Negative variance components for non-negative hierarchical data with correlation, over-, and/or underdispersion

Peer-reviewed author version

Oliveira, I. R. C.; MOLENBERGHS, Geert; VERBEKE, Geert; DEMETRIO, Clarice & Dias, C. T. S. (2017) Negative variance components for non-negative hierarchical data with correlation, over-, and/or underdispersion. In: JOURNAL OF APPLIED STATISTICS, 44(6), p. 1047-1063.

DOI: 10.1080/02664763.2016.1191624
Handle: http://hdl.handle.net/1942/24102

Negative Variance Components for Nonnegative Hierarchical Data

Journal:	*Journal of Applied Statistics*
Manuscript ID	CJAS-2015-0868
Manuscript Type:	Original Article
Date Submitted by the Author:	18-Oct-2015
Complete List of Authors:	Oliveira, Izabela; ESALQ/University of São Paulo, Department of Exact Sciences Molenberghs, Geert; Hasselt University, Center for Statistics; Verbeke, Geert; Katholieke Universiteit Leuven, Biostatistical Centre G.B.Demetrio, Clarice ; University of São Paulo, Dias, Carlos; ESALQ/USP,
Keywords:	Gamma distribution, Generalized linear mixed model, Overdispersion, Poisson distribution, Combined model
2010 Mathematics Subject Classification:	62A01

To appear in the *Journal of Applied Statistics*
Vol. 00, No. 00, Month 20XX, 1–16

Negative Variance Components for Nonnegative Hierarchical Data

I.R.C. Oliveira[a,b]*, G. Molenberghs[b,c], G. Verbeke[c], C.G.B. Demétrio[a] and C.T.S. Dias[a]

[a] *Department of Exact Sciences, ESALQ, 13418-900 Piracicaba, São Paulo, Brazil*; [b] *Interuniversity Institute for Biostatistics and Statistical Bioinformatics, Universiteit Hasselt, Hasselt, Belgium*; [c] *Interuniversity Institute for Biostatistics and Statistical Bioinformatics, KU Leuven, Leuven, Belgium*

(*v4.1 released December 2013*)

The concept of negative variance components in linear mixed-effects models, while confusing at first sight, has received considerable attention in the literature, for well over half a century, following the early work of [20] and [7]. Broadly, negative variance components in linear mixed models are allowable if inferences are restricted to the implied marginal model. When a hierarchical view-point is adopted, in the sense that outcomes are specified conditionally upon random effects, the variance-covariance matrix of the random effects must be positive-definite (positive-semi-definite is also possible, but raises issues of degenerate distributions). Many contemporary software packages allow for this distinction. Less work has been done for generalized linear mixed models. Here, we study such models, with extension to allow for overdispersion, for nonnegative outcomes (counts). Using a study of trichomes counts on tomato plants, it is illustrated how such negative variance components play a natural role in modeling both the correlation between repeated measures on the same experimental unit and overdispersion.

Keywords: Combined model; Gamma distribution; Generalized linear mixed model; Overdispersion; Poisson distribution

1. Introduction

The need for inference on variance components arises in a variety of applied fields. Existing tools to this effect encompass random-effects ANOVA models [20], linear mixed models [27], generalized linear mixed models [16], and other models that also accommodate overdispersion and clustering [6]. By definition, when variance components are interpreted as variances, they are nonnegative quantities, but the occurrence of negative estimates is a reasonably well understood phenomenon in the context of linear models for hierarchical data. Of course, then the interpretation as variance is dropped; rather, such components play a role in the induced marginal model only.

In studies involving grouped data, it is common that the observations within the same cluster are positively correlated, which implies that such observations belonging to the same cluster are more similar to one another than observations from different clusters. Such dependence is sometimes measured by the intraclass correlation. Occasionally, though, observations within clusters may be dissimilar, e.g., when there are competition effects. An example of negative influence in grouped data is when there is a fixed resource and cluster members have to compete for it, leading to a negative intraclass

*Corresponding author. Email: izabela.oliveira@dex.ufla.br

correlation [12], such as when plants in the same plot compete for the same nutrients, water and/or light. The resulting within-plot negative correlation is then captured by negative variance-component estimates.

Note that a negative variance component (generically referring to a model parameter in a variance-covariance matrix) should not be confused with a negative variance; the latter does not exist. By a negative variance component, we mean here a parameter that would be a variance should a hierarchical (also referred to as conditional) view be adopted, but merely is a parameter in the implied marginal variance-covariance matrix (obtained by integrating over all random effects; e.g., in a compound-symmetry model, a covariance term), which is still positive-definite. Further, by hierarchical model we mean one where outcomes are modeled conditional upon covariates and random effects [cf. 8, 16]. The hierarchical view is then one where the parameters corresponding to the distribution of the random effects retain their meaning throughout the inference process. In contrast, a marginal view is one where the parameters should merely meaningfully describe the distribution of the outcomes, after integrating over the random effects.

The occurrence of negative variance components in linear mixed models was reviewed by [18]. As alluded to before, whenever inference for variance components is required, one will have to make a choice between a hierarchical and a marginal view. Under a marginal interpretation, the variance component can be negative as long as the resulting marginal variance-covariance matrix of the observations is positive definite. On the other hand, when a hierarchical view is adopted, random effects retain their interpretation and, hence, their variances must be nonnegative. [23] focused on negative variance components in generalized linear mixed models, specifically on binary and count outcomes. In this paper, we also allow for overdispersion, through the use of additional, usually conjugate, random effects. Negative variance components can then occur related to either the normal random effects, or the conjugate random effects, or both.

In this paper, our focus is on variance components of the so-called Poisson combined model, presented by [19] for modeling overdispersion and cluster-induced correlation in count data through two separate sets of random effects. Assuming gamma and normal distributions for the random effects leads to the Poisson-Gamma-Normal (PGN) model.

The marginal variance is made up of contributions from both random effects, as well as from the mean-variance relationship of the underlying generalized linear model. The counterpart for time-to-event data is the Weibull-Gamma-Normal (WGN) model. However, we will focus on the PGN, for brevity.

In Section 2, we review the count data case, from the simple, purely Poisson, to the PGN. Important additional expressions related to these models are presented in Appendix A. The variance components related to the gamma and normal random effects in the PGN are further studied in Section 3. Comments regarding estimation are provided in Section 4. Section 5 reports on the application of the PGN model, from which negative variance components arise.

2. Background on Poisson Models

The Poisson model is a natural choice for count data. This model is one of the prominent members of the exponential family [14]. The latter provides an elegant and encompassing mathematical framework within the generalized linear modeling context [14, 21].

Let Y_i be Poisson distributed with mean λ_i, denoted by $Y_i \sim\text{Poi}(\lambda_i)$. The probability mass function can be written as

$$f(y_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} = \exp\{y_i \ln \lambda_i - \lambda_i - \ln y_i!\}.$$

The variance function equals $\nu_i(\mu_i) = \mu_i = \lambda_i$. The logarithm is the natural link function, leading to the classical Poisson regression model with $\ln \lambda_i = \boldsymbol{x_i'}\boldsymbol{\xi}$, where $\boldsymbol{x}_i$ represent the p-dimensional vector of covariate values and $\boldsymbol{\xi}$ a vector of p fixed unknown regression coefficients.

