Coefficients of Determination for Mixed-Effects Models

Dabao Zhang

The coefficient of determination is well defined for linear models and its extension is long wanted for mixed-effects models in agricultural, biological, and ecological research. We revisit its extension to define measures for proportions of variation explained by the whole model, fixed effects only, and random effects only. We propose to calculate unexplained variations conditional on individual random and/or fixed effects so as to keep individual heterogeneity brought by available predictors. While these measures were naturally defined for linear mixed models, they can be defined for a generalized linear mixed model using a distance measured along its variance function, accounting for its heteroscedasticity. We demonstrate the promising performance and utility of our proposed methods via simulation studies as well as applications to real data sets in agricultural and ecological studies.

Key Words: Exponential family distribution; Generalized linear mixed model; Linear mixed model; Quasi-model; $R^2$; Variance function.

1. INTRODUCTION

Mixed-effects models are widely used in agricultural, biological, and ecological research to understand the variation components of a response variable (Gbur et al. 2012; Zuur et al. 2009). The law of total variance provides a theoretical basis for defining the coefficient of determination, also known as $R^2$, for linear models and sheds light on defining similar measures for mixed-effects models.

Suppose that a linear mixed model (McCulloch et al. 2008) is considered for observed response variable $Y_{ij}$ from the $j$th individual inside the $i$th cluster, with $j = 1, \ldots, n_i$ and $i = 1, \ldots, m$. In general, we write the corresponding linear mixed model as,

$$Y_{ij} = \eta^F_{ij} + \eta^R_{ij} + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2),$$

(1)
where $\eta^F_{ij}$ and $\eta^R_{ij}$, respectively, summarize all fixed and random effects on the response variable with

$$\eta^R_{ij} \mid \tau^2_{ij} \sim N(0, \tau^2_{ij}).$$

Note that $\{\epsilon_{ij}, j = \ldots, \ldots, m\}$ may be an autocorrelated series in, for example, longitudinal studies (Laird and Ware 1982). For construction simplicity, we will not emphasize such correlation as it does not affect the unbiasedness of estimated variances in the below although the estimation may not be optimal.

Similar to linear models, the law of total variance provides a clear path to extend $R^2$ for linear mixed models, measuring the proportion of variation in the dependent variable modeled by fixed effects, random effects, or both (Xu 2003; Nakagawa and Schielzeth 2013; Nakagawa et al. 2017; Jaeger et al. 2017, 2019). However, unlike usual calculation that heavily relies on estimated error variance $\sigma^2$, we instead propose to calculate unexplained variations conditional on individual random and/or fixed effects so as to keep individual heterogeneity brought by available predictors which also shed lights on their extensions to more general models.

The inherent heteroscedasticity makes it difficult to properly define $R^2$ for generalized linear mixed models. Indeed, such heteroscedasticity also challenges the proper definition of $R^2$ for generalized linear models (McCullagh and Nelder 1989). Therefore, many different measures have been proposed to define $R^2$ for generalized linear models from different aspects of view (Cameron and Windmeijer 1997; Cox and Snell 1989; Maddala 1983; Magee 1990; Nagelkerke 1991; Zhang 2017). However, it is difficult to extend these measures to account for random effects involved in generalized linear mixed models.

A common strategy to define $R^2$ for generalized linear mixed models, adopted by Nakagawa and Schielzeth (2013) and Nakagawa et al. (2017), is to recognize the linear function presented by the link function $g(\cdot)$, i.e.,

$$g(E[Y_{ij} \mid \eta^F_{ij}, \eta^R_{ij}]) = \eta^F_{ij} + \eta^R_{ij},$$

and instead construct $R^2$ for a transformed linear mixed model for $g(Y_{ij})$. However, such measures rely on the specified link function and even the approximation method used to calculate the error variance (Nakagawa et al. 2017). On the other hand, the link function does not necessarily provide homoscedastic variance on the error term, which is the primary challenge in extending classical $R^2$ from linear models to generalized linear models, although it may describe a linear relationship of all effects on $g(E[Y_{ij} \mid \eta^F_{ij}, \eta^R_{ij}])$, and present additive variance components in $g(Y_{ij})$.

