  $\mathbf{X}_\phi$  and  $\mathbf{X}_\beta$  are the design matrices for the dispersion and mean parameters respectively; and  $\boldsymbol{\phi}$  and  $\boldsymbol{\beta}$  unknown parameter vectors. This yields

$$\boldsymbol{\Sigma}_i = \begin{pmatrix} \phi_{i1}\mu_{i1} & 0 & \dots & 0 \\ 0 & \phi_{i2}\mu_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \phi_{in}\mu_{in} \end{pmatrix},$$

where  $i \in \{1, 2\}$  is the index corresponding to each response variable and  $j = 1, \dots, n$  is the index corresponding to each observation.

#### 4. Estimation and Inference

We first show how estimation works for a single response, and in the next subsection extend it to the bivariate case of interest here.

##### 4.1 Univariate modelling

Taking one of the response outcomes  $\mathbf{Y}_i$ , from here we will omit the index  $i$  to improve readability. We assume that asymptotically  $\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma} = \mathbf{D}_\phi \mathbf{D}_\mu)$ , and write the log-likelihood as

$$\log \mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{y}) = l(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{y}) = -\frac{1}{2} \{ \log |\boldsymbol{\Sigma}| + (\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \} + \text{constant} \quad (2)$$

The estimating equations can be obtained using the chain rule. Differentiating (2) w.r.t.  $\boldsymbol{\Sigma}$  yields



$$\frac{\partial l}{\partial \Sigma} = -\frac{1}{2}(\Sigma^{-1})' + \frac{1}{2}(\Sigma^{-1})' + (\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'(\Sigma^{-1})' + . \quad (3)$$

Now further differentiating eq. (3) w.r.t.  $\mathbf{D}_\phi$  and w.r.t.  $\phi$  yields

$$-\frac{1}{2}\text{diag}(\mathbf{X}_\phi \phi)^{-1} \circ \mathbb{I} + \frac{1}{2}\text{diag}(\mathbf{X}_\phi \phi)^{-1} \mathbf{D}_\mu^{-1}(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})' \text{diag}(\mathbf{X}_\phi \phi)^{-1} \circ \mathbb{I}, \quad (4)$$

where  $\mathbb{I}$  is an  $n \times n$  identity matrix and  $\mathbb{J}$  is an  $n \times 1$  vector of 1s, and  $\circ$  the Hadamard product. Finally, solving eq. (4) w.r.t.  $\phi$  yields

$$\hat{\phi} = (\mathbf{X}'_\phi \mathbf{X}_\phi)^{-1} \mathbf{X}'_\phi \{(\mathbf{D}_\mu^{-1}(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})' \circ \mathbb{I}) \mathbb{J}\}, \quad (5)$$

see Appendix A for details on calculations. In this case, for positive values, the dispersion parameter ( $\phi$ ) is always greater than zero, with  $\phi < 1$  resulting in underdispersion and  $\phi > 1$  in overdispersion.

It is not as straightforward to obtain the maximum likelihood estimator for  $\boldsymbol{\beta}$ , because  $\boldsymbol{\mu}$  is also a function of  $\Sigma$  in the likelihood. Therefore we employ an iterative procedure, using estimates of the regression coefficients using the known estimator for the regression parameters in linear models as starting values, and then updating  $\hat{\boldsymbol{\beta}}$  using  $\hat{\boldsymbol{\mu}}$  estimated in the preceding stage. The maximum likelihood estimator (MLE) of  $\boldsymbol{\beta}$ , conditional on  $\boldsymbol{\mu}, \phi, \boldsymbol{\rho}$  is given by (Laird and Ware, 1982; Molenberghs and Verbeke, 2000)

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\mu}, \phi, \boldsymbol{\rho}) = (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' \Sigma^{-1} \mathbf{Y}.$$

Because we are including different link functions, not limited to the identity, we include a matrix of weights  $\mathbf{W}$  in the estimation of  $\boldsymbol{\beta}$ , following the estimation procedure of the GLM framework. The estimation for the vector  $\boldsymbol{\beta}$  is given considering

$$\boldsymbol{\eta} = \mathbf{g}(\boldsymbol{\mu}), \quad \mathbf{W} = \frac{1}{\Sigma[\mathbf{g}'(\boldsymbol{\mu})]^2} \quad \text{and} \quad \mathbf{z} = \boldsymbol{\eta} + \mathbf{G}(\mathbf{y} - \boldsymbol{\mu}), \quad (6)$$

where  $\mathbf{W}$  is a diagonal matrix of weights,  $\Sigma = \mathbf{D}_\phi \mathbf{D}_\mu$  and  $\mathbf{G} = \text{diag}\{g'(\mu_1), \dots, g'(\mu_n)\}$  (Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989).

Using  $\boldsymbol{\eta}^{(1)} = \mathbf{g}(\mathbf{y})$  for the first iteration, then

$$\hat{\boldsymbol{\beta}}^{(1)} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\eta}^{(1)} \quad \text{and} \quad \hat{\boldsymbol{\beta}}^{(n)} = (\mathbf{X}'\mathbf{W}^{(n-1)}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{(n-1)}\mathbf{z}^{(n-1)}. \quad (7)$$

This algorithm yields the same results as numerically maximising the likelihood function using, e.g. the BFGS algorithm, however it is much faster in terms of computational burden. We will show that these estimators are unbiased in the simulation studies in Section 6.

#### 4.2 Bivariate modelling

To model both outcomes jointly it is necessary to estimate the covariance matrix  $\boldsymbol{\Sigma}_{12}$  (eq. 1). Martinez and Benedito (2013) introduced a method to construct the multivariate covariance matrix using Kronecker products. Bonat and Jørgensen (2016) applied the same method to develop the technique called Multivariate Covariance Generalized Linear Models. Let  $\mathbf{Y}_{n(2 \times 1)} = \{\mathbf{Y}_1, \mathbf{Y}_2\}$ , where the response variables are stacked, with  $n$  = number of observations and  $2$  = number of outcomes;  $\boldsymbol{\Sigma}_i$  the  $n \times n$  covariance matrix within the outcome  $i$  for  $i = 1, 2$ . Let  $\boldsymbol{\Sigma}_b$  be the  $R \times R$  correlation matrix between outcomes, with the diagonal elements equal to 1 and off-diagonal elements equal to  $\rho$ . Then

$$\mathbf{E}(\mathbf{Y}) = \{g^{-1}(\mathbf{X}_1\boldsymbol{\beta}_1), g^{-1}(\mathbf{X}_2\boldsymbol{\beta}_2)\}$$

$$\text{Var}(\mathbf{Y}) = \boldsymbol{\Sigma} = \text{Bdiag}(\tilde{\boldsymbol{\Sigma}}_1, \tilde{\boldsymbol{\Sigma}}_2)(\boldsymbol{\Sigma}_b \otimes \mathbf{I})\text{Bdiag}(\tilde{\boldsymbol{\Sigma}}_1', \tilde{\boldsymbol{\Sigma}}_2') \quad (8)$$

where the matrix  $\tilde{\boldsymbol{\Sigma}}_i$  denotes the lower triangular matrix of the Cholesky decomposition of  $\boldsymbol{\Sigma}_i$ . The operator Bdiag denotes a block diagonal matrix and  $\mathbf{I}$  denotes an  $n \times n$  identity matrix. Then, the covariance matrix, presented in the (eq. 1)., will be defined as

$$\boldsymbol{\Sigma}_{12} = \rho \tilde{\boldsymbol{\Sigma}}_1 \tilde{\boldsymbol{\Sigma}}_2^T.$$