The model imposes equality of mean and variance, although empirical research has abundantly demonstrated that this assumption is often not met in real data scenarios. Therefore, a number of extensions have been proposed [4, 13]. An elegant way to deal with overdispersion, that is, when the variability is greater than predicted by the mean-variance relationship, is through a random-effects approach: $Y_i|\lambda_i \sim \text{Poi}(\lambda_i)$ where λ_i is a random variable with $\text{E}(\lambda_i) = \mu_i$ and $\text{Var}(\lambda_i) = \sigma_i^2$. Using iterated expectations, it follows that

$$E(Y_i) = E[E(Y_i|\lambda_i)] = E(\lambda_i) = \mu_i,$$
$$\text{Var}(Y_i) = E[\text{Var}(Y_i|\lambda_i)] + \text{Var}[E(Y_i|\lambda_i)] = E(\lambda_i) + \text{Var}(\lambda_i) = \mu_i + \sigma_i^2.$$

It is common to assume a gamma distribution for λ_i, leading to the Poisson-Gamma (PG-) model, also known as the negative-binomial model [9, 10].

This model can easily be extended to the case of repeated measures. For this, let us assume a hierarchical data structure, where Y_{ij} denotes the jth outcome measured for cluster i $(i = 1,\ldots,N; j = 1,\ldots,n_i)$ and $\boldsymbol{Y_i}$ is the n_i-dimensional vector of all measurements available for cluster i. The vector of parameters is then $\boldsymbol{\lambda}_i = (\lambda_{i1},\ldots,\lambda_{in_i})'$, with $E(\boldsymbol{\lambda}_i) = \boldsymbol{\mu}_i$ and $\text{Var}(\boldsymbol{\lambda}_i) = \Sigma_i$. Then, $E(\boldsymbol{Y_i}) = \boldsymbol{\mu_i}$ and $\text{Var}(\boldsymbol{Y_i}) = M_i + \Sigma_i$ where M_i is a diagonal matrix with the vector $\boldsymbol{\mu}_i$ along the main diagonal. For example, assuming the components of $\boldsymbol{\lambda_i}$ to be independent, a pure overdispersion model results, without correlation between the repeated measures. Also, assuming $\lambda_{ij} = \lambda_i$, then $\text{Var}(\boldsymbol{Y_i}) = M_i + \sigma_i^2 J_{ni}$, where J_{n_i} is an $n_i \times n_i$ dimensional matrix of ones. Such a structure can be seen as a count-data version of compound symmetry in the scale of the canonical log link. In many applications these assumptions will not apply and then more general versions, or other sub-models thereof, can be used without any problem.

In hierarchical data modeling, the generalized linear mixed model (GLMM) [5, 16] has become a standard tool in the context of non-Gaussian measures. For the specific case of count data, the parameters become $\lambda_{ij} = \exp(\boldsymbol{x}_{ij}'\boldsymbol{\xi} + \boldsymbol{z}_{ij}'\boldsymbol{b}_i)$, with $\boldsymbol{b}_i \sim N(\boldsymbol{0}, D)$. Owing to the use of the logarithmic link and the normality of the random effects, the mean vector and variance-covariance matrix of $\boldsymbol{Y}_i$ can be derived in closed form [19]. The expressions are presented in Appendix A.

An extended version of the aforementioned models was presented by [19], in line with [3]. These extensions accommodate correlated count data with overdispersion, simultaneously combining two separate sets of random effects.

Assuming normal and gamma distributions for these random effects, the so-called combined model, that is, the Poisson-Gamma-Normal (PGN) model follows. It yields the Poisson-Normal (P-N), the Poisson-Gamma (PG-), and the purely Poisson model (P--) as special cases. The P-N is the standard generalized linear mixed model with the Poisson distribution and logarithmic link function and does not contain the gamma random effects. As mentioned in the introduction, the PG- is the negative binomial model and P-- is the ordinary Poisson, without both gamma and normal random effects.

The PGN model elements are:

$$\begin{aligned} Y_{ij} &\sim \text{Poi}(\theta_{ij}\lambda_{ij}), \\ \lambda_{ij} &= \exp(\boldsymbol{x}_{ij}'\boldsymbol{\xi} + \boldsymbol{z}_{ij}'\boldsymbol{b}_i), \\ \boldsymbol{b}_i &\sim N(\boldsymbol{0}, D), \\ \theta_{ij} &\sim \text{Gamma}(\alpha_{ij}, \alpha_{ij}^{-1}), \\ \text{Var}(\boldsymbol{\theta}_i) &= \Sigma_i. \end{aligned} \tag{1}$$

It is often assumed that the components θ_{ij} of $\boldsymbol{\theta}_i$ are independent, which is usually reasonable because the $\boldsymbol{b}_i$ components induce association between repeated measures, while θ_{ij} capture additional dispersion. However, the θ_{ij} can be assumed dependent as well. In the independence case, Σ_i reduces to a diagonal matrix with the variances of the gamma random effects along the main diagonal. Note that, because of the parameterization of the gamma random effects, their mean is equal to 1. This avoids aliasing with the intercept term in the linear predictor. Further, the variances in Σ_i take the form $(\alpha_{ij}+1)/\alpha_{ij}$. As a special case, the α_{ij} can be chosen to be independent of j, or even constant across independent replications. These and the components of D play a crucial role in what follows.

We note in passing that model assessment tools for the PGN have been studied as well [24]. This will not be pursued further in this paper.

The PGN and its sub-models admit closed-form expressions for means, variances, and higher-order moments. As a result, the correlations too have closed-form expressions [26]. All of these are presented in Appendix A.

3. The Case of Random Intercepts and Independent Gamma Variables

We focus on the variance components of the PGN, for the special and important case where the random-effects structure is reduced to random intercepts only, and with a constant mean function, thereby reducing the linear predictor to merely ξ_0. We will also assume that the gamma variables are identically and independently distributed, that is, Σ_i is a diagonal matrix, with elements α along the diagonal. In such case, the components of the mean vector, $\boldsymbol{\mu}_i = E(\boldsymbol{Y}_i)$, presented in (A1) reduce to

$$\mu = \exp\left(\xi_0 + \frac{1}{2}d\right) = \exp(\xi_0)\exp\left(\frac{1}{2}d\right) \equiv \mu_0 \exp\left(\frac{1}{2}d\right),$$

where d is the scalar version of D in case there is only one normal random effect. To simplify notation, let $\Delta \equiv \exp(d/2)$. Now, we can rewrite the mean, variance, covariance, and correlation expressions of the PGN, presented in Appendix A, in terms of Δ:

$$\begin{aligned} \mu &= \mu_0\Delta, \\ \sigma^2 &= \mu_0\Delta + \mu_0^2\Delta^2(\Delta^2\alpha + \Delta^2 - 1), \\ \rho &= \frac{\mu_0\Delta(\Delta^2-1)}{1+\mu_0\Delta(\Delta^2\alpha+\Delta^2-1)}, \end{aligned} \tag{2}$$

where μ_0, Δ and α are unknowns. Assume now that $\widetilde{\mu}$, $\widetilde{\sigma}^2$ and $\widetilde{\rho}$, the mean, variance and correlation, are given, and solve μ_0, Δ and α, where:

$$\widetilde{\mu} = \mu_0\Delta, \tag{3}$$

$$\widetilde{\sigma}^2 = \mu_0\Delta + \mu_0^2\Delta^2(\Delta^2\alpha + \Delta^2 - 1), \tag{4}$$

$$\widetilde{\rho} = \frac{\mu_0\Delta(\Delta^2 - 1)}{1 + \mu_0\Delta(\Delta^2\alpha + \Delta^2 - 1)}. \tag{5}$$