A critical concern of defining $R^2$ based on a transformed linear model is that proportions of different variance components in $g(Y_{ij})$ may not represent the genuine proportions of different variance components in $Y_{ij}$. For example, it is well known that the latent linear model of a probit model may present a very high $R^2$, but the binomial response still holds a lot of uncertainty, as illustrated by the probit model $Y = 1[Y^* > 0]$ with $Y^* | X \sim N(X, 1)$ and $X/10 \sim Bernoulli(0.5)$ in which the theoretical $R^2$ for $Y^*$ is approximately 0.96 but the classical $R^2$ for $Y$ is 1/3. On the other hand, if $X/6 + .5 \sim Bernoulli(0.05)$, then the theoretical $R^2$ for $Y^*$ is approximately 0.63 but the classical $R^2$ for $Y$ is approximately 0.97.
Such inconsistency is well recognized in the study of genetic heritability, see Dempster and Lerner (1950).

Recently, Zhang (2017) has showed that quantifying the variation change along the variance function can measure explained variation of a heteroscedastic response variable and hence proposed to define a variance function-based $R^2$. Unlike other likelihood-based measures, such a measure neither overstates the proportion of explained variation, nor demands the specification of likelihood functions. While it only requires specification of the link function and variance function, it reduces to classical $R^2$ for linear models so it is conceptually consistent with classical $R^2$. We thus use the variance function-based measures to define $R^2$ for generalized linear mixed models.

In the next section, we revisit the definition of coefficients of determination to account for explained variation by fixed effects, random effects, or both in linear mixed models, and propose our calculation emphasizing individual heterogeneity. In Sect. 3, we extend these measures for generalized linear mixed models by using the variance function-based distance (Zhang 2017). Our simulation studies to compare different measures are shown in Sect. 4. We also compare these measures in Sect. 5 by applying them to three sets of real data from agricultural and ecological studies, and conclude with a discussion in Sect. 6.

2. LINEAR MIXED MODELS

For the linear mixed model (1), we can follow the law of total variance and define the proportion of variation in $Y_{ij}$ modeled by the fixed effects as

$$
\rho^2_F = 1 - \frac{E[\text{var}(Y_{ij} | \eta_{ij}^F)]}{\text{var}(Y_{ij})} = 1 - \frac{E[(Y_{ij} - E[Y_{ij} | \eta_{ij}^F])^2]}{E[(Y_{ij} - E[Y_{ij}])^2]}.
$$

(3)

With $\eta_{ij}^F$ estimated by $\hat{\eta}_{ij}^F$ and $E[Y_{ij}]$ estimated by the sample average $\bar{Y}_{..}$, we have the following estimate of $\rho^2_F$,

$$
R^2_F = 1 - \frac{\sum_{i,j}(Y_{ij} - \hat{\eta}_{ij}^F)^2}{\sum_{i,j}(Y_{ij} - \bar{Y}_{..})^2}.
$$

(4)

Since

$$
E[(Y_{ij} - E[Y_{ij} | \eta_{ij}^F])^2] = E[\tau_{ij}^2] + \sigma^2,
$$

we may also take estimated variance components to construct $R^2_F$ to estimate $\rho^2_F$, see, e.g., Xu (2003), Nakagawa and Schielzeth (2013); Nakagawa et al. (2017), and Jaeger et al. (2017).