Here we propose an extension to allow more than one correlation coefficient, for example, we could have one for each treatment in our case study data (Section 2). This requires increasing the flexibility of the  $(\boldsymbol{\Sigma}_b \otimes \mathbf{I})$  matrix to include more than one correlation coefficient. Let  $\mathbf{X}_\rho$  denote a  $N \times t$  design matrix, with  $t$  the number of treatment levels, and  $\boldsymbol{\rho}$  a vector for

the correlation coefficients with  $t \times 1$  dimension, then writing  $\Sigma_\rho = \text{diag}(\mathbf{X}_\rho \boldsymbol{\rho})$  and replacing  $(\Sigma_b \otimes \mathbf{I})$  in eq. (8), we have

$$\text{Var}(\mathbf{Y}) = \Sigma = \text{Bdiag}(\tilde{\Sigma}_1, \tilde{\Sigma}_2) \begin{bmatrix} \mathbf{I} & \Sigma_\rho \\ \Sigma_\rho & \mathbf{I} \end{bmatrix} \text{Bdiag}(\tilde{\Sigma}_1', \tilde{\Sigma}_2').$$

For  $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2]$ , the full log likelihood is

$$l(\boldsymbol{\mu}, \Sigma(\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\rho}) | \mathbf{y}) = -\frac{1}{2} \{ \log |\Sigma| + (\mathbf{y} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}) \} + \text{constant}. \quad (9)$$

The covariance matrix is more flexible, allowing the modelling of data sets that have subgroups with different correlation profiles. A likelihood ratio test can be used to test if correlations coefficients differ, for example by levels of a treatment in an experiment.

Jørgensen and Knudsen (2004) suggested a method to jointly estimate all parameters of the  $\Sigma$  matrix using the Newton scoring algorithm, based on quasi-likelihood and the Pearson estimating functions, applying second-moment assumptions. As an estimation alternative approach, here, the matrices  $\Sigma_1$  and  $\Sigma_2$  are estimated first separately, using the maximum likelihood method, presented in section (4.1). These estimates are then used as initial values to obtain the estimates  $\hat{\boldsymbol{\rho}}$  by profiling the log-likelihood function w.r.t.  $\boldsymbol{\rho}$  and maximizing (9) numerically. At each step, the estimates for  $\boldsymbol{\mu}$  and  $\boldsymbol{\phi}$  are updated, using the estimation algorithm described in Section 4.1. Since the derivatives of  $l$  w.r.t.  $\boldsymbol{\rho}$  cannot be obtained in closed form, the L-BFGS-B algorithm (Zhu et al., 1997) was used. Standard errors for the regression parameters are obtained based on the observed information matrix  $\mathbf{I}(\boldsymbol{\theta})$ , where  $\mathbf{I}(\boldsymbol{\theta}) = -\mathbf{H}(\boldsymbol{\theta})$  (hessian matrix) is calculated by numerical approximation of  $l(\boldsymbol{\theta})$ , using Richardson extrapolation, as implemented in the `hessian()` function of the `numDeriv` package (Gilbert et al., 2006). Since  $\hat{\boldsymbol{\phi}}$  depends on  $\hat{\boldsymbol{\beta}}$ , the standard errors for the dispersion parameters are obtained using a two-stage algorithm. First, considering  $\hat{\boldsymbol{\rho}}$ , we obtain the standard errors for  $\hat{\boldsymbol{\beta}}$ . In the next stage considering  $\hat{\boldsymbol{\rho}}$  and  $\hat{\boldsymbol{\beta}}$  we obtain the standard errors for  $\hat{\boldsymbol{\phi}}$ .

We applied these methods to the case study data described in Section 2. Diagnostic analyses and goodness-of-fit assessment for the models fitted to the case study data set were carried out by producing bivariate residual plots with simulation polygons (Moral et al., 2020). Likelihood-ratio tests were used to compare treatments and to assess the significance of the estimated effects.

## 5. Case study analysis

We fitted the bivariate model to the data set described in Section 2, with the log and identity link functions for the count and continuous outcome respectively, including the effects of diet (treatment) in the linear predictors for the mean, dispersion and correlation. Table 1 shows the estimated parameters and standard errors. The females fed with *D. saccharalis* emerged with greater weight ( $\hat{\beta}_{22}$ ) and had greater fertility, producing more eggs ( $\hat{\beta}_{12}$ ). The estimates  $\hat{\phi}_{11}$  and  $\hat{\phi}_{12}$  indicate overdispersion for the number of eggs, however since their standard errors are large we tested for their equality. The estimates  $\hat{\phi}_{21}$  and  $\hat{\phi}_{22}$  are also similar.

[Table 1 about here.]

The bivariate residual plot with simulation polygons with 99 simulations showed 3 out of 27 points out of their respective polygons, which indicates a reasonably good fit (Figure 2).

[Figure 2 about here.]

The selection of the model was performed using likelihood-ratio tests. We assessed the treatment effect on the mean and the dispersion by each outcome and jointly (Table 2 and 3). The effect of treatment on the correlation was also tested and presented no significance, but here due to the differences in correlations observed per treatment (Figure 1), the biological assumption, and the small sample size, we chose to maintain a different correlation effect per treatment. The *D. saccharalis* presented an estimated value for the correlation of 0.5192 indicating that the insects that received this diet produced more eggs as they gained more

weight, contributing to their development, but for the *A. gemmatalis* diet we did not see this same relationship, since the estimated value for the correlation was -0.0389.

Considering the dispersion parameters, the test gave evidence for equal dispersion coefficients, but different regression coefficients for each diet. The regression parameters showed a difference between treatments for both outcomes, hence, the diet *D. saccharalis* contributed to a better development of the predator.

[Table 2 about here.]

[Table 3 about here.]

## 6. Simulation studies

Simulation studies were performed to assess the properties of the maximum likelihood estimators and the flexibility of the bivariate framework. The scenarios proposed here compare the performance of the proposed model with different sample size; under, over and equidispersion for the count outcome; heterogeneity of variances for the continuous outcome and negative, positive, low and high correlation between outcomes.

