A mean-dispersion bivariate framework to model counts and continuous responses

Maíra 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

Departament 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, including entomology. Generally,

the interest is in more than one response variable and in how factors and covariates affect them, simultaneously. However, when the interest 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 count response were observed simultaneously, a mean-dispersion bivariate model was developed. The proposed model is based on the bivariate normal distribution, 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, whilst allowing for modelling different magnitudes of correlation between the two responses.

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

1. Introduction

In many observational or experimental studies, there is a researcher interest in measuring more than one response or characteristic. To extract information from these measurements, besides how covariates or factors of interest impact them, statistical modelling can be used. In modelling techniques, the response variables can be analyzed in isolation using univariate methods where the dependence between them is not considered in the modelling. However, the correlation between the response variables can bring important information about the study and contribute significantly to the tests and inferences performed, especially when this correction is of high magnitude. The multivariate analysis techniques allow us deal with this problem. They are useful to make inferences multiple variables, and for exploring data patterns that may exist in one or more dimensions (Raykov and Marcoulides, 2008; Everitt and Hothorn, 2011).

When it comes to univariate analysis, many inferences are based on the normal distribution, and once 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 continous 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 provides 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 exponential family distributions is the *mean-variance* relationship, i.e., the fact that the variance is a function of the mean.

Johnson (1997) presented a range of discrete multivariate models that bring in details on how to work with 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 or bounded continuous variables and discrete variables). They created the multivariate covariance generalized linear modelling framework (McGLM), implemented as `mcglm` package for `R software` (Team, 2020). In their approach, marginal models are fitted through quasi-likelihood functions based on first and second-moment assumptions (Bonat et al., 2017; Bonat, 2018). Another possibility for modelling discrete and continuous multivariate response variables jointly are Copula models (Nikoloulopoulos and Karlis, 2009; Krupskii and Joe, 2013). The framework McGLM is easy to fit, allows us to have freedom of choice for the covariance structure, besides it is possible to test the fixed effects present in the model. However, this framework does not have associated probability distribution, that is a drawback. The copula model allows us to incorporate the correlation between the variables and is also flexible for choosing the correlation structure. However, the use of different copulas generates completely different results. In this paper, we present another option for modelling jointly count and continuous data using an approximation based on the central limit theorem and a one-parameter extension of the variance-covariance structure.

When analyzing count data, the Poisson model is a natural first choice, but as it follows from the Central Limit Theorem, the Poisson distribution can be approximated asymptotically by the normal distribution (Mood, 1973). To model through normal distribution makes it becomes easier to work with closed formulas, especially in the presence of more than one response variable. A drawback for the Poisson model is that the impose equality of mean and variance, which doesn't happen very often in count data. This extra or 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-variance to obtain reliable inference about the parameters (Hinde et al., 1998; Ver Hoef and Boveng, 2007). These causes of variation can also impact continuous data. However, the normal distribution, often used to fit to this data, assumes homoscedastic variance, which could not reflect 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 estimations. This proposal brought a powerful approach to deal with the heterogeneity of variance, which was modeled rather than being transformed away. McCullagh and Nelder (1989) presented the joint modelling of mean and dispersion in the generalized linear models, however 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, to accommodate two counts, or two continuous response variables, or a combination of one count and one continuous variable using the bivariate normal approximation. To obtain a versatile framework to deal with count data, we included an extra parameter in the variance-covariance matrix to capture different characteristics present in this type of data, making it very useful also for continuous data. The normal bivariate model proposed allows us to model jointly mean and dispersion dealing with over- and underdispersion, and heteroscedastic variance. Furthermore, it is possible to fit the model considering different correlations that can change depending on a covariate of interest. To exemplify the method developed, a case study applied to the area of entomology was used.

This paper is organized as follows. Section 2 presents the case-study that is a motivation for this paper. Section 3 presents the mean-dispersion bivariate framework to model counts and continuous responses. Section 4 presents estimation and inference for the novel regression

model based on the likelihood paradigm. The properties of the maximum likelihood and profile likelihood estimators are assessed in Section 5 through simulation studies. The results of the simulations and the application of the new mean-dispersion bivariate framework to model count and continuos response is presented in Section 6.

2. Case study

The data is from an experiment examining the development of *Podisus nigrispinus* (Hemiptera: Pentatomidae), an important predator, found in agricultural and forest systems in several countries of Central and South America, when fed two different prey *(Anticarsia gemmatalis* and *Diatraea saccharalis).*

This stinkbug has an important role as a control agent biological for different cultures, for example, in the forest area, the predator can be useful in the biological control of defoliating caterpillars in Eucalyptus plantations. Given this importance, experiments aimed at assessing the dynamics of prey consumption by the *Podisus nigrispinus* have been developed. In these studies, the response variables can be the development, and the reproduction of the predator, being this information, essential for the choice of biological control strategies in pest management programs (Parra, 2002). The development can be measured by several characteristics, including the weight of the insect, being this the main goal of this study, besides their fertility under the effect of two different diets.

The experiment was carried out at the Laboratory of Forest Ecology and Entomology, Department of Entomology and Acarology, ESALQ - USP. The insects were placed in an incubator chamber and individualized in recipients, where were fed daily. The two diets or treatments consisted of offering three caterpillars of either *Anticarsia gemmatalis* or *Diatraea saccharalis* daily. The experiment started with 50 insects for each diet, however, as there was a high mortality rate, the final measurements were made only with 9 replicates of *Anticarsia gemmatalis* and 18 of *Diatraea saccharalis.* 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 formed couples were kept until the death of the females. All eggs laid by the females were placed in Petri dishes and counted.

The correlation between these two outcomes (female weight and number of eggs) was verified by a scatterplot (Figure 1). They presented some correlation for the treatment *Diatraea saccharalis*, then, could be a good idea used the approach proposed here, to model both outcomes jointly.

[Figure 1 about here.]

3. Joint Modelling of a Count and a Continuous Response

Bivariate models may be applied to analyze bivariate data jointly. However, when the outcomes have different distributions, the useful bivariate normal or others can not capture all information present in these data. In this way, as a methodology to analyze count and continuous outcomes, in this work, we proposed a joint modelling approach based on the multivariate normal distribution.