Conditions must be imposed: $\widetilde{\sigma}^2 \geq 0$ and $1 \geq \widetilde{\rho} \geq -1/(n-1)$, the latter to ensure the matrix be positive definite. Also, write $\widetilde{\mu}\widetilde{\theta} = \widetilde{\sigma}^2$, where $\widetilde{\theta}$ is the overdispersion effect. The solution to the system of equations (3)–(5) is:

$$\mu_0 = \frac{\widetilde{\mu}^2}{\sqrt{\widetilde{\mu}^2 + \widetilde{\rho}\widetilde{\sigma}^2}} = \sqrt{\frac{\widetilde{\mu}^3}{\widetilde{\mu} + \widetilde{\rho}\widetilde{\theta}}}, \tag{6}$$

$$\alpha = \frac{\widetilde{\sigma}^2 - \widetilde{\mu} - \widetilde{\rho}\widetilde{\sigma}^2}{\widetilde{\mu}^2 + \widetilde{\rho}\widetilde{\sigma}^2} = \frac{\widetilde{\theta} - (1 + \widetilde{\rho}\widetilde{\theta})}{\widetilde{\mu} + \widetilde{\rho}\widetilde{\theta}}, \tag{7}$$

$$\Delta^2 = 1 + \frac{\widetilde{\rho}\widetilde{\sigma}^2}{\widetilde{\mu}^2} = \frac{\widetilde{\mu} + \widetilde{\rho}\widetilde{\theta}}{\widetilde{\mu}}, \tag{8}$$

leading to

$$d = \ln\left(1 + \frac{\widetilde{\rho}\widetilde{\sigma}^2}{\widetilde{\mu}^2}\right) = \ln\left(\frac{\widetilde{\mu} + \widetilde{\rho}\widetilde{\theta}}{\widetilde{\mu}}\right). \tag{9}$$

In the following, we will study this solution in some detail.

It should be noted that, should a hierarchical interpretation be desired, α is the shape parameter of the gamma distribution and should be positive. In a merely marginal view, α is free of this interpretation and can be viewed simply as an additional model parameter, to add flexibility to the variance and correlation functions.

3.1 *Variance Component Induced by the Gamma Random Effect*

In this and the next section we provide additional insight into when negative variance components and/or negative correlation occurs. This is an aid for the researcher to interpret their findings from a particular data analysis.

First, we study the variance component related to the gamma random effect, α. If $\widetilde{\rho} = 0$, that is, when there is no intraclass correlation, $d = 0$, because $\Delta^2 = \exp(d) = 1$, which is obvious given the normal random effect captures the association between repeated measurements.

Turning attention to the gamma variance component where $\widetilde{\rho} = 0$, we have that $\alpha = (\widetilde{\sigma}^2 - \widetilde{\mu})/\widetilde{\mu}^2 = (\widetilde{\theta} - 1)/\widetilde{\mu}$. Hence, if there is no overdispersion, that is, $\widetilde{\theta} = 1$, then $\alpha = 0$. If there is overdispersion ($\widetilde{\sigma}^2 > \widetilde{\mu}$), $\widetilde{\theta} > 1$ and $\alpha > 0$. On the other hand, $\alpha < 0$ when $\widetilde{\theta} < 1$, that is, when underdispersion occurs ($\widetilde{\sigma}^2 < \widetilde{\mu}$).

When there is perfect positive intraclass correlation, $\widetilde{\rho} = 1$. In such a case, d is positive because $\Delta^2 = 1 + \widetilde{\sigma}^2/\widetilde{\mu}^2 = 1 + \widetilde{\theta}/\widetilde{\mu}$. On the other hand, $\alpha = -\widetilde{\mu}/(\widetilde{\mu}^2 + \widetilde{\sigma}^2) = -\left(\widetilde{\mu} + \widetilde{\theta}\right)^{-1}$ and its sign depends on whether $\widetilde{\mu} + \widetilde{\theta} > 0$ or $\widetilde{\mu} + \widetilde{\theta} < 0$. The latter is not possible, implying that for perfect positive correlation, α must be negative. Clearly then, a hierarchical interpretation is not possible, but, importantly, the model is valid from a purely marginal point of view. This development describes a situation where negative values for α will occur. While we concentrated on a limiting case, negative α values will occur for correlations that are sufficiently large. While perfect correlations are rarely encountered in practice, large correlations are not uncommon. It is then important to have available this more flexible marginal model.

Another useful scenario is when there is no overdispersion ($\widetilde{\theta} = 1$) and $\widetilde{\rho}$ is arbitrary. In such case $\Delta^2 = (\widetilde{\mu} + \widetilde{\rho})/\widetilde{\mu}$, implying $d = \ln[(\widetilde{\mu} + \widetilde{\rho})/\widetilde{\mu}]$ and $\alpha = -\widetilde{\rho}(\widetilde{\mu} + \widetilde{\rho})$. Observe that, if there is positive intraclass correlation ($\widetilde{\rho} > 0$) then α must be negative, that is, $\widetilde{\rho} \geq 0$, where $\widetilde{\rho} \in [0, 1]$, or $\widetilde{\rho} < -\widetilde{\mu}$, where $\widetilde{\rho} \in [-1, -\widetilde{\mu}[$, which can happen only for $\widetilde{\mu}$ in the unit interval. Thus, if there is positive correlation but no overdispersion, the overdispersion that is forced upon the model by the said positive correlation should be compensated for; this can be done only through a negative α.

On the other hand, in the special case without overdispersion and for $\widetilde{\rho} = 0$, then $\Delta^2 = 1$, $d = 0$ and $\alpha = 0$.

Observing the solution presented in (6)–(9), it is clear that α can be infinity if $\widetilde{\mu} + \widetilde{\rho}\widetilde{\theta} = 0$, implying that $\widetilde{\rho}\widetilde{\theta} = -\widetilde{\mu}$. Then, μ_0 tends to $+\infty$ and $\Delta^2 = 0$, that is, d tends to $-\infty$.