We conducted the simulation studies based on the experimental structure described in our case study (Section 2). Four scenarios were designed considering one treatment factor with two levels and the numbers of experimental units equal to with half of the experiments units assigned to each treatment. The count outcome was generated from a normal distribution with a logarithmic link function, fixing the regression coefficients at the values  $\beta_{11} = 2$  and  $\beta_{12} = 3$  (the first index represents the response variable and the second index the treatment level). The dispersion parameters were fixed at the values  $\phi_{11} = 0.5$  and  $\phi_{12} = 2$ . For the second continuous outcome, we considered  $\beta_{21} = 10$ ,  $\beta_{22} = 300$ ,  $\phi_{21} = 1$ ,  $\phi_{22} = 10$  and an identity link function. We produced simulations based on four different values for a single correlation parameter ( $\rho = -0.8, 0.2, 0.5$  and  $0.8$ ). These scenarios present negative, weak,

moderate, and strong positive correlation. To check the consistency of the estimators, four sample sizes were considered: 50, 100, 300, and 600. We generated 1000 data sets for each scenario. The simulation process is summarised as follows

- (1) Define  $\beta_1 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ ;  $\beta_2 = \begin{bmatrix} 10 \\ 300 \end{bmatrix}$ ;  $\phi_1 = \begin{bmatrix} 0.5 \\ 2 \end{bmatrix}$ ;  $\phi_2 = \begin{bmatrix} 1 \\ 10 \end{bmatrix}$ ;
- (2) Set up  $\mathbf{X}_{\beta_1}$  and  $\mathbf{X}_{\beta_2}$  as  $n \times 2$  design matrices;
- (3) Set up  $\mathbf{X}_{\phi_1}$  and  $\mathbf{X}_{\phi_2}$  as  $n \times 2$  dispersion design matrices;
- (4) Calculate  $\mu_1 = g^{-1}(\mathbf{X}_{\beta_1}\beta_1) = \exp(\mathbf{X}_{\beta_1}\beta_1)$  and  $\mu_2 = g^{-1}(\mathbf{X}_{\beta_2}\beta_2) = \mathbf{X}_{\beta_2}\beta_2$ ;
- (5) Obtain  $\Sigma_1 = \text{diag}(\mathbf{X}_{\phi_1}\phi_1)\text{diag}(\mu_1)$  and  $\Sigma_2 = \text{diag}(\mathbf{X}_{\phi_2}\phi_2)\text{diag}(\mu_2)$ ;
- (6) Set up  $\mathbf{X}_\rho$  as an  $n \times 1$  unit matrix;
- (7) Obtain  $\Sigma_\rho = \text{diag}(\mathbf{X}_\rho\rho)$ , where  $\rho$  is a scalar ( $-0.8, 0.2, 0.5$  or  $0.8$ );
- (8) Calculate  $\Sigma = \text{Bdiag}(\Sigma_1, \Sigma_2) \begin{bmatrix} \mathbf{I} & \Sigma_\rho \\ \Sigma_\rho & \mathbf{I} \end{bmatrix} \text{Bdiag}(\Sigma_1^T, \Sigma_2^T)$ ;
- (9) Simulate from  $\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} \sim N \left( \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \Sigma \right)$ ;
- (10) Fit the proposed model to the simulated data sets and evaluate estimates.

This process was followed for each simulation scenario. Further, we also simulated a scenario with two correlations parameters, one for each treatment. The study considered 1,000 simulations, sample sizes equal to 50, 100 and 300, and four scenarios combining correlation coefficients ( $\rho_1 = 0.2$  and  $\rho_2 = 0.5$ ;  $\rho_1 = 0.2$  and  $\rho_2 = 0.8$ ;  $\rho_1 = -0.8$  and  $\rho_2 = 0.2$ ;  $\rho_1 = -0.2$  and  $\rho_2 = 0.5$ ).

The complete code used for the simulations and model fitting can be found in the supplementary materials.

### 6.1 Simulation results

To visualize estimation bias for each simulation scenario Figure 3 was constructed. It presents the average bias, defined as the average estimate of 1,000 simulations minus the true value and their respective confidence interval calculated as the average bias plus and minus 1.96 times the standard error.

[Figure 3 about here.]

The results of the simulation studies showed that for all correlation levels, the average bias and the standard errors tend to zero as the sample size increases, meaning that all parameters have unbiased and consistent estimators. This is also observed for the regression parameters when the correlation is very high ( $\rho = 0.8$ ). For the correlation parameters, the bias is low even for small samples, but it is smaller when the sample size increases. The average standard error is lower for strong correlations, positive or negative ( $\rho = -0.8$  and  $\rho = 0.8$ ); this is clear for the correlation coefficient, but it is also possible to observe a slightly smaller average standard error for the dispersion coefficients.

Figure 4 shows the empirical coverage rate of the asymptotic confidence intervals. The results showed that for the regression parameters the empirical coverage rates are close to the nominal level of 95% for all sample sizes and all simulation scenarios. For the dispersion and correlation parameters, the empirical coverage rates are lower than the nominal level, however, for  $\rho = 0.2$  and  $\rho = 0.5$ , they become closer to the nominal level for large samples. The worst scenario is with a small sample and strong correlation, in this case, even for large samples the coverage rate is still lower than the nominal level. An alternative here is to consider other methods to construct the confidence intervals such as the bootstrap approach.

[Figure 4 about here.]

## 6.2 Bootstrap approach

To verify the coverage rate considering bootstrap confidence intervals, a parametric bootstrap simulation was performed, considering 1,000 bootstrap samples and 100 simulations, following the same previous scenarios with a sample size equal to 50 and correlation coefficients equal to  $-0.8$  and  $0.8$ . These simulations showed some asymmetry and bias, mainly for the dispersion and correlation coefficients, therefore bias-corrected percentile confidence intervals (BCP) were considered. Efron (1982) presented this method, in which the correction relies upon transforming the distribution to one that satisfies the symmetry condition (Diciccio and Romano, 1988; Buckland, 1984). Using this method, the empirical coverage rates increased, becoming very close to the 95% nominal coverage rate (Table 4).

[Table 4 about here.]

Table 5 shows the results of the simulation study considering two correlation coefficients, one for each treatment. The observed average bias was low, and the larger the sample size, the lower the average bias and associated standard errors. We observed that for stronger correlations, the standard error was lower, implying empirical confidence intervals with smaller coverage than the nominal rate. Further, it is clear that the scenarios that presented better coverage, mainly for larger samples, were scenarios 1 and 4.

[Table 5 about here.]