The marginal distribution structure is the same for either a count or a continuous variable. Both assume a normal distribution, with mean $\boldsymbol{\mu_i}; i = 1, 2$ and standard deviation depending on the mean and a dispersion parameter ($\boldsymbol{\phi_i \mu_i}; i = 1, 2$). Besides, it is also possible to capture the heterogeneity of variances, overdispersion, or underdispersion through the modelling of the ϕ parameter. Here, since the normal distribution belongs to the exponential family, it is possible to model the linear predictor $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ through a monotonic and differentiable link function $\boldsymbol{\eta} = \boldsymbol{g(\mu)}$.

Let $\mathbf{Y_1}$ and $\mathbf{Y_2}$ be vectors of count or continuous random variables. We then assume 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.

$$\begin{bmatrix} \mathbf{Y_1} \\ \mathbf{Y_2} \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma_1} & \boldsymbol{\Sigma_{12}} \\ \boldsymbol{\Sigma_{12}} & \boldsymbol{\Sigma_2} \end{bmatrix} \right) \tag{1}$$

where $\boldsymbol{\Sigma_1}$ denote the covariance matrix within outcome 1; $\boldsymbol{\Sigma_2}$ denote the covariance matrix within outcome 2 and $\boldsymbol{\Sigma_{12}}$ denote the covariance matrix between outcomes, where $\boldsymbol{\Sigma_{12}}$ depends on the correlation parameter ρ, $\boldsymbol{\Sigma_1}$ and $\boldsymbol{\Sigma_2}$.

In this framework we proposed to write $\boldsymbol{\Sigma_i} = \text{diag}(\mathbf{X}_\phi \boldsymbol{\phi})\text{diag}(g^{-1}(\mathbf{X}_\beta \boldsymbol{\beta})) = \mathbf{D}_\phi \mathbf{D}_\mu$, where 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 dispersion and regression parameters respectively; and $\boldsymbol{\phi}$ and $\boldsymbol{\beta}$ unknown parameter vectors. We may write

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

where the variance-covariance matrix ($\boldsymbol{\Sigma_i}$) will capture the mean and dispersion of the data.

4. Estimation and Inference

4.1 *Univariate modelling*

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

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

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

$$\frac{\partial l}{\partial \boldsymbol{\Sigma}} = -\tfrac{1}{2}(\boldsymbol{\Sigma}^{-1})^T + \tfrac{1}{2}(\boldsymbol{\Sigma}^{-1})^T(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'(\boldsymbol{\Sigma}^{-1})^T. \tag{3}$$

Now further diffferentiating eq. (3) w.r.t. $\mathbf{D}_\phi$ and w.r.t. $\boldsymbol{\phi}$ yields (see Appendix A for calculations):

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

Finally, solving eq. (4) w.r.t. $\boldsymbol{\phi}$ yields (see Appendix A for calculations):

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

where $\mathbb{I}$ is an $n \times n$ identity matrix and $\mathbb{J}$ is an $n \times n$ vector of 1s, and $\circ$ the Hadamard product. In this case, for positive values, the dispersion parameter (ϕ) is always greater than zero. If $\phi < 1$ there is underpersion and $\phi > 1$ overdispersion.

It is not as straightforward to obtain the maximum likelihood estimator for $\boldsymbol{\beta}$, because $\boldsymbol{\mu}$ is also a function of $\boldsymbol{\Sigma}$ in the likelihood. Then, the idea was, in the first stage, approximate the estimate of the regression coefficients using the known estimation for the regression parameters in the linear models and then updated $\hat{\boldsymbol{\beta}}$ using $\hat{\boldsymbol{\mu}}$ estimated in the preceding stage. The maximum likelihood estimator (MLE) of $\boldsymbol{\beta}$, conditional on $\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\rho}$, with $\boldsymbol{\Sigma}(\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\rho})$ is given by (Laird and Ware, 1982; Molenberghs and Verbeke, 2000)

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

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

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

where $\mathbf{W}$ is a diagonal matrix of weights, $\mathbf{V}(\boldsymbol{\mu}) = \mathbf{D}_\phi\mathbf{D}_\mu$ and $\mathbf{G} = \text{diag}\{g'(\mu_1), \ldots, g'(\mu_n)\}$ (Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989). Using $\boldsymbol{\eta}^{(1)} = g(\mathbf{y})$ for the first iteration, $\mathbf{X}$ as model matrix, then

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

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. The good behavior of this estimators also can be verified in the simulations studies in Section 6.1.

4.2 *Bivariate modelling*

To model both outcomes jointly it is necessary to estimate the matrix $\mathbf{\Sigma}_{12}$ (1). Martinez and Benedito (2013) introduced a method to construct the multivariate covariance matrix using Kronecker product. Bonat and Jorgersen (2016) applied the same method to develop the technique called Multivariate Covariance Generalized Linear Models. Let $\mathbf{Y}_{NR\times 1} = \{\mathbf{Y}_1,\ldots,\mathbf{Y}_R\}$, with N = number of observations and R = number of outcomes; $\mathbf{\Sigma}_i$ the $N \times N$ covariance matrix within the outcome i for $i = 1,\ldots,R$. Let $\mathbf{\Sigma}_b$ the $R \times R$ correlation matrix between outcomes, with the diagonal elements equal to 1 and off-diagonal elements equal to ρ. Then

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

$$\mathrm{Var}(\mathbf{Y}) = \mathbf{\Sigma} = \mathrm{Bdiag}(\tilde{\mathbf{\Sigma}}_1,\ldots,\tilde{\mathbf{\Sigma}}_R)(\mathbf{\Sigma}_b \otimes \mathbf{I})\mathrm{Bdiag}(\tilde{\mathbf{\Sigma}}_1^T,\ldots,\tilde{\mathbf{\Sigma}}_R^T)$$

where the matrix $\tilde{\mathbf{\Sigma}}_i$ denotes the lower triangular matrix of the Cholesky decomposition of $\mathbf{\Sigma}_i$. The operator Bdiag denotes a block diagonal matrix and $\mathbf{I}$ denotes an $N \times N$ identity matrix. Then, the covariance matrix, could be defined as

$$\mathbf{\Sigma}_{12} = \rho\tilde{\mathbf{\Sigma}}_i\tilde{\mathbf{\Sigma}}_{i'}^T$$

with $i \neq i'$.

As a proposal for this work, it is possible to estimate more than one correlation coefficient, for example, one for each treatment under study. This proposed method will need an extension of the $\mathbf{\Sigma}_b$ matrix, hence, it may incorporate more than one correlation coefficient.