3.2 *Variance Component Induced by the Normal Random Effect*

The variance component d, associated with the normal random effects in the PGN is nonzero if and only if $\Delta^2 \geq 0$. Then,

$$1 + \frac{\widetilde{\rho}\widetilde{\sigma}^2}{\widetilde{\mu}^2} \geq 0 \Longleftrightarrow \widetilde{\rho} \geq -\frac{\widetilde{\mu}^2}{\widetilde{\sigma}^2} = -\frac{\widetilde{\mu}}{\widetilde{\theta}}.$$

Depending on the values of $\widetilde{\mu}$ and $\widetilde{\theta}$, this will or will not be a genuine condition. Precisely, if $\widetilde{\mu} \geq \widetilde{\theta}$ the above condition is sufficient. On the other hand, there is an additional restriction if $\widetilde{\mu} \leq \widetilde{\theta}$. For d to be nonnegative:

$$\begin{aligned} 1 + \frac{\widetilde{\rho}\widetilde{\sigma}^2}{\widetilde{\mu}^2} \geq 1 &\Longleftrightarrow \frac{\widetilde{\rho}\widetilde{\sigma}^2}{\widetilde{\mu}^2} \geq 0 \\ &\Longleftrightarrow \widetilde{\rho} \geq 0, \end{aligned}$$

because $\widetilde{\sigma}^2, \widetilde{\mu}^2 \geq 0$. So, nonnegative intraclass correlation ($\widetilde{\rho}$) implies nonnegative d and negative intraclass correlation implies negative d, in line with results that hold for the linear mixed model.

3.3 *Existence of an Extended Marginal Model*

When a hierarchical model is formulated and its hierarchical interpretation is preserved, then the implied marginal model is valid in the sense that it rests upon a valid probability density function. When the model is marginalized and the so-obtained model is considered on its own terms (the marginal view), the question arises as to which parameter combinations make up a valid model.

Prior to addressing this for the PGN, we examine a few pivotal standard situations. Starting from the linear mixed model, and in particular the random-intercepts model $Y_{ij} \sim N(\boldsymbol{x}_{ij}'\boldsymbol{\xi} + b_i, \sigma^2)$, with $b_i \sim N(0, d)$, $\boldsymbol{\xi}$ the fixed effects, and $\boldsymbol{x}_{ij}$ the covariate vector for subject i at occasion j, where $i = 1, \ldots, N$ and $j = 1, \ldots, n_i$. The induced marginal model is multivariate normal $\boldsymbol{Y}_i \sim N(X_i\boldsymbol{\xi}, \sigma^2 I_{n_i} + dJ_{n_i})$, with $\boldsymbol{Y}_i$ the vector of Y_{ij},

X_i the design matrix for subject i with rows made up by the $\boldsymbol{x}_{ij}$, and I_{n_i} and J_{n_i} the $n_i \times n_i$ identity and one matrices, respectively. The only condition for this model to be valid is that $V_i = \sigma^2 I_{n_i} + dJ_{n_i}$ be positive-definite. This is known to be satisfied if $\rho = d/(d+\sigma^2) \geq -(n_i-1)^{-1}$. While this is a simple condition, it should be noticed that there remain subtle differences between positive and negative correlation. When the correlation is positive, a valid model is obtained regardless of the cluster sizes n_i, but for negative correlation, the above condition places a bound on the maximal cluster size. Note that exactly the same condition applies to the marginalized beta-binomial model [see 16, 22, 25]. For the linear mixed model with general random-effects design, the condition on the marginal variance-covariance matrix is that $V_i = Z_i D Z_i' + \Sigma_i$ is positive-definite (with Z_i the random-effects design, D the variance of the random-effects vector, and Σ_i the residual variance-covariance matrix). It should be clear already that the condition that D and Σ_i be positive-definite is easier than that V_i be positive-definite over a relevant set of Z_i and Σ_i. Even when $\Sigma_i = \sigma^2 I_{n_i}$, there is still a dependence on the cluster size.

While the above marginalized hierarchical models are still relatively easy to study, it is much worse for a model like the Bahadur model, a directly specified marginal model for multivariate binary data, because the restrictions that apply to its parameter space (consisting of pairwise and higher-order correlations) are to this day only partially studied. Sufficient work has been done to know that the parameter space is highly restricted [1, 2, 16].

It is therefore unrealistic to expect that one can very easily establish that a given parameter combination for the marginalized PGN leads to a valid model. We have several tools and arguments at our disposition, though.

First, as shown in Appendix A, not only the marginal mean, variance, covariance, and correlations are known in explicit form, the same is true for the marginal joint probability mass function, although it takes the form of an infinite series (see also Appendix A). Admittedly, this probability mass function is not so easy to examine. It is more fruitful to study the marginal cumulants or moments. The cumulants are especially simple because, in the standard Poisson model, all cumulants are equal to the Poisson parameter. [19] established the higher-order moments as well:

$$E(Y_{ij}^k) = \sum_{\ell=0}^{k} S(k,\ell) \frac{\beta^\ell \Gamma(\alpha+\ell)}{\Gamma(\alpha)} \exp\left[\ell \boldsymbol{x}_{ij}' \boldsymbol{\xi} + \frac{1}{2}\ell^2 \boldsymbol{z}_{ij}' D \boldsymbol{z}_{ij}\right],$$

where $S(k,\ell)$ is the so-called Stirling number of the second kind. Moments of a mixed type, i.e., involving products of various outcomes of the same subject, can be derived in a similar fashion.

Second, and very important, it is actually not necessary to study the higher-order moments or cumulants. Rather, it is sufficient that the mean and variance of the marginalized PGN exist. As soon as this is satisfied, there exists a model, though not necessarily of PGN form in all of its moments, that is valid. This assertion is based on the work by [15]. These authors show that generalized estimating equations, when producing a valid marginal mean function and variance-covariance structure, can be thought of as coming from a valid joint distribution function. Their argument is based on considering *conditional* higher-order moments, rather than marginal ones. Fortunately, the same argument can be invoked here. In Sections 3.1 and 3.2, we derived sets of parameters that are marginally valid, even though they do not correspond to a hierarchical PGN, indicating that there is value in the marginal extension.

Third, [11] showed how the combined model, using a general random-effects structure combined with residual marginal association, can be used to broadly and flexibly generate correlated count data. These authors start from a given marginal structure, and then use normal random effects and/or multivariate gamma variables to derive a pre-specified multivariate Poisson random variable. In case one wants to generate such variables purely marginally, it is sufficient to work with the multivariate gamma distribution.

All of this, taken together, provides sufficient credibility to the marginalized PGN, of the same nature as available for other multivariate distributions, regardless of whether or not they are marginalized versions of hierarchical models.

4. Estimation and Inference

Standard generalized linear mixed models can be fitted to data with a variety of software tools, such as the SAS procedures GLIMMIX and NLMIXED. These and other tools offer a variety of numerical optimization algorithms, a key component of which is the method for integrating over the normal random effects. It has been shown [16] that Taylor-series-expansion based methods, such as MQL and PQL, perform poorly, especially with binary data, but that the quality of the approximation used, especially for PQL, improves with count and time-to-event data. Further methods are based on Laplace approximations and Gauss-Hermite quadrature. The PGN, formulated in (1), can be fitted using the SAS procedure NLMIXED, because it allows to flexibly use program statements for the conditional likelihood. The conditional likelihood here is understood as the likelihood integrated over the conjugate but not over the normal random effects. Using the example in the next section, we will assess the relative ease/complexity with which boundary and/or negative estimates can or cannot be accommodated using the various methods. When a negative estimate is allowed for by the user, it is still possible that it cannot be found purely because an algorithm is used that does not allow for it (i.e., that requires a hierarchical interpretation). For example, we will note that the Laplace approximation allows for negative normal random-effects variance components. There is then a tradeoff between the accuracy of a method on the one hand and its capability of extending the parameter space of the variance components on the other.