## 7. Discussion

In this work, we proposed a bivariate framework to jointly model count and continuous responses. Our approach represents an attractive and flexible framework for studying two variables jointly and is widely applicable. This alternative considered a bivariate normal distribution, modelling the variance-covariance matrix so that the characteristics of the data, such as over- or underdispersion for count data, and heterogeneity of variances for contin-



uous data can be modelled. As expected, the normal approximation performed better with larger sample sizes. The advantage of this methodology is that it is possible to incorporate correlation, to jointly model the variable responses and to assume an associated probability distribution. One drawback of our approach is that issues may arise for small sample sizes, including biased dispersion estimates and a true coverage rate of the confidence intervals for the parameters smaller than the nominal rate. So it is necessary to carefully analyse very small samples, doing descriptive analyses before fitting the model, and checking if the obtained results have a practical interpretation.

We observed in the simulation studies that even for large samples, when  $\rho = -0.8$  and  $\rho = 0.8$ , the coverage rate of the correlation parameters is close to 90% instead of 95%. This is possibly because the true correlation value is close to the boundary of the parametric space. When  $\rho \approx 1$ , the range limits are automatically limited by 1 or  $-1$ , then the variation of the parameter space is smaller. This became more evident by observing Figure 3, which showed that the standard errors for the correlation coefficients are lower for higher correlations. This is not reflected in the regression coefficients, due to the fact that  $\hat{\beta}$  varies very little for different values of  $\phi$  and  $\rho$ . However, the coverage rate for the dispersion parameters also depends on the correlation parameter value, and the higher the correlation, the lower the coverage rate. These problems are expected when it comes to estimates at the limit of the parametric space, especially with smaller samples. Despite these limitations, we concluded that the model generally behaves as expected, with unbiased, consistent, and accurate estimates.

Several extensions to the proposed bivariate model are possible (and currently in progress). These extensions include modelling correlation between observations taken on the same experimental unit, e.g., longitudinal data or repeated measurements. In this case, the variance-covariance matrix needs to be extended, allowing to incorporate the dependent measures

taken on the same subject. This extension will be carried out by generalising  $\Sigma = \mathbf{D}_\phi \mathbf{D}_\mu + \mathbf{ZGZ}'$ , where  $\mathbf{Z}$  is the matrix of known grouping factors reflecting correlated observations or how the response evolves overtime for each subject and  $\mathbf{G}$  a general variance-covariance matrix. However, it is also necessary to include a link function for the dispersion parameter, so that it does not yield a negative maximum likelihood estimate.

## References

- Aitkin, M. (1987). Modelling variance heterogeneity in normal regression using glim. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **36**, 332–339.
- Bonat, W., Olivero, J., Grande-Vega, M., Farfán, M., and Fa, J. (2017). Modelling the covariance structure in marginal multivariate count models: hunting in bioko island. *Journal of Agricultural, Biological and Environmental Statistics* **22**, 446–464.
- Bonat, W. H. (2018). Multiple response variables regression models in r: The mcglim package. *Journal of Statistical Software* **84**,.
- Bonat, W. H. and Jørgensen, B. (2016). Multivariate covariance generalized linear models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **65**, 649–675.
- Buckland, S. T. (1984). Monte carlo confidence intervals. *Biometrics* pages 811–817.
- Diciccio, T. J. and Romano, J. P. (1988). A review of bootstrap confidence intervals. *Journal of the Royal Statistical Society: Series B (Methodological)* **50**, 338–354.
- Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. SIAM.
- Everitt, B. and Hothorn, T. (2011). *An introduction to applied multivariate analysis with R*. Springer Science & Business Media.
- Gilbert, P., Gilbert, M. P., and Varadhan, R. (2006). The numderiv package.
- Haslett, J., Parnell, A. C., Hinde, J., and de Andrade Moral, R. (2021). Modelling excess zeros in count data: A new perspective on modelling approaches. *International Statistical Review* .

- Hinde, J. and Demétrio, C. G. (1998). Overdispersion: models and estimation. *Computational Statistics and Data Analysis* **27**, 151–170.
- Jobson, J. D. (2012). *Applied multivariate data analysis: regression and experimental design*. Springer Science & Business Media.
- Johnson, N., Kotz, S., and Balakrishnan, N. (1997). *Discrete multivariate distributions*. Wiley, New York.
- Jørgensen, B. and Knudsen, S. J. (2004). Parameter orthogonality and bias adjustment for estimating functions. *Scandinavian Journal of Statistics* **31**, 93–114.
- Krupskii, P. and Joe, H. (2013). Factor copula models for multivariate data. *Journal of Multivariate Analysis* **120**, 85–101.
- Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* pages 963–974.
- Martinez-Beneito, M. A. (2013). A general modelling framework for multivariate disease mapping. *Biometrika* **100**, 539–553.
- McCullagh, P. and Nelder, J. (1989). *Generalized linear models*. Chapman and Hall.
- Molenberghs, G. and Verbeke, G. (2000). *Linear mixed models for longitudinal data*. Springer.
- Mood, A. M. (1973). *Introduction to the Theory of Statistics.*, volume 3. McGraw-hill.
- Moral, R. A., Hinde, J., and Demétrio, C. G. (2020). Bivariate residual plots with simulation polygons. *Journal of Computational and Graphical Statistics* **29**, 203–214.
- Nelder, J. A. and Pregibon, D. (1987). An extended quasi-likelihood function. *Biometrika* **74**, 221–232.
- Nelder, J. A. and Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)* **135**, 370–384.
- Nikoloulopoulos, A. K. and Karlis, D. (2009). Modeling multivariate count data using

- copulas. *Communications in Statistics-Simulation and Computation* **39**, 172–187.
- Parra, J. R. P. (2002). *Controle biológico no Brasil: parasitóides e predadores*. Editora Manole Ltda.
- Raykov, T. and Marcoulides, G. A. (2008). *An introduction to applied multivariate analysis*. Routledge.
- Team, R. C. (2022). R: A language and environment for statistical computing. r foundation for statistical computing.
- Ver Hoef, J. M. and Boveng, P. L. (2007). Quasi-poisson vs. negative binomial regression: how should we model overdispersed count data. *Ecology* **88**, 2766–2772.
- Zhu, C., Byrd, R. H., Lu, P., and Nocedal, J. (1997). Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software (TOMS)* **23**, 550–560.