Let $\mathbf{X}_\rho$ denote a $N \times t$ design matrix, with t the to number of treatment levels, and $\boldsymbol{\rho}$ a vector for the correlation coefficients with $t \times 1$ dimension, then

$$\boldsymbol{\Sigma}_\rho = \mathrm{diag}(\mathbf{X}_\rho \boldsymbol{\rho})$$

and

$$\boldsymbol{\Sigma} = \mathrm{Bdiag}(\tilde{\boldsymbol{\Sigma}}_1, \tilde{\boldsymbol{\Sigma}}_2) \begin{bmatrix} \mathbf{I} & \boldsymbol{\Sigma}_\rho \\ \boldsymbol{\Sigma}_\rho & \mathbf{I} \end{bmatrix} \mathrm{Bdiag}(\tilde{\boldsymbol{\Sigma}}_1^T, \tilde{\boldsymbol{\Sigma}}_2^T).$$

Hence, the proposed model is considering R=2 outcomes. Then, for $\mathbf{Y} = [\mathbf{Y_1}, \mathbf{Y_2}]$, the full log likelihood is

$$l(\boldsymbol{\mu}, \boldsymbol{\Sigma}(\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\rho})|\mathbf{y}) = -\frac{1}{2}\{\ln|\boldsymbol{\Sigma}| + (\mathbf{y} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})\}. \tag{8}$$

Here the covariance matrix is more flexible, allowing to model data sets with different correlation profiles. Besides, it is also possible to test, using the likelihood ratio test, whether the correlation coefficients differ, for example for each treatment understudy.

Jørgensen and Knudsen (2004) suggested a method to estimate all parameters of matrix $\boldsymbol{\Sigma}$ jointly, using the Newton scoring algorithm, based on quasi-likelihood and the Pearson estimating functions applying second-moment assumptions.

As an alternative approach, here, the matrices $\boldsymbol{\Sigma_1}$ and $\boldsymbol{\Sigma_2}$ will be estimated first separately, using the maximum likelihood method, presented in section (4.1). After that, it is possible to use these matrices as initial values to obtain the estimates $\hat{\boldsymbol{\rho}}$ by profiling log-likelihood function w.r.t. $\boldsymbol{\rho}$ and maximizing (8) 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 implemented in the function `mle2()` for `R` software (Team, 2020). Standard errors for the regression parameters are obtained based on the observed information matrix $\mathbf{I}(\boldsymbol{\theta})$, where $\mathbf{I}(\theta) = -\mathbf{H}(\theta)$ (hessian matrix) is computed numerically, using the `hessian()` function of the package `numDeriv` (Gilbert et al., 2006). As $\hat{\boldsymbol{\phi}}$ depend on $\hat{\boldsymbol{\beta}}$, the standard errors for

the dispersion parameters were obtained using a two-stage algorithm, considering $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\rho}}$ obtained from the previous stages.

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).The likelihood-ratio test was used to compare treatments and to assess the significance of the parameters.

5. Simulation studies

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

We conducted the simulation study based on the experimental structure described in the case study. Four simulation scenarios were designed considering two treatments with two levels and the numbers of plots equal to half of the sample size. 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 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 simulation scenario. The simulation process is summarised as follows

(1) Define $\boldsymbol{\beta_1} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$; $\boldsymbol{\beta_2} = \begin{bmatrix} 10 \\ 300 \end{bmatrix}$; $\boldsymbol{\phi_1} = \begin{bmatrix} 0.5 \\ 2 \end{bmatrix}$; $\boldsymbol{\phi_2} = \begin{bmatrix} 1 \\ 10 \end{bmatrix}$;

(2) Define $\mathbf{X}_{\beta_1}$ and $\mathbf{X}_{\beta_2}$ as $n \times 2$ design matrix;

(3) Define $\mathbf{X}_{\phi_1}$ and $\mathbf{X}_{\phi_2}$ as $n \times 2$ dispersion design matrix;

(4) Define $\boldsymbol{\mu_1} = g^{-1}(\mathbf{X}_{\beta_1}\boldsymbol{\beta_1}) = \exp(\mathbf{X}_{\beta_1}\boldsymbol{\beta_1})$ and $\boldsymbol{\mu_2} = g^{-1}(\mathbf{X}_{\beta_2}\boldsymbol{\beta_2}) = \mathbf{X}_{\beta_2}\boldsymbol{\beta_2}$;

(5) Define $\boldsymbol{\Sigma_1} = \mathrm{diag}(\mathbf{X}_{\phi_1}\boldsymbol{\phi_1})\mathrm{diag}(\boldsymbol{\mu_1})$ and $\boldsymbol{\Sigma_2} = \mathrm{diag}(\mathbf{X}_{\phi_2}\boldsymbol{\phi_2})\mathrm{diag}(\boldsymbol{\mu_2})$;

(6) Define $\mathbf{X}_\rho$ as a $n \times 1$ unit matrix

(7) Define $\boldsymbol{\Sigma}_\rho = \mathrm{diag}(\mathbf{X}_\rho\rho)$, where ρ is a scalar (-0.8, 0.2, 0.5 or 0.8);

(8) Define $\boldsymbol{\Sigma} = \mathrm{Bdiag}(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) \begin{bmatrix} \mathbf{I} & \boldsymbol{\Sigma}_\rho \\ \boldsymbol{\Sigma}_\rho & \mathbf{I} \end{bmatrix} \mathrm{Bdiag}(\boldsymbol{\Sigma}_1^T, \boldsymbol{\Sigma}_2^T)$;

(9) Simulate from $\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \ \boldsymbol{\Sigma} \right)$;

(10) Fit the proposed model to the simulated data sets. The complete code used for the simulations and model fitting can be found in the supplementary materials.

6. Results

6.1 *Model performance*

In this section, we present the results of the simulation studies described above.

To show the bias of estimation for each simulation scenario (combination between values of correlation parameter and sample sizes) Figure 2 was constructed. It presents the average bias, defined as the average estimate of 1000 simulations minus the true value and their respective confidence interval calculated as the average bias plus and minus 1.96 times the average standard error of 1000 simulations.

[Figure 2 about here.]