Drawing inferences from estimated variance components is not trivial, regardless of whether linear mixed models or extensions of the type discussed in this paper are considered. When non-negativity constraints are lifted, the so-called boundary problem is removed and standard asymptotic inference tools can be used (likelihood ratio, score, and Wald tests asymptotically follow χ^2 distributions). In the reverse case, mixtures of χ^2's should be used instead. Which mixtures apply in a particular cases depends on the geometry of the null and alternative parameter space, and fortunately not on the particular hierarchical model used [17, 28].

5. Inheritance Study of Trichomes Density in Tomato

We consider data from an experiment that was implemented at the Biology Department of the Federal University of Lavras, Brazil, during the first semester of 2014, to study the inheritance of some types of trichomes in tomato. Trichomes are epidermal outgrowths of diverse structure and function on plants. The glandular ones are of particular importance. They secrete oils, essences, enzymes, urticant substances and, for this reason, some of them are associated with resistance to some insect-pests.

In this study, plants from populations P_1, P_2, F_1, F_2, $BC_{1(1)}$ and $BC_{1(2)}$ were evaluated. On each plant three locations were sampled and at each of them an area of 1 mm^2 was defined. Furthermore, at each location, both front (adaxial) and back (abaxial) faces of

Table 1. Tomatoes study. Descriptive statistics for different trichomes, considering 19 plants from P_2 population.

Face	Glandular	Type	Code	$\bar{x}$	s^2
Abaxial	Yes	I	G-I-ab	0	0
		IV	G-IV-ab	20.28	35.03
		VI	G-VI-ab	0	0
		VII	G-VII-ab	0	0
		I, IV, VI and VII	G-ab	20.28	35.03
	No	V	NG-V-ab	1.93	3.21
		II and III	NG-II-III-ab	0.11	0.17
		II, III and V	NG-ab	2.04	3.18
Adaxial	Yes	I	G-I-ad	0.04	0.04
		IV	G-IV-ad	7.11	30.13
		VI	G-VI-ad	0.09	0.22
		VII	G-VII-ad	0.14	0.23
		I, IV, VI and VII	G-ad	7.37	29.63
	No	V	NG-V-ad	3.33	2.98
		II and III	NG-II-III-ad	0.19	0.55
		II, III and V	NG-ad	3.53	3.08

Table 2. Tomatoes study. Parameter estimates (standard errors) for the PGN, considering random intercepts, for some trichome counts. The estimation method used was penalized quasi-likelihood (PQL).

Effect	Par.	G-IV-ab	G-IV-ad	NG-V-ab	NG-V-ad
Intercept	ξ_0	3.01 (0.05)	1.84 (0.13)	0.55 (0.17)	1.20 (0.08)
Overdispersion	α	0.01 (0.01)	0.02 (0.03)	-0.05 (0.08)	-0.08 (0.05)
Compound symmetry	d	0.03 (0.02)	0.25 (0.10)	0.33 (0.17)	0.05 (0.04)
Correlation	ρ	0.32 (0.16)	0.62 (0.14)	0.48 (0.21)	0.18 (0.17)
		G-ab	G-ad	NG-ab	NG-ad
Intercept	ξ_0	3.01 (0.05)	1.89 (0.12)	0.61 (0.16)	1.25 (0.08)
Overdispersion	α	0.01 (0.01)	0.02 (0.03)	-0.06 (0.07)	-0.07 (0.05)
Compound symmetry	d	0.03 (0.02)	0.22 (0.09)	0.31 (0.16)	0.04 (0.04)
Correlation	ρ	0.32 (0.16)	0.60 (0.14)	0.48 (0.21)	0.16 (0.16)

the leaf were examined. On each face, trichomes of eight different types were counted. So these count responses were measured repeatedly (in a nested sampling scheme) on each plant.

For illustrative purposes we will use data from the P_2 population, which consists of 19 plants. Let y_{ij} be the number of a trichoma type counted in the j-th location of the i-th plant, where $i = 1, 2, \ldots, 19$ and $j = 1, 2, 3$, and the choice $\ln \lambda_{ij} = \xi_0 + b_i$, where ξ_0 is the overall effect and b_i is the random effect that captures the plant-level variability, assumed to be normally distributed with mean 0 and variance d.

The P_2 individuals are inbreeds, that is, they have the same genotype. Under homogeneous environmental conditions, it is expected that the overall variability between individuals with respect to a given characteristic is low and, therefore, this implies that underdispersion may be expected. As we have seen, such underdispersion in hierarchical data is captured by negative variance components. Indeed, while the correlation will likely remain positive, the α parameter might become negative. Table 1 presents descriptive statistics for the number of different trichomes in such a population.

In line with Section 4, we fitted the PGN to the tomatoes data, employing PQL, Laplace, and adaptive Gauss-Hermite quadrature, as implemented in the SAS procedure GLIMMIX. The results for the PQL method are displayed in Table 2. Through this method, convergence was achieved in most cases and standard errors for both fixed and random effects estimates were consistent. This numerical stability did not occur for the other estimation methods considered. Results from these are presented and briefly discussed in Appendix C. Recall that the PQL method is based on a relatively coarse Taylor series expansion, which is a factor to be taken into consideration next to the numerical stability.

The estimates for intraclass correlations displayed in Table 2 were obtained from (2) and their standard errors were calculated using the delta method [29]. Details on the calculation are presented in Appendix B.

In all cases, positive correlations were obtained. This is somewhat expected since measures within the same plant are more similar than measures between different plants. This leads to positive estimates for the variance component d. For the trichome NG-ad, the correlation is not significantly different from 0 and a modest underdispersion is observed, which leads to a negative estimate of α. A similar case is that of the NG-V-ad trichome.

In line with the calculated intraclass correlations and with the results presented in Section 3, positive correlations and no overdispersion implies negative α, given that ρ induces extra variance, which is removed by the negative estimate of the α component. This occurs for trichomes NG-V-ab and NG-ab.

Although the PQL method has shown the best performance among the three estimation methods, it failed while fitting the PGN model for trichomes G-I-ad, G-VI-ad, G-VII-ad, NG-II-III-ab and NG-II-III-ad. In all these cases convergence has not been achieved. Ideally, additional methods should be developed for negative variance component estimation, especially when interest lies in inference for such effects.

6. Concluding Remarks

Hierarchical data are common in empirical research. For the analysis of continuous data, the linear mixed model is a flexible tool while the generalized linear mixed model is commonly used to model non-Gaussian data. Beyond inferences on the fixed effects, such models allow inferences about variance components. While it seems natural to interpret the parameters in hierarchically formulated models from a purely hierarchical standpoint, there are practically relevant situations that cannot be captured by such an hierarchical interpretation. For example, the random-intercepts version of the linear mixed models induces a compound-symmetry marginal model with constant, non-negative correlation. In a context where cluster members experience correlation, however, negative correlations are not uncommon. These can be captured by the implied marginal model but not by the purely hierarchical formulation. In linear models, it is clear that what hierarchically is a variance, becomes a covariance marginally, thus allowing for negative correlation. These issues have been documented in the literature [7, 18, 20, 23].