## Appendix A: Estimation and Inference for Section 3

Taking one of the response outcomes  $\mathbf{Y}_i$ , from here we will omit the index  $i$  to improve readability. Asymptotically  $\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma} = \mathbf{D}_\phi \mathbf{D}_\mu)$ , we can write the log-likelihood as

$$\log \mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{y}) = l(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{y}) = -\frac{1}{2} \{ \log |\boldsymbol{\Sigma}| + (\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \} + \text{const.} \quad (\text{A.1})$$

Differentiating with respect to  $\boldsymbol{\Sigma}$  is straightforward:

$$\frac{\partial l}{\partial \boldsymbol{\Sigma}} = \frac{\partial l}{\partial \boldsymbol{\Sigma}} \left\{ -\frac{1}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right\} \quad (\text{A.2})$$

as  $\frac{\partial}{\partial A} \log |A| = A'$  and  $\frac{\partial \mathbf{a}' \mathbf{X}^{-1} \mathbf{b}}{\partial \mathbf{X}} = -(\mathbf{X}^{-1})' (\mathbf{a} \mathbf{b})' (\mathbf{X}^{-1})'$ . Then,

$$\frac{\partial l}{\partial \boldsymbol{\Sigma}} = -\frac{1}{2} (\boldsymbol{\Sigma}^{-1})' + \frac{1}{2} (\boldsymbol{\Sigma}^{-1})' (\mathbf{y} - \boldsymbol{\mu}) (\mathbf{y} - \boldsymbol{\mu})' (\boldsymbol{\Sigma}^{-1})' \quad (\text{A.3})$$

and

$$\begin{aligned}
\frac{\partial l}{\partial \Sigma} \frac{\partial \Sigma}{\partial \mathbf{D}_\phi} &= \left[ -\frac{1}{2}(\Sigma^{-1})' + \frac{1}{2}(\Sigma^{-1})'(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'(\Sigma^{-1})' \right] \circ \mathbf{D}_\mu \\
&= \left[ -\frac{1}{2}((\mathbf{D}_\phi \mathbf{D}_\mu)^{-1})' + \frac{1}{2}((\mathbf{D}_\phi \mathbf{D}_\mu)^{-1})'(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'((\mathbf{D}_\phi \mathbf{D}_\mu)^{-1})' \right] \circ \mathbf{D}_\mu \\
&= -\frac{1}{2}((\mathbf{D}_\phi \mathbf{D}_\mu)^{-1})' \circ \mathbf{D}_\mu + \frac{1}{2}((\mathbf{D}_\phi \mathbf{D}_\mu)^{-1})'(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'((\mathbf{D}_\phi \mathbf{D}_\mu)^{-1})' \circ \mathbf{D}_\mu \\
&= -\frac{1}{2}(\mathbf{D}_\phi^{-1})'(\mathbf{D}_\mu^{-1})' \circ \mathbf{D}_\mu + \frac{1}{2}(\mathbf{D}_\phi^{-1})'(\mathbf{D}_\mu^{-1})'(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'(\mathbf{D}_\phi^{-1})'(\mathbf{D}_\mu^{-1})' \circ \mathbf{D}_\mu \\
&= -\frac{1}{2}(\mathbf{D}_\phi^{-1})' + \frac{1}{2}(\mathbf{D}_\phi^{-1})(\mathbf{D}_\mu^{-1})'(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'(\mathbf{D}_\phi^{-1})'.
\end{aligned}$$

where  $\mathbb{I}$  is an  $n \times n$  identity matrix and  $\circ$  the Hadamard product. If  $\mathbf{A}$  is diagonal, then  $\mathbf{A}' = \mathbf{A}$  and

$$\begin{aligned}
\frac{\partial l}{\partial \Sigma} \frac{\partial \Sigma}{\partial \mathbf{D}_\phi} \frac{\partial \mathbf{D}_\phi}{\partial \phi} &= \left[ -\frac{1}{2}\mathbf{D}_\phi^{-1} + \frac{1}{2}\mathbf{D}_\phi^{-1}\mathbf{D}_\mu^{-1}(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'\mathbf{D}_\phi^{-1} \right] \circ \mathbb{I} \\
&= -\frac{1}{2}\text{diag}(\mathbf{X}_\phi \hat{\phi})^{-1} \circ \mathbb{I} + \frac{1}{2}\text{diag}(\mathbf{X}_\phi \hat{\phi})^{-1}\mathbf{D}_\mu^{-1}(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'\text{diag}(\mathbf{X}_\phi \hat{\phi})^{-1} \circ \mathbb{I}. \quad (\text{A.4})
\end{aligned}$$

Equating (A.4) to zero yields

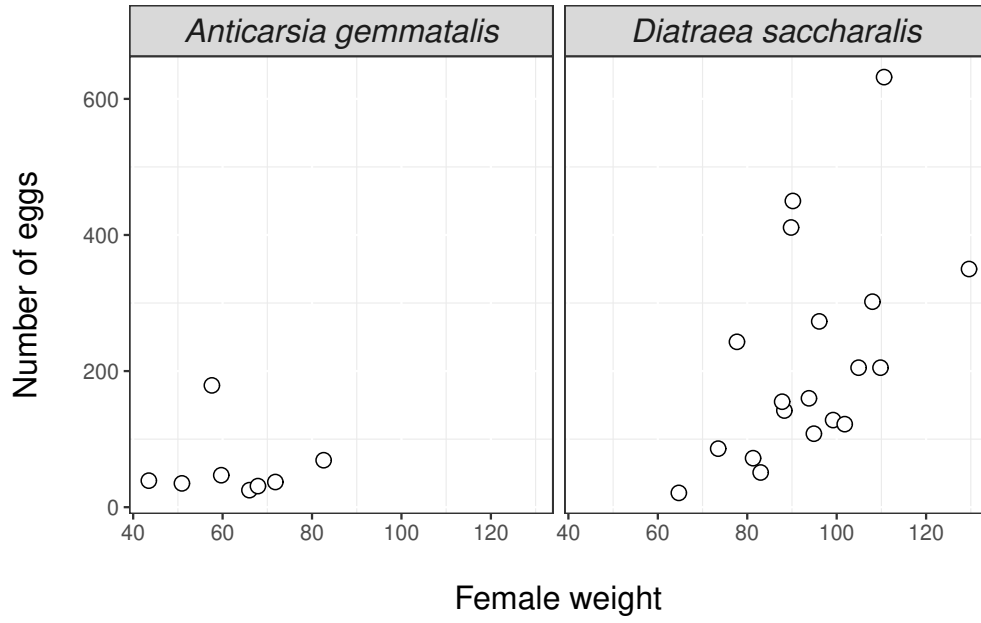
$$\begin{aligned}
&-\frac{1}{2}\text{diag}(\mathbf{X}_\phi \hat{\phi})^{-1} \circ \mathbb{I} + \frac{1}{2}\text{diag}(\mathbf{X}_\phi \hat{\phi})^{-1}\mathbf{D}_\mu^{-1}(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'\text{diag}(\mathbf{X}_\phi \hat{\phi})^{-1} \circ \mathbb{I} = 0 \\
&\Rightarrow \text{diag}(\mathbf{X}_\phi \hat{\phi})^{-1} \circ \mathbb{I} = \text{diag}(\mathbf{X}_\phi \hat{\phi})^{-1}\mathbf{D}_\mu^{-1}(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'\text{diag}(\mathbf{X}_\phi \hat{\phi})^{-1} \circ \mathbb{I} \\
&\Rightarrow \text{diag}(\mathbf{X}_\phi \hat{\phi}) = \mathbf{D}_\mu^{-1}(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})' \circ \mathbb{I}.
\end{aligned}$$