The results of the simulation studies showed that for all correlation levels, the average bias and the standard errors tend to 0 as the sample size increase, meaning that all parameters have unbiased and consistent estimators. For small samples, the dispersion parameters are underestimated. This is also possible to observe 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 increase. Another interesting point to note is that the average standard error is lower for high 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 3 was assembled to exhibit 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, that even for large samples the coverage rate is still lower than the nominal level. An alternative here is to considerer other methods to construct the confidence intervals such as the bootstrap approach.

[Figure 3 about here.]

To verify the coverage rate considering bootstrap confidence intervals, a parametric bootstrap simulation was performed, considering 1000 bootstrap samples and 100 simulations, respecting 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 to dispersion and correlation coefficients, then bias-corrected percentile confidence intervals

(BCP) were considered. Efron (1982) showed this method, in which the correction relies upon transforming the distribution to one that satisfies the symmetry condition (Diciccio and Romano, 1988; Buckland, 1984). Empirical coverage rates increased, becoming very close to 95% nominal coverage (Table 1).

[Table 1 about here.]

Table 2 shows the results of the simulation study considering two correlations coefficients, in this example, one for each treatment. The study considered 1000 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 $\rho_2 = 0.5$). It was also possible to observe that the average bias is low, and larger the sample lower the average bias and the average standard errors. As for the average standard error, it was possible to observe that for stronger correlations, the standard error was lower, implying empirical confidence intervals with smaller coverage than the nominal. Further, it is clear that the scenarios that presented better coverage, mainly for larger samples, were scenarios 1 and 4.

[Table 2 about here.]

6.2 *Podisus data analysis*

We fitted the bivariate model to the data set described in Section 2, including the effects of diet (treatment) in the linear predictors for the mean, dispersion and correlation. Table 3 shows the estimated parameters and standard errors. The females fed with *D. saccharalis* emerged with greater weight (β_{22}) and had greater fertility, producing more eggs (β_{12}). The estimates for ϕ_{11} and ϕ_{12} indicate overdispersion for the number of eggs, however since their standard erros are large we tested for their equality. The estimates for ϕ_{21} and ϕ_{22} are also similar.

[Table 3 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 4).

[Figure 4 about here.]

The selection of the model was performed using likelihood ratio tests. We verify the treatment effect on the mean and the dispersion by each outcome and jointly (Table 4 and 5). The effect of treatment on the correlation can also be tested, but here due to the differences in correlations observed per treatment (Figure 1), the biological assumption, and the small sample size, we choice to maintain a correlation effect per treatment. The test presented equal dispersion coefficients, but different regression coefficients for each diet. The regression parameters showed the difference between treatments for both outcomes, hence, the diet *Diatraea saccharalis* contributed to better development for the insect.

[Table 4 about here.]

[Table 5 about here.]

7. Discussion

In this work, we proposed a mean-dispersion bivariate framework to model counts and continuous responses. Our approach represents an attractive and flexible framework for studying two variables jointly and may be applied in many cases. This alternative considered a normal approach, 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 continuous data can be modeled. As this considered the multivariate normal approximation the estimation process presents better results with larger samples. The advantage of this methodology is that it is possible to incorporate correlation, jointly modeling the variable responses and assuming 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 analyze very small samples.

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 it is smaller. This became more evident by observing Figure 2, 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 $\hat{\boldsymbol{\beta}}$ varies very little for different values of ϕ and ρ. However, the coverage rate for the dispersion parameters also showed a vary depending on the correlation parameter level, and the higher the correlation, the lower the coverage rate. These cited problems are expected when it comes to estimates at the limit of the parametric space, especially with small samples. Even that, we concluded that the model behaves as expected, with unbiased, consistent, and accurate estimates.

Finally, natural extensions to the proposed bivariate model are in progress. These extensions include modeling 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 adding in the current $\boldsymbol{\Sigma} = \mathbf{D}_\phi \mathbf{D}_\mu$ the term $\mathbf{ZGZ}'$, where $\mathbf{Z}$ is the matrix of known covariates indicating the correlated observations or how the response evolves overtime for the ith subject and $\mathbf{G}$ a general covariance matrix. However, it is also necessary to include a link function for the dispersion parameter, so that it does not have a negative maximum likelihood estimate.

References

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 mcglm 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.

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.

Hinde, J., Demétrio, C. G., et al. (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.

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.*

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. (2020). 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 an Inference for Chapter 3

Univariate modelling

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

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

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

$$\frac{\partial l}{\partial \boldsymbol{\Sigma}} = \frac{\partial l}{\partial \boldsymbol{\Sigma}}\{-\tfrac{1}{2}\ln|\boldsymbol{\Sigma}| - \tfrac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\} \tag{A.2}$$

as $\dfrac{\partial}{\partial A}\ln|A| = A^{-T}$ and $\dfrac{\partial \mathbf{a^T X^{-1} b}}{\partial \mathbf{X}} = -\mathbf{X}^{-T}\mathbf{ab^T X^{-T}}$. Then,

$$\frac{\partial l}{\partial \boldsymbol{\Sigma}} = -\tfrac{1}{2}\boldsymbol{\Sigma}^{-T} + \tfrac{1}{2}\boldsymbol{\Sigma}^{-T}(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-T} \tag{A.3}$$

and

$$\frac{\partial l}{\partial \boldsymbol{\Sigma}}\frac{\partial \boldsymbol{\Sigma}}{\partial \mathbf{D}_\phi} = \left[-\tfrac{1}{2}\boldsymbol{\Sigma}^{-T} + \tfrac{1}{2}\boldsymbol{\Sigma}^{-T}(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-T}\right] \circ \mathbf{D}_\mu$$

$$= \left[-\tfrac{1}{2}(\mathbf{D}_\phi \mathbf{D}_\mu)^{-T} + \tfrac{1}{2}(\mathbf{D}_\phi \mathbf{D}_\mu)^{-T}(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})'(\mathbf{D}_\phi \mathbf{D}_\mu)^{-T}\right] \circ \mathbf{D}_\mu$$

$$= -\tfrac{1}{2}(\mathbf{D}_\phi \mathbf{D}_\mu)^{-T} \circ \mathbf{D}_\mu + \tfrac{1}{2}(\mathbf{D}_\phi \mathbf{D}_\mu)^{-T}(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})'(\mathbf{D}_\phi \mathbf{D}_\mu)^{-T} \circ \mathbf{D}_\mu$$

$$= -\tfrac{1}{2}\mathbf{D}_\phi^{-T}\mathbf{D}_\mu^{-T} \circ \mathbf{D}_\mu + \tfrac{1}{2}\mathbf{D}_\phi^{-T}\mathbf{D}_\mu^{-T}(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})'\mathbf{D}_\phi^{-T}\mathbf{D}_\mu^{-T} \circ \mathbf{D}_\mu$$

$$= -\tfrac{1}{2}\mathbf{D}_\phi^{-T} + \tfrac{1}{2}\mathbf{D}_\phi^{-T}\mathbf{D}_\mu^{-T}(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})'\mathbf{D}_\phi^{-T}.$$