In this work, we investigated non-Gaussian hierarchical data, where both overdispersion/underdispersion and correlation occur. There are various reasons why negative variance components may be needed: the occurrence of underdispersion, the occurrence of correlation, and the simultaneous occurrence of high correlation and low overdispersion. We studied these situations through the Poisson-Gamma-Normal (PGN) model. This model, developed for count data, accommodates hierarchies as well as overdispersion in the data, through normal and gamma distributed random effects, respectively. The variance components associated with these distributions were studied theoretically and a real data were used for illustration purposes. Negative estimates of variance components can occur, especially when the variability is low and there is an expected negative intraclass correlation perhaps due to intra-specific competition. Enforcing non-negative variance components and/or non-negative correlation, when data or design suggest that the reverse is likely may result in misleading conclusions. This is the case already for Gaussian data [18], but the problem exacerbates with non-Gaussian data, because of the non-linear relationships between mean, variance, and correlation functions. Of course, when interest is in marginal functions only, once might consider, for example, the use of generalized estimating equations (GEE) [8]. While these definitely have advantages, their moment-based nature precludes the use of certain inferential tools (e.g., likelihood ratio tests), and the estimation of certain fucntions, such as correlation and high-order association functions. Also, the non-likelihood basis of GEE leads to additional complications when data are incomplete.

Acknowledgment

The authors thank professor Wilson Roberto Maluf from Federal University of Lavras for having kindly provided the dataset used as illustration.

Funding

Geert Molenberghs gratefully acknowledges support from IAP research Network P7/06 of the Belgian Government (Belgian Science Policy). This work was partially supported by grants from CNPq, National Council for Scientific and Technological Development, Brazilian science funding agency.

References

[1] M. Aerts, H. Geys, G. Molenberghs, and L. Ryan, *Topics in Modelling of Clustered Data*, Chapman & Hall, London, 2002.
[2] R.R. Bahadur, *A representation of the joint distribution of responses to n dichotomous items*, in *Studies in Item Analysis and Prediction*, S. H, ed., Stanford Mathematical Studies in the Social Sciences, Vol. VI, Stanford University Press, 1961.
[3] J.G. Booth, G. Casella, H. Friedl, and J.P. Hobert, *Negative binomial loglinear mixed models*, Statistical Modelling 3 (2003), pp. 179–181.
[4] N.E. Breslow, *Extra-poisson variation in log-linear models*, Journal of the Royal Statistical Society 33 (1984), pp. 38–44.
[5] N.E. Breslow and D.G. Clayton, *Approximate inference in generalized linear mixed models*, Journal of American Statistical Association 88 (1993), pp. 9–25.
[6] T. Britton, *Tests to detect clustering of infected individuals within families*, Biometrics 53 (1997), pp. 98–109.
[7] H. Chernoff, *On the distribution of the likelihood ratio*, Annals of Mathematical Statistics 25 (1954), pp. 573–578.
[8] P.J. Diggle, P.J. Heagerty, K.Y. Liang, and S.L. Zeger, *Analysis of Longitudinal Data*, Oxford University Press, Oxford, 2002.
[9] J. Hinde and C.G.B. Demétrio, *Overdispersion: Models and estimation*, Computational Statistics and Data Analysis 27 (1998a), pp. 151–170.
[10] J. Hinde and C.G.B. Demétrio, *Overdispersion: Models and Estimation.*, Sinape, São Paulo, 1998b.
[11] G. Kalema, S. Iddi, and G. Molenberghs, *The combined model: A tool for simulating correlated counts with overdispersion*, Communications in Statistics 00 (2014), pp. 000–000.
[12] D.A. Kenny, L. Mannetti, A. Pierro, S. Livi, and D.A. Kashy, *The statistical analysis of data from small groups*, Journal of Personality and Social Psychology 83 (2002), pp. 126–137.
[13] J.F. Lawless, *Negative binomial and mixed poisson regression*, The Canadian Journal of Statistics 15 (1987), pp. 209–225.
[14] P. Mccullagh and J.A. Nelder, *Generalized Linear Models*, Chapman & Hall, London, 1989.
[15] G. Molenberghs and M.G. Kenward, *Semi-parametric marginal models for hierarchical data and their corresponding full models*, Computational Statistics and Data Analysis 54 (2010), pp. 585–597.
[16] G. Molenberghs and G. Verbeke, *Models for Discrete Longitudinal Data*, Springer, New York, 2005.
[17] G. Molenberghs and G. Verbeke, *Likelihood ratio, score, and wald tests in a constrained parameter space*, The American Statistician 61 (2007), pp. 1–6.
[18] G. Molenberghs and G. Verbeke, *A note on a hierarchical interpretation for negative variance components*, Statistical Modelling 11 (2011), pp. 389–408.

[19] G. Molenberghs, G. Verbeke, and C.G.B. Demétrio, *An extended random-effects approach to modeling repeated, overdispersed count data*, Lifetime Data Analysis 13 (2007), pp. 513–531.
[20] J.A. Nelder, *The interpretation of negative components of variance*, Biometrika 41 (1954), pp. 544–548.
[21] J.A. Nelder and R.W.M. Wedderburn, *Generalized linear models*, Journal of the Royal Statistical Society 135 (1972), pp. 370–384.
[22] R.L. Prentice, *Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors*, Journal of the American Statistical Association 81 (1986), pp. 321–327.
[23] A. Pryseley, C. Tchonlafi, G. Verbeke, and G. Molenberghs, *Estimating negative variance components from gaussian and non-gaussian data: A mixed models approach*, Computational Statistics & Data Analysis 55 (2011), pp. 1071–1085.
[24] T. Rakhmawati, G. Molenberghs, G. Verbeke, and C. Faes, *Local influence diagnostics for generalized linear mixed models with overdispersion*, Submitted for publication .
[25] D.M. Smith and M.S. Ridout, *Remark as r93: A remark on algorithm as 189: Maximum likelihood estimation of the parameters of the beta binomial distribution*, Journal of the Royal Statistical Society, Series C 44 (1995), pp. 545–547.
[26] T. Vangeneugden, G. Molenberghs, G. Verbeke, and C.G.B. Demétrio, *Marginal correlation from an extended random-effects model for repeated and overdispersed counts*, Journal of Applied Statistics 38 (2011), pp. 215–232.
[27] G. Verbeke and G. Molenberghs, *Linear mixed models for longitudinal data*, Springer-Verlag, New York, 2000.
[28] G. Verbeke and G. Molenberghs, *The use of score tests for inference on variance components*, Biometrics 59 (2003), pp. 254–262.
[29] A.H. Welsh, *Aspects of Statistical Inference*, Wiley Series in Probability and Statistics, Wiley, Hoboken, NJ, 2011.