The proportion of variation in $Y_{ij}$ modeled by both fixed and random effects can be similarly defined as

$$
\rho^2_M = 1 - \frac{E[\text{var}(Y_{ij} | \eta_{ij}^F, \eta_{ij}^R)]}{\text{var}(Y_{ij})} = 1 - \frac{E[(Y_{ij} - E[Y_{ij} | \eta_{ij}^F, \eta_{ij}^R])^2]}{E[(Y_{ij} - E[Y_{ij}])^2]}.
$$

(5)
With \( \text{var}(Y_{ij} \mid \eta_{ij}^F, \eta_{ij}^R) = \sigma^2 \), it is tempting to estimate \( \rho_M^2 \) on the basis of \( \rho_M^2 = 1 - \sigma^2 / \text{var}(Y_{ij}) \) as in Xu (2003), Nakagawa and Schielzeth (2013), Nakagawa et al. (2017), Jaeger et al. (2017), and Ives (2019). However, the total variation in the response variable is described by \( \text{SST} = \sum_{i,j} (Y_{ij} - \bar{Y})^2 \), and we thus would rather to calculate the total unexplained variation by emphasizing individual heterogeneity of \( \tau_{ij} \) and the contribution of individual observation, which also helps the extension to generalized linear models.

Note that,

\[
E\left[ (Y_{ij} - E(Y_{ij} \mid \eta_{ij}^F, \eta_{ij}^R))^2 \right] = E\left[ E\left( (Y_{ij} - E(Y_{ij} \mid \eta_{ij}^F, \eta_{ij}^R))^2 \mid Y_{ij}, \eta_{ij}^F, \tau_{ij}^2, \sigma^2 \right) \right],
\]

implying that each observation contributes \( E\left( (Y_{ij} - E(Y_{ij} \mid \eta_{ij}^F, \eta_{ij}^R))^2 \mid Y_{ij}, \eta_{ij}^F, \tau_{ij}^2, \sigma^2 \right) \) with observed value \( Y_{ij} \) and estimable parameters in \( \eta_{ij}^F, \tau_{ij}^2, \) and \( \sigma^2 \). That is, the expectation is on the random variable \( \eta_{ij}^R \) conditional on the observed values and these estimable parameters. With the conditional distribution

\[
\eta_{ij}^R \mid Y_{ij}, \eta_{ij}^F, \tau_{ij}^2, \sigma^2 \sim N\left( \frac{\tau_{ij}^2}{\sigma^2 + \tau_{ij}^2} (Y_{ij} - \eta_{ij}^F), \frac{\sigma^2 \tau_{ij}^2}{\sigma^2 + \tau_{ij}^2} \right),
\]

we have

\[
E\left[ (Y_{ij} - E(Y_{ij} \mid \eta_{ij}^F, \eta_{ij}^R))^2 \mid Y_{ij}, \eta_{ij}^F, \tau_{ij}^2, \sigma^2 \right] = \left( \frac{\sigma^2}{\sigma^2 + \tau_{ij}^2} \right)^2 (Y_{ij} - \eta_{ij}^F)^2 + \frac{\sigma^2 \tau_{ij}^2}{\sigma^2 + \tau_{ij}^2}.
\]

Therefore, \( \rho_M^2 \) will be estimated by

\[
R_M^2 = 1 - \frac{\sum_{i,j} \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{\tau}_{ij}^2} \left\{ \hat{\tau}_{ij}^2 + \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{\tau}_{ij}^2} (Y_{ij} - \hat{\eta}_{ij}^F)^2 \right\}}{\sum_{i,j} (Y_{ij} - \bar{Y})^2}.
\]  \( (6) \)

The proportion of variation in \( Y_{ij} \) modeled by random effects can be simply defined as

\[
\rho_R^2 = \rho_M^2 - \rho_F^2 = \frac{E((Y_{ij} - E(Y_{ij} \mid \eta_{ij}^F, \tau_{ij}^2))^2) - E((Y_{ij} - E(Y_{ij} \mid \eta_{ij}^F, \eta_{ij}^R))^2)}{\text{var}(Y_{ij})},
\]  \( (7) \)

and can be estimated as

\[
R_R^2 = R_M^2 - R_F^2.
\]  \( (8) \)