If $\mathbf{A}$ is diagonal, then $\mathbf{A}^T = \mathbf{A}$ and

$$\frac{\partial l}{\partial \boldsymbol{\Sigma}} \frac{\partial \boldsymbol{\Sigma}}{\partial \mathbf{D}_\phi} \frac{\partial \mathbf{D}_\phi}{\partial \boldsymbol{\phi}} = [-\tfrac{1}{2}\mathbf{D}_\phi^{-1} + \tfrac{1}{2}\mathbf{D}_\phi^{-1}\mathbf{D}_\mu^{-1}(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'\mathbf{D}_\phi^{-1}] \circ \mathbb{I}$$

$$= -\frac{1}{2}\mathrm{diag}(\mathbf{X}_\phi\boldsymbol{\phi})^{-1} \circ \mathbb{I} + \frac{1}{2}\mathrm{diag}(\mathbf{X}_\phi\boldsymbol{\phi})^{-1}\mathbf{D}_\mu^{-1}(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'\mathrm{diag}(\mathbf{X}_\phi\boldsymbol{\phi})^{-1} \circ \mathbb{I}. \qquad (A.4)$$

Equalizing (A.4) to zero

$$-\tfrac{1}{2}\mathrm{diag}(\mathbf{X}_\phi\hat{\boldsymbol{\phi}})^{-1} \circ \mathbb{I} + \tfrac{1}{2}\mathrm{diag}(\boldsymbol{X}_\phi\hat{\boldsymbol{\phi}})^{-1}\mathbf{D}_\mu^{-1}(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'\mathrm{diag}(\mathbf{X}_\phi\hat{\boldsymbol{\phi}})^{-1} \circ \mathbb{I} = 0$$

$$\therefore \mathrm{diag}(\boldsymbol{X}_\phi\hat{\boldsymbol{\phi}})^{-1} \circ \mathbb{I} = \mathrm{diag}(\mathbf{X}_\phi\hat{\boldsymbol{\phi}})^{-1}\mathbf{D}_\mu^{-1}(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'\mathrm{diag}(\mathbf{X}_\phi\hat{\boldsymbol{\phi}})^{-1} \circ \mathbb{I}$$

$$\therefore \mathrm{diag}(\mathbf{X}_\phi\hat{\boldsymbol{\phi}}) = \mathbf{D}_\mu^{-1}(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})' \circ \mathbb{I}.$$

Applyng $\mathbb{J}_{n\times 1}$ in both sides and leaving $\hat{\boldsymbol{\phi}}$ in evidence

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

where $\mathbb{I}$ is an $n \times n$ identity matrix and $\mathbb{J}$ is an $n \times n$ vector of 1s, and $\circ$ the Hadamard product.
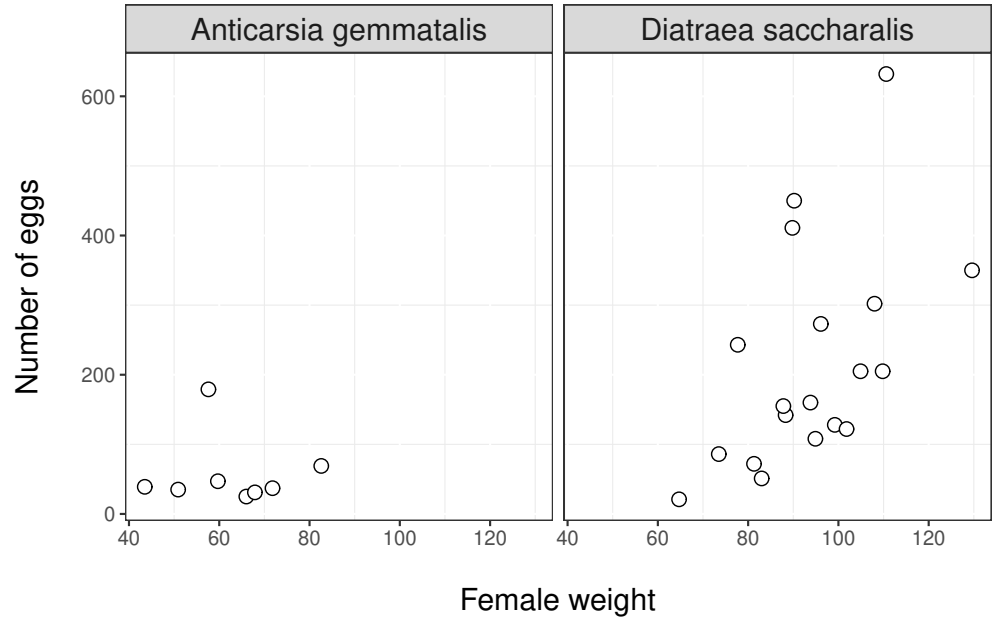
Figure 1: Scatterplot of female weight and number of eggs by treatment.