Appendix A. Model Elements for the Poisson-Gamma-Normal and Poisson-Normal

The mean and variance expressions for the PGN (1) were presented by [19]. The mean vector $\boldsymbol{\mu}_i = E(\boldsymbol{Y}_i)$ has components

$$\mu_{ij} = \phi \exp\left(\boldsymbol{x}_{ij}'\boldsymbol{\xi} + \frac{1}{2}\boldsymbol{z}_{ij}' D \boldsymbol{z}_{ij}\right) \tag{A1}$$

and the variance-covariance matrix is given by

$$\text{Var}(\boldsymbol{Y}_i) = M_i + M_i(P_i - J_{n_i})M_i, \tag{A2}$$

where M_i is a diagonal matrix with the vector $\boldsymbol{\mu}_i$ along the diagonal and the $(j,k)^{th}$ element of P_i equals

$$p_{i,jk} = \exp\left(\frac{1}{2}\boldsymbol{z}_{ij}' D \boldsymbol{z}_{ik}\right) \frac{\sigma_{i,jk} + \phi_{ij}\phi_{ik}}{\phi_{ij}\phi_{ik}} \exp\left(\frac{1}{2}\boldsymbol{z}_{ik}' D \boldsymbol{z}_{ij}\right).$$

Note that $\sigma_{i,jk}$ is the (j,k)th element of Σ_i.

[26] presented closed-form expression for the correlation function for the general case of the combined model and its specific cases. Considering the combined model with arbitrary fixed- and random-effects structures, the variance, deriving from (A2) equals:

$$\begin{aligned}\text{Var}(Y_{ij}) = \phi_{ij} \exp\left(\boldsymbol{x}_{ij}'\boldsymbol{\xi} + \tfrac{1}{2}\boldsymbol{z}_{ij}' D \boldsymbol{z}_{ij}\right) + \sigma_{i,jj} \exp(2\boldsymbol{x}_{ij}'\boldsymbol{\xi} + 2\boldsymbol{z}_{ij}' D \boldsymbol{z}_{ij}) \\ +\phi_{ij}^2 \exp(2\boldsymbol{x}_{ij}'\boldsymbol{\xi} + \boldsymbol{z}_{ij}' D \boldsymbol{z}_{ij})[\exp(\boldsymbol{z}_{ij}' D \boldsymbol{z}_{ij}) - 1].\end{aligned}$$

Likewise, the covariance can be written as:

$$\mathrm{Cov}(Y_{ij}, Y_{ik}) = \phi_{ij} \exp\left(\boldsymbol{x}'_{ij}\boldsymbol{\xi} + \frac{1}{2}\boldsymbol{z}'_{ij}D\boldsymbol{z}_{ij}\right) \times \left\{\left(\frac{\sigma_{i,jk}}{\phi_{ij}\phi_{ik}} + 1\right)\exp\left[\frac{1}{2}(\boldsymbol{z}'_{ij}D\boldsymbol{z}_{ik} + \boldsymbol{z}'_{ik}D\boldsymbol{z}_{ij})\right] - 1\right\} \times\phi_{ik}\exp\left(x'_{ik}\boldsymbol{\xi} + \frac{1}{2}\boldsymbol{z}'_{ik}D\boldsymbol{z}_{ik}\right).$$

The correlation between two measures j and k on the same cluster (subject) i then is:

$$\mathrm{Corr}(Y_{ij}, Y_{ik}) = \frac{\mathrm{Cov}(Y_{ij}, Y_{ik})}{\sqrt{\mathrm{Var}(Y_{ij})\mathrm{Var}(Y_{ik})}}.$$

These expressions also produce their simplified counterparts for important special cases. For the P-N, when only normal random effects are present, the mean vector components slightly simplify:

$$\mu_{ij} = \exp\left(\boldsymbol{x}'_{ij}\boldsymbol{\xi} + \frac{1}{2}\boldsymbol{z}'_{ij}D\boldsymbol{z}_{ij}\right),$$

and the variance-covariance matrix is

$$\mathrm{Var}(\boldsymbol{Y}_i) = M_i + M_i[\exp(Z_iDZ'_i) - J_{n_i}]M_i.$$

Similar logic as in the PGN leads to the correlation expression for this special case, considering:

$$\mathrm{Var}(Y_{ij}) = \exp\left(\boldsymbol{x}'_{ij}\boldsymbol{\xi} + \frac{1}{2}\boldsymbol{z}'_{ij}D\boldsymbol{z}_{ij}\right) + \exp(2\boldsymbol{x}'_{ij}\boldsymbol{\xi} + \boldsymbol{z}'_{ij}D\boldsymbol{z}_{ij})[\exp(\boldsymbol{z}'_{ij}D\boldsymbol{z}_{ij}) - 1],$$

and

$$\mathrm{Cov}(Y_{ij}, Y_{ik}) = \exp\left(\boldsymbol{x}'_{ij}\boldsymbol{\xi} + \frac{1}{2}\boldsymbol{z}'_{ij}D\boldsymbol{z}_{ij}\right)\left[\exp(\boldsymbol{z}'_{ij}D\boldsymbol{z}_{ik}) - 1\right]\exp\left(x'_{ik}\boldsymbol{\xi} + \frac{1}{2}\boldsymbol{z}'_{ik}D\boldsymbol{z}_{ik}\right).$$

It is useful to recall that [19] derived the marginal joint distribution of the PGN, taking the form:

$$P(\boldsymbol{Y}_i = \boldsymbol{y}_i) = \sum_{\boldsymbol{t}}\left[\prod_{j=1}^{n_i}\binom{y_{ij}+t_j}{y_{ij}}\cdot\binom{\alpha_j + y_{ij} + t_j - 1}{\alpha_j - 1}\cdot(-1)^{t_j}\cdot\alpha_j^{-y_{ij}-t_j}\right] \times\exp\left(\sum_{j=1}^{n_i}(y_{ij}+t_j)\boldsymbol{x}'_{ij}\boldsymbol{\xi}\right) \times\exp\left\{\frac{1}{2}\left[\sum_{j=1}^{n_i}(y_{ij}+t_j)\boldsymbol{z}'_{ij}\right]D\left[\sum_{j=1}^{n_i}(y_{ij}+t_j)\boldsymbol{z}_{ij}\right]\right\}. \quad \text{(A3)}$$

In the above equation, the vector-valued index $\boldsymbol{t} = (t_1, \ldots, t_{n_i})$ ranges over all non-negative integer vectors.

Appendix B. Precision Estimation of Correlation in the Poisson-Gamma-Normal Model

Standard errors for the intraclass correlations in the Poisson-Gamma-Normal (2) model were obtained using the delta method. We consider $\rho = \zeta_N/\zeta_D$, apply the delta method first to numerator and denominator, and then to the ratio. Assume W is the variance-covariance matrix of $\boldsymbol{\zeta} = \begin{pmatrix} \zeta_N \\ \zeta_D \end{pmatrix}$, then $\mathrm{Var}(\zeta) \cong T'WT$, where

$$T = \frac{\partial \boldsymbol{\zeta}}{\partial(\zeta_N, \zeta_D)} = \begin{pmatrix} 1/\zeta_D \\ -\zeta_N/\zeta_D^2 \end{pmatrix}.$$

To estimate W, we also use the delta method. At this stage, let $\boldsymbol{\phi}$ be the parameter vector relevant for ζ_N and ζ_D and V be its variance-covariance matrix. Then, $W \cong S'VS$, with $S = \dfrac{\partial(\zeta_N, \zeta_D)}{\partial \boldsymbol{\phi}}$.