Such a simple definition on \( \rho_R^2 \) assumes that \( \rho_F^2 \) has a priority over \( \rho_R^2 \), that is, when fixed and random effects overlapped, we would rather know the proportion of variation explained by available fixed effects, and therefore use \( \rho_R^2 \) to measure the proportion of variable additionally explained by random effects.
Table 1. Difference between the Euclidean distance and $d_V(\cdot, \cdot)$

| Distribution   | $V(\mu)$ | $\mu_1$ | $\mu_2$ | $Y$ | $(Y-\mu)^2$ | $d_V(Y, \mu_2)$ | $d_V(Y, \mu_1)$ |
|---------------|----------|---------|---------|-----|-------------|----------------|----------------|
| Binomial      | $\mu (1 - \mu)$ | 0.5     | 0.75    | 1   | 0.25        | 0.2991         | 0.25           |
| Poisson       | $\mu$    | 1       | 2       | 3   | 0.25        | 0.25           | 0.25           |
| Gamma         | $\mu^2$  | 1       | 2       | 3   | 0.25        | 0.3805         | 0.25           |
| Inverse Gaussian | $\mu^{3/2}$ | 0       | 1       | 2   | 0.25        | 0.6735         | 0.25           |

3. GENERALIZED LINEAR MIXED MODELS

For the generalized linear mixed model (2), the variance of $Y_{ij}$, given both fixed and random effects, can be specified via a dispersion parameter $\phi$ and a known variance function $V(\cdot)$, i.e.,

$$\text{var}(Y_{ij} \mid \eta^F_{ij}, \eta^R_{ij}) = \phi V(g^{-1}(\eta^F_{ij} + \eta^R_{ij})).$$

In general, as long as the mean $g^{-1}(\eta^F_{ij} + \eta^R_{ij})$ can be modeled well and linked appropriately to a set of predictors, a generalized linear model with known variance function $V(\cdot)$ can be investigated for the utility of the involved predictors.

The variance function describes the effect of the mean on the variation of the response variable besides the dispersion parameter. For a response variable with its mean moving from $a$ to $b$, its variation changes accordingly along the variance function from $\phi V(a)$ to $\phi V(b)$. Taking advantage of relationship between mean and variance defined by the variance function, we measure the variation change along the variance function (Zhang 2017). That is, the variation change of the response variable should be measured using, instead of $(a-b)^2$, the squared length of the variance function $V(\cdot)$ between $V(a)$ and $V(b)$,

$$d_V(a, b) = \left\{ \int_a^b \sqrt{1 + [V'(t)]^2} dt \right\}^2.$$ 

As shown in Table 1, $d_V(\cdot, \cdot)$ may dramatically differ than the Euclidean distance when the variance function is nonlinear.

Our definition of $R^2$ for generalized linear mixed models will proceed by replacing the Euclidean distance by the above manifold distance along the variance function.

Replacing $(a-b)^2$ with $d_V(a, b)$ in (3), we can define the proportion of variation in $Y_{ij}$ modeled by the fixed effects as

$$\rho^2_F = 1 - \frac{E[d_V(Y_{ij}, E[Y_{ij} \mid \eta^F_{ij}])]}{E[d_V(Y_{ij}, E[Y_{ij}])]}.$$ (9)

Recall that for the linear mixed model in (1), $E[Y_{ij} \mid \eta^F_{ij}]$ is estimated via $\hat{\eta}^F_{ij}$ from the fitted mixed model. Suppose that, for the generalized linear mixed model we similarly obtained $\hat{\eta}^F_{ij}$ from the fitted model in (2). Unlike for the linear mixed model case, this may result in
an estimate different from \( \hat{\eta}_{ij}^F \), which is obtained from fitting the model in (2) without the random effect \( \eta_{ij}^R \). Since \( \hat{\eta}_{ij}^F \), instead of \( \hat{\eta}_{ij}^F \), represents better the contribution of fixed effects to explaining the variation in the response variable, we propose to estimate \( \rho^2_F \) by

\[
R^2_F = 1 - \frac{\sum_{i,j} dV(Y_{ij}, g^{-1}(\hat{\eta}_{ij}^F))}{\sum_{i,j} dV(Y_{ij}, \bar{Y})}. \tag{10}
\]