Post-multiplying both sides by  $\mathbb{J}_{n \times 1}$ , where  $\mathbb{J}$  is an  $n \times 1$  vector of 1s yields

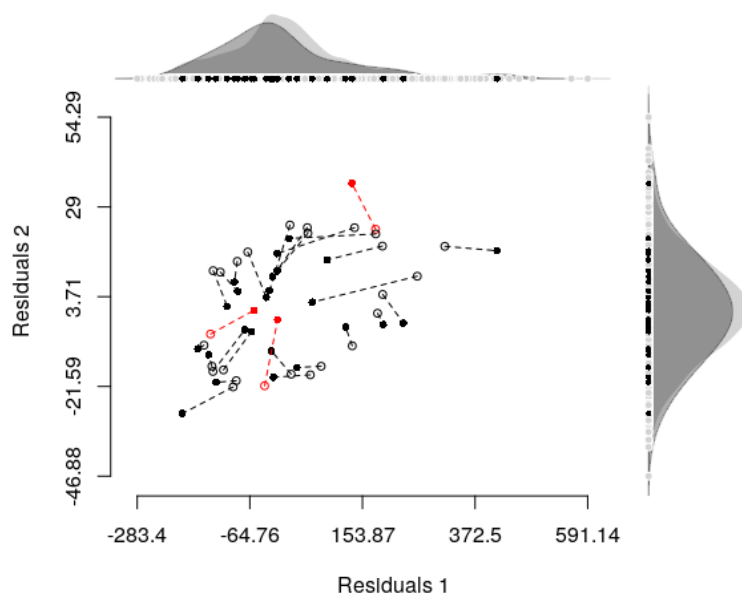
$$\mathbf{X}_\phi \hat{\boldsymbol{\phi}} = (\mathbf{D}_\mu^{-1}(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})' \circ \mathbb{I})\mathbb{J} \quad (\text{A.5})$$

Finally, pre-multiplying both sides by  $(\mathbf{X}'_\phi \mathbf{X}_\phi)^{-1} \mathbf{X}'_\phi$  yields

$$\hat{\boldsymbol{\phi}} = (\mathbf{X}'_\phi \mathbf{X}_\phi)^{-1} \mathbf{X}'_\phi \{(\mathbf{D}_\mu^{-1}(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})' \circ \mathbb{I})\mathbb{J}\} \quad (\text{A.6})$$

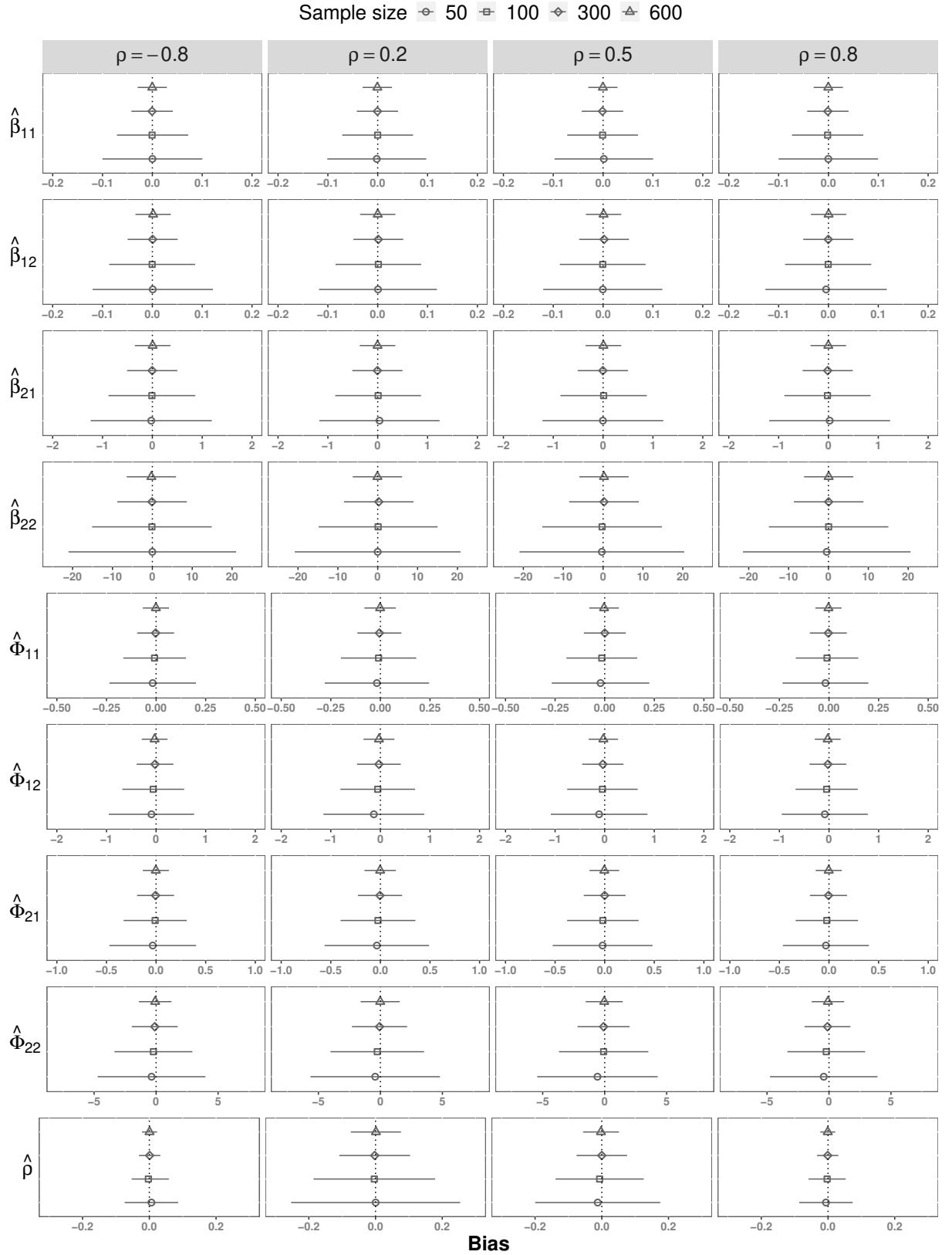


**Figure 1:** *Podisus nigrispinus* female weight after 18 days and total number of eggs laid throughout the insect's life cycle by treatment (diet consisting of *Anticarsia gemmatalis* or *Diatraea saccharalis*).