The S matrix is

$$\begin{pmatrix} \dfrac{\partial \zeta_N}{\partial d} & \dfrac{\partial \zeta_D}{\partial d} \\ \dfrac{\partial \zeta_N}{\partial \alpha} & \dfrac{\partial \zeta_D}{\partial \alpha} \\ \dfrac{\partial \zeta_N}{\partial \xi_0} & \dfrac{\partial \zeta_D}{\partial \xi_0} \end{pmatrix},$$

where

$$\frac{\partial \zeta_N}{\partial d} = \exp\left(\xi_0 + \frac{1}{2}d\right)\left\{\frac{1}{2}[\exp(d) - 1] + \exp(d)\right\},$$

$$\frac{\partial \zeta_D}{\partial d} = \exp\left(\xi_0 + \frac{1}{2}d\right)\left\{\frac{1}{2}[3\exp(d)\alpha + 3\exp(d) - 1]\right\},$$

$$\frac{\partial \zeta_N}{\partial \alpha} = 0,$$

$$\frac{\partial \zeta_D}{\partial \alpha} = \exp\left(\xi_0 + \frac{3}{2}d\right),$$

$$\frac{\partial \zeta_N}{\partial \xi_0} = \exp\left(\xi_0 + \frac{1}{2}d\right)[\exp(d) - 1],$$

$$\frac{\partial \zeta_D}{\partial \xi_0} = \exp\left(\xi_0 + \frac{1}{2}d\right)[\exp(d)\alpha + \exp(d) - 1].$$

The correlation expression depend on estimates of fixed and random effects, that is, d, α and ξ_0, and the V matrix should contain all their variances and covariances. However, the SAS procedure GLIMMIX provides a variance-covariance matrix for the random effects and another variance-covariance matrix for the fixed effects, separately. Then, the covariances between random and fixed estimates in the V matrix were set to zero, which

Table C1. Tomatoes study. Parameter estimates (standard errors) for the PGN, considering random intercepts, for some trichome counts. The estimation method used was Laplace.

Effect	Par.	G-IV-ab	G-IV-ad	NG-V-ab	NG-V-ad
Intercept	ξ_0	3.00 (0.05)	1.82 (0.13)	0.66 (0)	1.19 (0)
Overdispersion	α	0.01 (0.01)	0.02 (0.03)	-0.05 (.)	-0.08 (.)
Compound symmetry	d	0.02 (0.02)	0.24 (0.10)	0.33 (.)	0.05 (.)
		G-ab	G-ad	NG-ab	NG-ad
Intercept	ξ_0	3.00 (0.05)	1.87 (0.12)	0.56 (0.32)	1.26 (0.07)
Overdispersion	α	0.01 (0.01)	0.02 (0.03)	1.89×10^{-7} (0.01)	0.00 (0.07)
Compound symmetry	d	0.02 (0.02)	0.21 (0.09)	0.28 (0.28)	-0.11 (.)

Table C2. Tomatoes study. Parameter estimates (standard errors) for the PGN, considering random intercepts, for some trichome counts. The estimation method used was adaptive Gauss-Hermite quadrature.

Effect	Par.	G-IV-ab	G-IV-ad	NG-V-ab	NG-V-ad
Intercept	ξ_0	3.00 (0.05)	1.82 (0.13)	0.50 (0.17)	1.19 (0.08)
Overdispersion	α	0.01 (0.01)	0.02 (0.03)	1.28×10^{-7} (.)	9.72×10^{-8} (.)
Compound symmetry	d	0.02 (0.02)	0.24 (0.10)	0.30 (0.17)	0.02 (0.04)
		G-ab	G-ad	NG-ab	NG-ad
Intercept	ξ_0	3.00 (0.05)	1.87 (0.12)	0.56 (0.17)	1.25 (0.08)
Overdispersion	α	0.01 (0.01)	0.02 (0.03)	8.67×10^{-8} (.)	1.52×10^{-7} (.)
Compound symmetry	d	0.02 (0.02)	0.21 (0.09)	0.27 (0.16)	0.01 (0.03)

are not very different from the true values.

When there are covariate effects, additional parameters, say ξ, need to be considered in the method. In such case and due to the linearity in the linear predictor of GLMs, the derivatives will be the same as those with respect to ξ_0, but depending on the covariate values.

Appendix C. Results from Other Estimation Methods in the Poisson-Gamma-Normal Model

In Tables C1 and C2 the estimates for the PGN, considering Laplace and adaptive Gauss-Hermite quadrature as estimation methods, are displayed. When compared to the PQL results, both methods showed poor performance for this application.

In some cases, the Laplace method failed to estimate the random effects and their standard errors. Specifically for trichomes NG-V-ab and NG-V-ad, the convergence was achieved, but the covariance matrix is the zero matrix. For trichome NG-ad, the convergence was also achieved, but the estimated G matrix is not positive definite.

In a simulation study, [23] investigated the performance of the PQL and Laplace methods in the face of negative variance components for binary clustered data, by means of generalized linear mixed models. In their study, Laplace approximations were more accurate than PQL and convergence was easier to reach, different from what we noticed in this application for count data and using the PGN.

Although quadrature methods are generally considered the most accurate ones, they adopt a hierarchical perspective and cannot be used when negative variance components are allowed. Confirming this, the method failed in all situations where negative variance components are expected.

Appendix D. SAS code

The Poisson-Gamma-Normal model can be fitted using the SAS procedure NLMIXED [19]. It allows the user to implement his own likelihood, using the so-called *general like-*

lihood. However, such procedure makes use of adaptive Gaussian quadrature as approximation method of the integral over the random effects. This method adopts a hierarchical perspective and does not allow for negative estimates of the variance components.

Alternatively, we have used the SAS procedure GLIMMIX and the built-in negative binomial likelihood. The normal random effects are included using the RANDOM statement, which leads to the PGN model. The nobound option requests the removal of boundary constraints on covariance and scale parameters, allowing variance components estimates to be negative. Note that this option can not be used for adaptive quadrature estimation.

```
proc glimmix data=tricp2 method=laplace nobound ASYCOV;
TITLE ‘Laplace PGN model: NG_V_ad - P2’;
CLASS plant;
MODEL NG_V_ad =/ s link=log dist=NB COVB;
RANDOM int / subject=plant;
run;

proc glimmix data=tricp2 method=rspl nobound ASYCOV;
TITLE ‘PQL PGN model: NG_V_ad - P2’;
CLASS plant;
MODEL NG_V_ad =/ s link=log dist=NB COVB;
RANDOM int / subject=plant;
run;

proc glimmix data=tricp2 method=quad;
TITLE ‘QUAD PGN model: NG_V_ad - P2’;
CLASS plant;
MODEL NG_V_ad =/ s link=log dist=NB COVB;
RANDOM int / subject=plant;
run;
```