With (5), we can similarly define the proportion of variation in \( Y_{ij} \) modeled by both fixed and random effects as

\[
\rho^2_M = 1 - \frac{E[dV(Y_{ij}, E[Y_{ij} | \eta_{ij}^F, \eta_{ij}^R])]}{E[dV(Y_{ij}, E[Y_{ij} | \eta_{ij}])]} = 1 - \frac{E[dV(Y_{ij}, g^{-1}((\hat{\eta}_{ij}^F) + \eta_{ij}^R))]}{E[dV(Y_{ij}, E[Y_{ij} | \eta_{ij}])]} . \tag{11}
\]

We can simply plug in \( \hat{\eta}_{ij}^F \) and \( \hat{\eta}_{ij}^R \), estimates of \( \eta_{ij}^F \) and \( \eta_{ij}^R \), respectively, from fitting (2), to estimate \( \rho^2_M \),

\[
R^2_M = 1 - \frac{\sum_{i,j} dV(Y_{ij}, g^{-1}((\hat{\eta}_{ij}^F) + \hat{\eta}_{ij}^R))}{\sum_{i,j} dV(Y_{ij}, \bar{Y})} . \tag{12}
\]

The proportion of variation in \( Y_{ij} \) modeled by random effects can be defined as

\[
\rho^2_R = \rho^2_M - \rho^2_F = \frac{E[dV(Y_{ij}, E[Y_{ij} | \eta_{ij}^F])]}{\text{var}(Y_{ij})} - E[dV(Y_{ij}, E[Y_{ij} | \eta_{ij}^F, \eta_{ij}^R])] . \tag{13}
\]

Accordingly, we can calculate the \( R^2 \) for random effects using \( R^2_R = \rho^2_M - R^2_F \) with \( R^2_M \) and \( R^2_F \) calculated in (12) and (10), respectively. Such a calculation of \( R^2_R \) based on the above definition of \( \rho^2_R \) follows the fact that we usually want to evaluate the contribution due to random effects by removing those attributable to the fixed effects. However, in the case that we prioritize the contribution due to the random effects over the contribution due to the fixed effects, we can accordingly first define \( \rho^2_R \) and \( R^2_R \) and then define \( \rho^2_F \) and \( R^2_F \) based on other measures.

Similar to classical \( R^2 \) defined for linear regression models, both \( R^2_F \) and \( R^2_M \) defined as above may increase as more predictors are included into the underlying model. We can define the adjusted \( R^2_F \) and adjusted \( R^2_M \) by dividing the numerators and denominators in (10) and (12) with proper degrees of freedom. The adjusted \( R^2_R \) can be accordingly calculated via the adjusted \( R^2_F \) and adjusted \( R^2_M \).

4. SIMULATION STUDIES

4.1. LINEAR MIXED MODELS

For linear mixed models, we compared the performance of our proposed \( R^2 \) to those proposed by Xu (2003) and Nakagawa et al. (2017), with a total of 1000 data sets simulated from each model under investigation. Each data set has a total of 200 random samples, evenly clustered inside \( m \) groups with \( m = 10 \) and 50, respectively. A binary covariate \( X_1 \)
was generated for each observation, with half observations within the same group taking 1 and the other half taking −1. A second variable $X_2$ was generated from the standard normal distribution, independent of $X_1$ and the response variable. The $j$th response value inside the $i$th cluster, i.e., $y_{ij}$, was generated by

$$y_{ij} = \mu_i + x_{1ij} \beta + \epsilon_{ij},$$

where $x_{1ij}$ is the corresponding value of the binary covariate $X_1$, the random effect $\mu_i \overset{iid}{\sim} N(0, 1)$, and $\epsilon_{ij} \overset{iid}{\sim} N(0, 1)$.