Sample size ⊖ 50 ⊟ 100 ◇ 300 △ 600
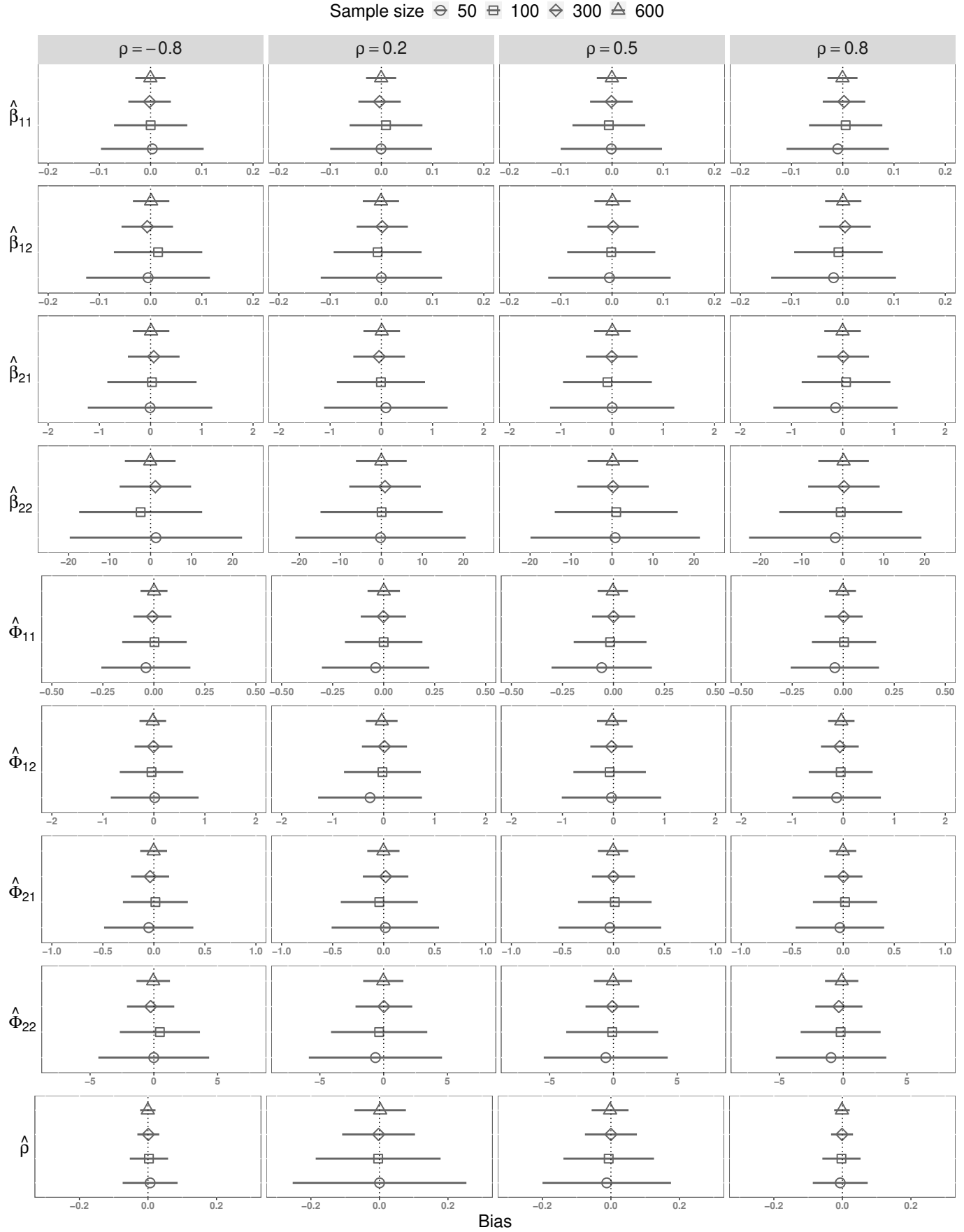
Figure 2: Estimation average bias of each simulation scenario (symbols) with their respective confidence interval of 95% obtained by the average standard error considering 1000 simulations.
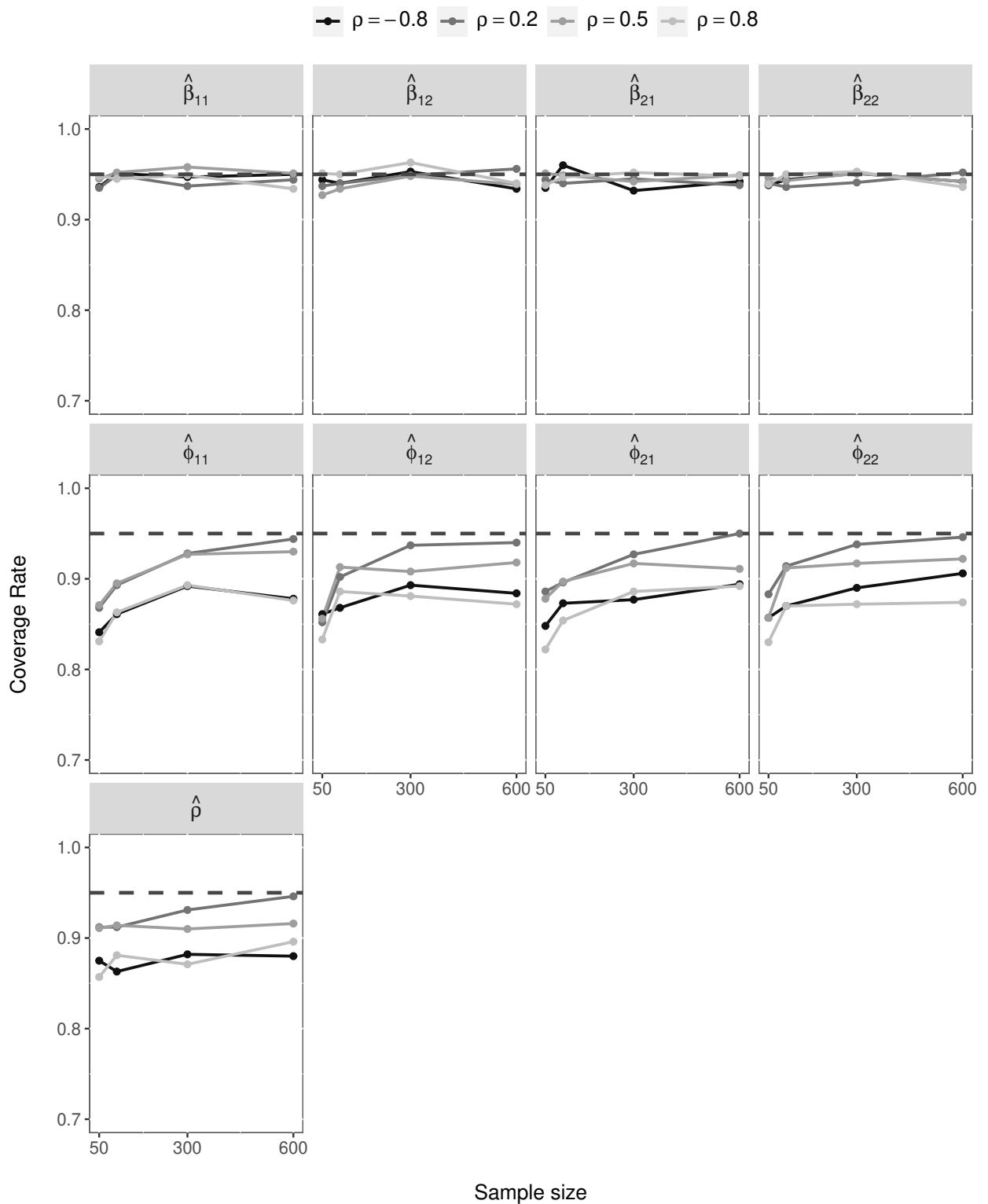
Figure 3: Coverage rate based on normal approximation confidence intervals with a nominal rate of 95% for different sample sizes (50, 100, 300, and 600) and correlation coefficients (-0.8, 0.2, 0.5 and 0.8).
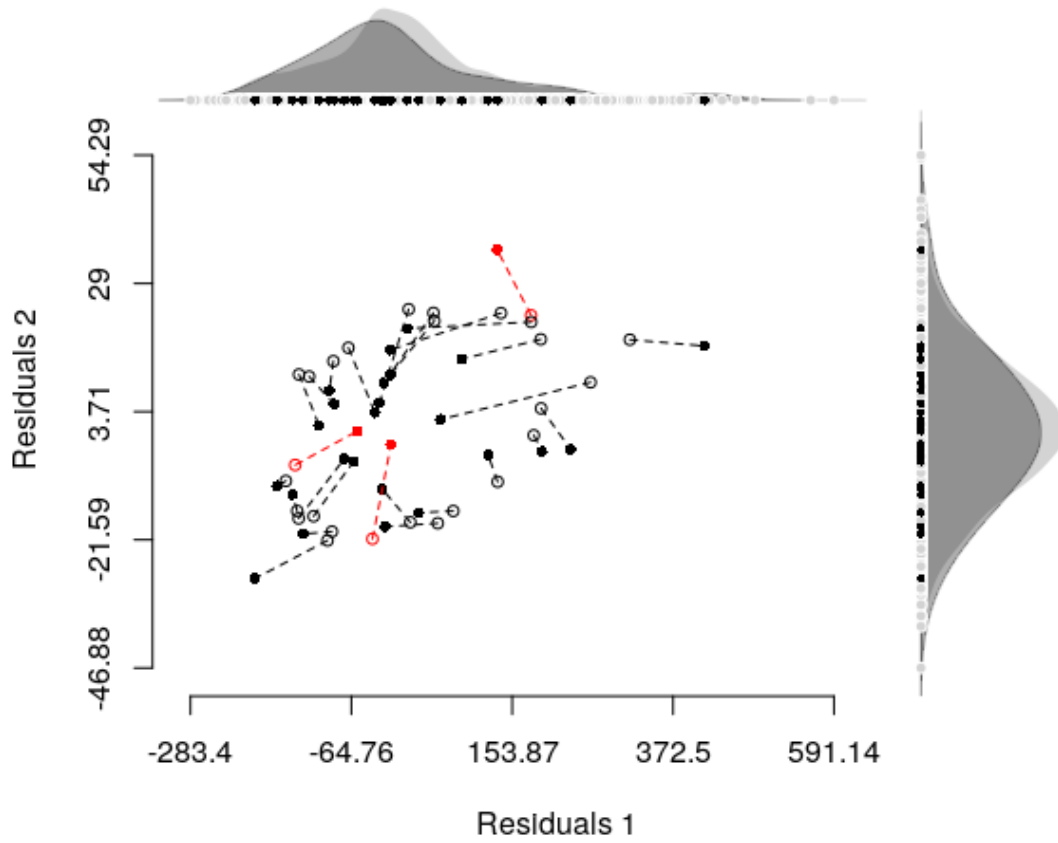
Figure 4: Bivariate plot with simulation polygons assuming a bivariate normal distribution for bivariate samples of size 99 simulated, with estimated marginal densities for residuals 1 and residuals 2. The an lighter shade corresponds to the estimated density of all simulated diagnostics, and the darker one corresponds to the estimated density of the observed diagnostics. The points outside their simulated polygons are displayed in red.