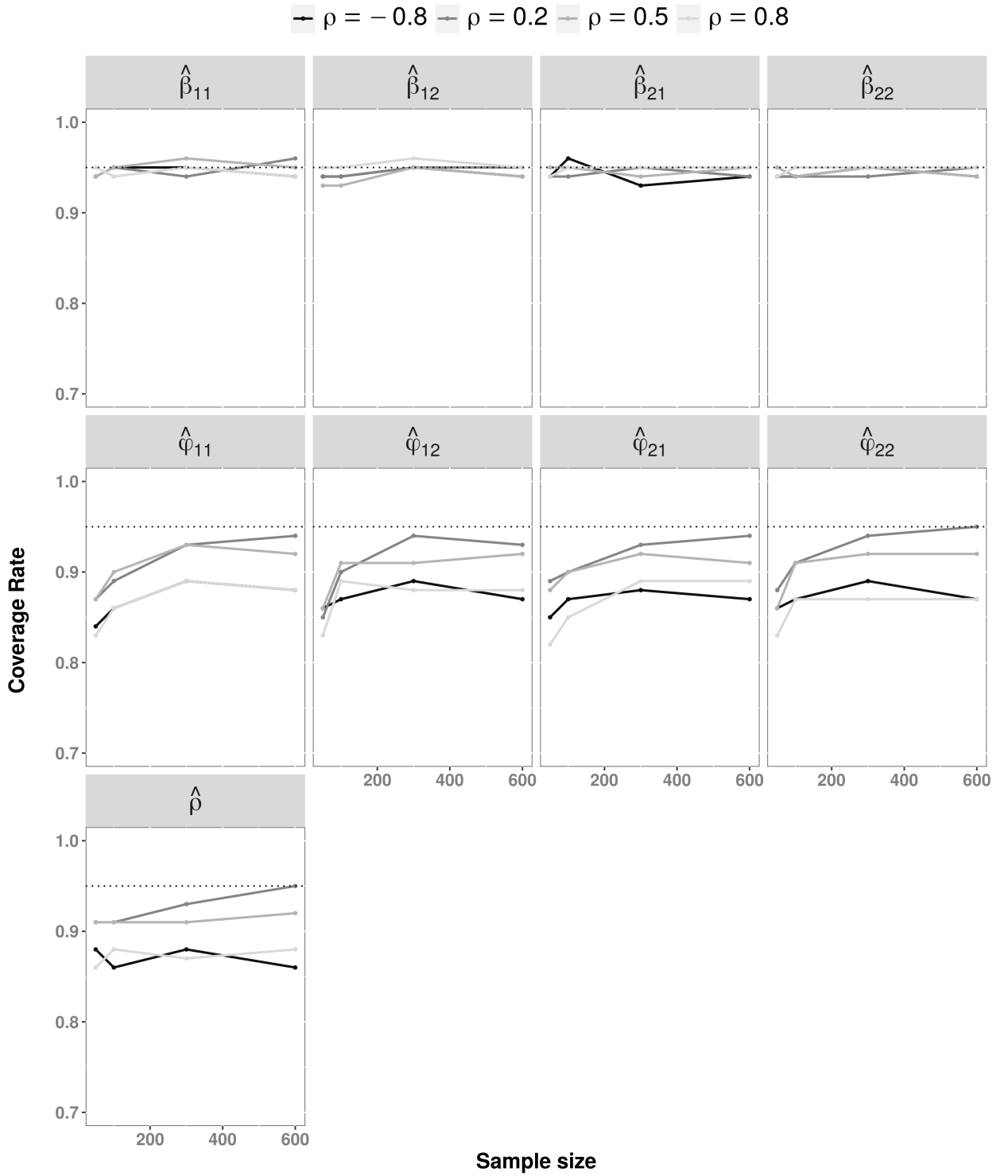


**Figure 2:** Bivariate plot with simulation polygons assuming a bivariate normal distribution for raw bivariate residuals, with estimated marginal densities at the top and right-hand side of the plot. The lighter shade corresponds to the estimated density of all simulated diagnostics, and the darker one corresponds to the estimated density of the observed diagnostics. The points outside their simulated polygons are displayed in red.





**Figure 3:** Estimation average bias of each simulation scenario (symbols) with their respective 95% confidence intervals, considering 1,000 simulations.



**Figure 4:** Coverage rate based on normal approximation confidence intervals with a nominal rate of 95% for different sample sizes (50, 100, 300, and 600) and correlation coefficients  $\rho \in \{-0.8, 0.2, 0.5, 0.8\}$ .

Table 1: Parameter estimates and standard errors (S.E.) for the normal bivariate model.  $\beta_{11}$  and  $\beta_{12}$  represent the mean number of eggs produced by stinkbugs on for diets *Anticarsia gemmatalis* and *Diatraea saccharalis* respectively;  $\beta_{21}$  and  $\beta_{22}$  represent the mean weight for stinkbugs on diets *A. gemmatalis* and *D. saccharalis* respectively.  $\phi_{11}$  and  $\phi_{12}$  represent the dispersion of the number of eggs for diets *A. gemmatalis* and *D. saccharalis* respectively;  $\phi_{21}$  and  $\phi_{22}$  represent the dispersion of weight for diets *A. gemmatalis* and *D. saccharalis* respectively.  $\rho_1$  represents the correlation between number of eggs and weight for *A. gemmatalis* diet and the parameter  $\rho_2$  represents the correlation between number of eggs and weight for *D. saccharalis*.

Parameter	Estimate	S.E.
$\beta_{11}$	4.0561	0.2907
$\beta_{12}$	5.3782	0.1609
$\beta_{21}$	62.5000	4.0684
$\beta_{22}$	93.9474	3.3824
$\phi_{11}$	39.0292	19.5049
$\phi_{12}$	106.6240	32.1814
$\phi_{21}$	2.1186	1.0588
$\phi_{22}$	2.3138	0.6983
$\rho_1$	-0.0389	0.3528
$\rho_2$	0.5192	0.1487
Log-Likelihood		-221.0125

Table 2: Model fit measurements and comparisons between the complete and reduced models to test the treatment effect on the mean and on the dispersion for each outcome.

Model	np	$l$	$2(\text{diff } l)$	$P(> \chi^2)$
Full model	10	-221.0125		
$\beta_{11} = \beta_{12}$	9	-227.4322	12.8394	0.0003
$\beta_{21} = \beta_{22}$	9	-236.2779	16.4631	< 0.0001
$\phi_{10} = \phi_{11}$	9	-222.2381	2.4511	0.2936
$\phi_{20} = \phi_{21}$	9	-221.0235	0.02209	0.9890

<sup>1</sup>np, number of parameters;  $l$ , log-likelihood; diff  $l$ , difference in log-likelihoods.

Table 3: Model fit measurements and comparisons between the complete and reduced models to test the treatment effect on the mean and on the dispersion for both outcomes jointly.

Model	np	$l$	$2(\text{diff } l)$	$P(> \chi^2)$
Full model	10	-221.0125		
$\beta_{11} = \beta_{12}$ and $\beta_{21} = \beta_{22}$	8	-232.1660	22.3070	< 0.0001
$\phi_{10} = \phi_{11}$ and $\phi_{20} = \phi_{21}$	8	-222.2347	2.4445	0.2946

<sup>2</sup>np, number of parameters;  $l$ , log-likelihood; diff  $l$ , difference in log-likelihoods.