For each data set, we fit different models, i.e., regressing against only $X_1$ or only $X_2$, with the maximum likelihood method, and then estimate both $\rho^2_M$ and $\rho^2_F$ with different approaches as shown in Fig. 1. Note that, when regressing against only $X_1$, we have

$$\rho^2_M = \frac{\beta^2 + 1}{\beta^2 + 2}, \quad \rho^2_F = \frac{\beta^2}{\beta^2 + 2}.$$
While Xu (2003) only estimates $\rho_M^2$ but not $\rho_F^2$, it may severely overestimate $\rho_M^2$ especially when the number of groups is large. Indeed, the bias can be more than 0.15 when there are 100 groups (the results are not shown). Our proposed $R_M^2$ and the methods by Nakagawa et al. (2017) estimate $\rho_M^2$ and $\rho_F^2$ well; however, there is slight bias especially when the number of groups is small. When regressing against only $X_2$, we have

$$
\rho_M^2 = \frac{1}{\beta^2 + 2}, \quad \rho_F^2 = 0.
$$

While our proposed $R_F^2$ and the method by Nakagawa et al. (2017) estimate $\rho_F^2$ very well, they both underestimate $\rho_M^2$, which may be a desirable property for model selection.

We also fitted all models using the restricted maximum likelihood method, and resultant coefficients of determinations are in general very close to the corresponding values calculated from the maximum likelihood estimators (the results are not shown). In particular, the method by Xu (2003) is rarely affected by the choice of the restricted maximum likelihood or maximum likelihood method estimator because the difference between restricted maximum likelihood and maximum likelihood estimators lies only in the estimated variances of the random effects, i.e., $\tau_{ij}^2$.

4.2. Logistic Mixed Models

For logistic models, we compared the performance of our proposed $R^2$ to those proposed by Nakagawa et al. (2017), with a total of 1,000 data sets simulated from each model under investigation. Each data set has a total of 400 random samples, evenly clustered inside $m$ groups with $m = 10$ and 50, respectively. The binary $X_1$ and continuous $X_2$ were generated similarly to those in the previous section. The $j$th response value inside the $i$th cluster, i.e., $y_{ij}$, was generated from a Bernoulli distribution with

$$
E[y_{ij} \mid \mu_i, x_{1ij}] = \{1 + \exp(-\mu_i - x_{1ij}\beta)\}^{-1},
$$

where $x_{1ij}$ is the corresponding value of the binary covariate $X_1$, and the random effect $\mu_i \sim iid N(0, 1)$.

For each data set, we fit different models with the maximum likelihood method, and then estimate both $\rho_M^2$ and $\rho_F^2$ with our approach and the method by Nakagawa et al. (2017) as shown in Fig. 2. With true values unknown, we fit a generalized linear model for each mixed-effects model by including fixed, instead of random, group effects, and calculated $R_{KL}^2$ by Cameron and Windmeijer (1997) and $R_V^2$ by Zhang (2017) as benchmarks. Note that $R_{KL}^2$ is defined on the basis of the Kullback–Leibler divergence and $R_V^2$ is defined based on quantifying the variation change along the variance function. Although $R_V^2$ may conceptually favor our proposed $R^2$, $R_{KL}^2$ can serve well as a benchmark by taking into account of whole distributional diffusion including quadratic variation. Figure 2 indeed shows that both $R_V^2$ and $R_{KL}^2$ have similar increasing patterns with the discrepancies due to their conceptual difference.

When the true predictor $X_1$ is included in the model, the estimated $\rho_M^2$ and $\rho_F^2$ by Nakagawa et al. (2017) are in general flatter than other estimates. All estimates of $\rho_M^2$ and
are slightly affected by the number of groups. It is not surprising that, in Fig. 2a, b, both $R^2_K$ and $R^2_V$ are significantly affected by the number of groups as including more fixed group effects in the model may inflate the corresponding $R^2$. Both $R^2_K$ and $R^2_V$ perform consistently between Fig. 2c, d as the group effects are excluded from the models. Note that $R^2_F$ and $R^2_V$ are overlapped in Fig. 2c, d due to their identity by definition. When regressing $Y$ vs. $X_2$ by excluding the true predictor $X_1$, our proposed measures and the measures by Nakagawa et al. (2017) perform consistently with $m = 10$ and $m = 50$. In general, $R^2_M$ and $R^2_F$ capture the increasing proportion of explained variation in a binary response variable, and perform better than the method by Nakagawa et al. (2017).