Table 1: Coverage rate based on bias-corrected percentile confidence intervals with 1000 bootstrap samples and 100 simulations, $n = 50$ and a nominal rate of 95%.

	Coverage rate	
Parameter	$\rho = 0.8$	$\rho = -0.8$
β_{11}	0.95	0.96
β_{12}	0.95	0.93
β_{21}	0.98	0.96
β_{22}	0.98	0.92
ϕ_{11}	0.94	0.95
ϕ_{12}	0.92	0.88
ϕ_{21}	0.93	0.95
ϕ_{22}	0.91	0.88
ρ	0.95	0.96

Table 2: Estimator mean bias, MSE and coverage rate for the proposed model under 4 scenarios, which varied true parameter value. Results are based on 1000 simulations.

Scenario 1:

Par.	True	Bias			MSE			Coverage		
		n=50	n=100	n=300	n=50	n=100	n=300	n=50	n=100	n=300
β_{11}	2	0.000	-0.004	0.000	0.050	0.036	0.021	0.94	0.94	0.94
β_{12}	3	-0.002	-0.001	0.000	0.061	0.044	0.026	0.94	0.93	0.95
β_{21}	10	-0.012	-0.018	-0.017	0.611	0.441	0.257	0.93	0.94	0.95
β_{22}	300	0.154	-0.059	-0.024	10.584	7.669	4.446	0.93	0.95	0.96
ϕ_{11}	0.5	-0.020	-0.013	-0.004	0.133	0.096	0.057	0.88	0.90	0.92
ϕ_{12}	2	-0.093	-0.054	-0.021	0.501	0.362	0.213	0.87	0.90	0.91
ϕ_{21}	1	-0.044	-0.015	-0.003	0.265	0.194	0.114	0.87	0.90	0.93
ϕ_{22}	10	-0.469	-0.080	0.083	2.501	1.844	1.069	0.87	0.90	0.91
ρ_1	0.2	-0.005	-0.004	-0.001	0.179	0.130	0.076	0.88	0.92	0.93
ρ_2	0.5	-0.013	-0.001	0.000	0.133	0.094	0.055	0.88	0.89	0.91