Table 4: Coverage rate based on bias-corrected percentile confidence intervals with 1,000 bootstrap samples and 100 simulations, and a nominal rate of 95%, for a sample size of  $n = 50$  and true values for the parameters as described in Section 6.

Parameter	Coverage rate	
	$\rho = 0.8$	$\rho = -0.8$
$\beta_{11}$	0.95	0.96
$\beta_{12}$	0.95	0.93
$\beta_{21}$	0.98	0.96
$\beta_{22}$	0.98	0.92
$\phi_{11}$	0.94	0.95
$\phi_{12}$	0.92	0.88
$\phi_{21}$	0.93	0.95
$\phi_{22}$	0.91	0.89
$\rho$	0.95	0.96

Table 5: Estimator mean bias, mean squared error (MSE) and coverage rate for the proposed model under four scenarios, which varied true parameter values. Results are based on 1,000 simulations.

Scenario 1:										
Parameter	True	Bias			MSE			Coverage		
		$n = 50$	$n = 100$	$n = 300$	$n = 50$	$n = 100$	$n = 300$	$n = 50$	$n = 100$	$n = 300$
$\beta_{11}$	2	0.000	-0.004	0.000	0.050	0.036	0.021	0.94	0.94	0.94
$\beta_{12}$	3	-0.002	-0.001	0.000	0.061	0.044	0.026	0.94	0.93	0.95
$\beta_{21}$	10	-0.012	-0.018	-0.017	0.611	0.441	0.257	0.93	0.94	0.95
$\beta_{22}$	300	0.154	-0.059	-0.024	10.584	7.669	4.446	0.93	0.95	0.96
$\phi_{11}$	0.5	-0.020	-0.013	-0.004	0.133	0.096	0.057	0.88	0.90	0.92
$\phi_{12}$	2	-0.093	-0.054	-0.021	0.501	0.362	0.213	0.87	0.90	0.91
$\phi_{21}$	1	-0.044	-0.015	-0.003	0.265	0.194	0.114	0.87	0.90	0.93
$\phi_{22}$	10	-0.469	-0.080	0.083	2.501	1.844	1.069	0.87	0.90	0.91
$\rho_1$	0.2	-0.005	-0.004	-0.001	0.179	0.130	0.076	0.88	0.92	0.93
$\rho_2$	0.5	-0.013	-0.001	0.000	0.133	0.094	0.055	0.88	0.89	0.91
Scenario 2:										
Parameter	True	Bias			MSE			Coverage		
		$n = 50$	$n = 100$	$n = 300$	$n = 50$	$n = 100$	$n = 300$	$n=50$	$n=100$	$n=300$
$\beta_{11}$	2	-0.002	-0.001	0.001	0.051	0.036	0.021	0.91	0.95	0.94
$\beta_{12}$	3	-0.002	0.001	0.001	0.062	0.044	0.026	0.92	0.94	0.94
$\beta_{21}$	10	0.021	0.006	-0.004	0.614	0.440	0.256	0.94	0.94	0.94
$\beta_{22}$	300	-0.093	0.207	0.078	10.682	7.599	4.435	0.92	0.94	0.95
$\phi_{11}$	0.5	-0.015	-0.014	-0.005	0.134	0.096	0.057	0.89	0.89	0.93
$\phi_{12}$	2	-0.059	-0.048	-0.031	0.450	0.321	0.187	0.86	0.87	0.87
$\phi_{21}$	1	-0.036	-0.021	-0.012	0.267	0.193	0.113	0.89	0.91	0.93
$\phi_{22}$	10	-0.283	-0.281	-0.130	2.252	1.597	0.939	0.86	0.86	0.88
$\rho_1$	0.2	-0.011	0.003	0.000	0.180	0.129	0.076	0.90	0.91	0.95
$\rho_2$	0.8	-0.006	-0.002	-0.002	0.058	0.040	0.023	0.83	0.86	0.87
Scenario 3:										
Parameter	True	Bias			MSE			Coverage		
		$n = 50$	$n = 100$	$n = 300$	$n = 50$	$n = 100$	$n = 300$	$n = 50$	$n = 100$	$n = 300$
$\beta_{11}$	2	-0.001	0.000	0.000	0.051	0.036	0.021	0.94	0.93	0.93
$\beta_{12}$	3	0.001	0.001	0.001	0.061	0.044	0.026	0.92	0.94	0.95
$\beta_{21}$	10	-0.013	-0.025	-0.007	0.615	0.442	0.257	0.92	0.93	0.95
$\beta_{22}$	300	-0.149	-0.321	0.245	10.699	7.599	4.444	0.94	0.93	0.94
$\phi_{11}$	0.5	-0.014	-0.014	-0.001	0.112	0.080	0.047	0.83	0.85	0.89
$\phi_{12}$	2	-0.109	-0.050	-0.029	0.524	0.384	0.225	0.86	0.91	0.93
$\phi_{21}$	1	-0.030	-0.010	-0.002	0.225	0.163	0.095	0.84	0.86	0.89
$\phi_{22}$	10	-0.239	-0.268	-0.100	2.701	1.918	1.130	0.88	0.91	0.94
$\rho_1$	-0.8	0.005	0.007	0.002	0.057	0.041	0.023	0.85	0.84	0.85
$\rho_2$	0.2	0.004	-0.003	-0.004	0.178	0.130	0.076	0.88	0.93	0.94
Scenario 4:										
Parameter	True	Bias			MSE			Coverage		
		$n = 50$	$n = 100$	$n = 300$	$n = 50$	$n = 100$	$n = 300$	$n = 50$	$n = 100$	$n = 300$
$\beta_{11}$	2	-0.003	0.001	-0.001	0.051	0.036	0.021	0.94	0.94	0.94
$\beta_{12}$	3	-0.001	0.002	0.001	0.061	0.044	0.025	0.94	0.94	0.95
$\beta_{21}$	10	-0.041	-0.013	-0.002	0.612	0.441	0.256	0.93	0.96	0.94
$\beta_{22}$	300	-0.137	-0.179	0.361	10.583	7.597	4.450	0.94	0.95	0.95
$\phi_{11}$	0.5	-0.017	-0.010	-0.004	0.134	0.097	0.057	0.88	0.91	0.92
$\phi_{12}$	2	-0.085	-0.062	-0.035	0.503	0.361	0.212	0.88	0.90	0.91
$\phi_{21}$	1	-0.035	-0.014	-0.009	0.267	0.194	0.113	0.88	0.92	0.91
$\phi_{22}$	10	-0.448	-0.264	-0.071	2.506	1.815	1.072	0.86	0.88	0.91
$\rho_1$	-0.2	0.004	0.000	0.002	0.179	0.130	0.076	0.88	0.94	0.93
$\rho_2$	0.5	-0.015	-0.008	-0.005	0.134	0.095	0.055	0.88	0.90	0.90