Scenario 2:

Par.	True	Bias			MSE			Coverage		
		n=50	n=100	n=300	n=50	n=100	n=300	n=50	n=100	n=300
β_{11}	2	-0.002	-0.001	0.001	0.051	0.036	0.021	0.91	0.95	0.94
β_{12}	3	-0.002	0.001	0.001	0.062	0.044	0.026	0.92	0.94	0.94
β_{21}	10	0.021	0.006	-0.004	0.614	0.440	0.256	0.94	0.94	0.94
β_{22}	300	-0.093	0.207	0.078	10.682	7.599	4.435	0.92	0.94	0.95
ϕ_{11}	0.5	-0.015	-0.014	-0.005	0.134	0.096	0.057	0.89	0.89	0.93
ϕ_{12}	2	-0.059	-0.048	-0.031	0.450	0.321	0.187	0.86	0.87	0.87
ϕ_{21}	1	-0.036	-0.021	-0.012	0.267	0.193	0.113	0.89	0.91	0.93
ϕ_{22}	10	-0.283	-0.281	-0.130	2.252	1.597	0.939	0.86	0.86	0.88
ρ_1	0.2	-0.011	0.003	0.000	0.180	0.129	0.076	0.90	0.91	0.95
ρ_2	0.8	-0.006	-0.002	-0.002	0.058	0.040	0.023	0.83	0.86	0.87

Scenario 3:

Par.	True	Bias			MSE			Coverage		
		n=50	n=100	n=300	n=50	n=100	n=300	n=50	n=100	n=300
β_{11}	2	-0.001	0.000	0.000	0.051	0.036	0.021	0.94	0.93	0.93
β_{12}	3	0.001	0.001	0.001	0.061	0.044	0.026	0.92	0.94	0.95
β_{21}	10	-0.013	-0.025	-0.007	0.615	0.442	0.257	0.92	0.93	0.95
β_{22}	300	-0.149	-0.321	0.245	10.699	7.599	4.444	0.94	0.93	0.94
ϕ_{11}	0.5	-0.014	-0.014	-0.001	0.112	0.080	0.047	0.83	0.85	0.89
ϕ_{12}	2	-0.109	-0.050	-0.029	0.524	0.384	0.225	0.86	0.91	0.93
ϕ_{21}	1	-0.030	-0.010	-0.002	0.225	0.163	0.095	0.84	0.86	0.89
ϕ_{22}	10	-0.239	-0.268	-0.100	2.701	1.918	1.130	0.88	0.91	0.94
ρ_1	-0.8	0.005	0.007	0.002	0.057	0.041	0.023	0.85	0.84	0.85
ρ_2	0.2	0.004	-0.003	-0.004	0.178	0.130	0.076	0.88	0.93	0.94

Scenario 4:

Par.	True	Bias			MSE			Coverage		
		n=50	n=100	n=300	n=50	n=100	n=300	n=50	n=100	n=300
β_{11}	2	-0.003	0.001	-0.001	0.051	0.036	0.021	0.94	0.94	0.94
β_{12}	3	-0.001	0.002	0.001	0.061	0.044	0.025	0.94	0.94	0.95
β_{21}	10	-0.041	-0.013	-0.002	0.612	0.441	0.256	0.93	0.96	0.94
β_{22}	300	-0.137	-0.179	0.361	10.583	7.597	4.450	0.94	0.95	0.95
ϕ_{11}	0.5	-0.017	-0.010	-0.004	0.134	0.097	0.057	0.88	0.91	0.92
ϕ_{12}	2	-0.085	-0.062	-0.035	0.503	0.361	0.212	0.88	0.90	0.91
ϕ_{21}	1	-0.035	-0.014	-0.009	0.267	0.194	0.113	0.88	0.92	0.91
ϕ_{22}	10	-0.448	-0.264	-0.071	2.506	1.815	1.072	0.86	0.88	0.91
ρ_1	-0.2	0.004	0.000	0.002	0.179	0.130	0.076	0.88	0.94	0.93
ρ_2	0.5	-0.015	-0.008	-0.005	0.134	0.095	0.055	0.88	0.90	0.90

Table 3: Parameter estimates (EST.) and standard errors (SE) for the normal bivariate model. β_{11} and β_{12} represent the mean of number of eggs for diets *Anticarsia* and *Diatraea* respectively; β_{21} and β_{22} represent the mean of weight for for diets *Anticarsia* and *Diatraea* respectively. ϕ_{11} and ϕ_{12} represent the dispersion of number of eggs for diets *Anticarsia* and *Diatraea* respectively; ϕ_{21} and ϕ_{22} represent the dispersion of weight for for diets *Anticarsia* and *Diatraea* respectively. ρ_1 represents the correlation between number of eggs and weight for *Anticarsia* diet and the parameter ρ_2 represents the correlation between number of eggs and weight for *Diatraea*.

	EST.	SE
β_{11}	4.0561	0.2907
β_{12}	5.3782	0.1609
β_{21}	62.5000	4.0684
β_{22}	93.9474	3.3824
ϕ_{11}	39.0292	19.5049
ϕ_{12}	106.6240	32.1814
ϕ_{21}	2.1186	1.0588
ϕ_{22}	2.3138	0.6983
ρ_1	-0.0389	0.3528
ρ_2	0.5192	0.1487
LogLik	-221.0125	

Table 4: 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(diff l)	$P(> \chi^2)$
Completely	10	-221.0125		
$\beta_{11} = \beta_{12}$	9	-227.4322	12.8394	0.0003
$\beta_{21} = \beta_{22}$	9	-236.2779	16.4631	0.0000
$\phi_{10} = \phi_{11}$	9	-222.2381	2.4511	0.2936
$\phi_{20} = \phi_{21}$	9	-221.0235	0.02209	0.9890

[1]np, number of parameters; l, log-likelihood; diff l, difference in log-likelihoods.

Table 5: 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(diff l)	$P(> \chi^2)$
Completely	10	-221.0125		
$\beta_{11} = \beta_{12}$ and $\beta_{21} = \beta_{22}$	8	-232.1660	22.3070	0.0000
$\phi_{10} = \phi_{11}$ and $\phi_{20} = \phi_{21}$	8	-222.2347	2.4445	0.2946

[2]np, number of parameters; l, log-likelihood; diff l, difference in log-likelihoods.