4.3. LOGLINEAR MIXED MODELS

For loglinear models, we compared the performance of our proposed $R^2$ to those proposed by Nakagawa et al. (2017), with a total of 1000 data sets simulated from each model under investigation. Each data set has a total of 400 random samples, evenly clustered inside $m = 50$ groups. The binary $X_1$ and continuous $X_2$ were generated similarly to those in the
Coefficients of Determination for Mixed-Effects Models

Figure 3. Medians of estimated $\rho^2_M$ and $\rho^2_F$ in loglinear mixed models. Shown in the plot are $R^2_M$ and $R^2_F$ (blue), and the method by Nakagawa et al. (2017) (red). For reference, we also plot $R^2_{KL}$ (green) and $R^2_V$ (black) for corresponding generalized linear models. The solid lines are for the models including $X_1$, and the dotted lines are for the models including $X_2$ instead (Color figure online).

previous sections. The $j$th response value inside the $i$th cluster, i.e., $y_{ij}$, was generated from a Poisson distribution with

$$E[y_{ij} \mid \mu_i, x_{1ij}] = \exp(\mu_i + x_{1ij}\beta),$$

where $x_{1ij}$ is the corresponding value of the binary covariate $X_1$, and the random effect $\mu_i \sim iid N(0, .25)$. We also simulated overdispersed counts from a negative binomial distribution with the success probability

$$p_{ij} = 1/\{1 + \exp(-\mu_i - x_{1ij}\beta)\}.$$

For each data set, we fit the different models with the maximum likelihood method, and then estimate both $\rho^2_M$ and $\rho^2_F$ with our approach and the method by Nakagawa et al. (2017) as shown in Figure 3. For each model, we also fit a generalized linear model by including fixed, instead of random, group effects, and calculated $R^2_{KL}$ by Cameron and Windmeijer (1997) and $R^2_V$ by Zhang (2017) as benchmarks.
When the true predictor \( X_1 \) is included in the model, as shown in Fig. 3a, b, \( R_M^2 \) overlays its reference \( R_V^2 \) for the Poisson counts data but is slightly smaller than \( R_M^2 \) for the overdispersed counts data. As conceptually identical, it is not surprising to observe overlaid \( R_F^2 \) and \( R_V^2 \) in Fig. 3c, d. While \( R_{KL}^2 \) measures overall distributional diffusion including quadratic variation, it reports slightly larger values than \( R_M^2 \) in Fig. 3a, b but much larger values than \( R_F^2 \) in Fig. 3c, d, in particular when \( \beta \) is large. On the other hand, the method by Nakagawa et al. (2017) reports slightly larger estimate of \( \rho_M^2 \) than \( R_{KL}^2 \) in Fig. 3a, b, but much larger estimate of \( \rho_F^2 \) than \( R_{KL}^2 \) in Fig. 3c, d.

When \( X_2 \), instead of the true predictor \( X_1 \), is included in the model, \( \rho_M^2 \) only measures the proportion of variation due to the random effects so it should decrease when \( \beta \) increases. While \( R_M^2 \) as well as \( R_V^2 \) and \( R_{KL}^2 \) decreases, the estimate by Nakagawa et al. (2017) increases and approach to one as \( \beta \) increases, as shown in Fig. 3a, b. Therefore, the method by Nakagawa et al. (2017) falsely evaluates the proportion of variation explained by \( X_2 \). On the other hand, all methods perform well in estimating \( \rho_F^2 \) in this scenario as shown in Fig. 3c, d. In general, the method proposed by Nakagawa et al. (2017) tends to overstate the proportion of explained variation in a count response while \( R_M^2 \) and \( R_F^2 \) capture well the increasing pattern as benchmarked by \( R_{KL}^2 \).

### 5. REAL DATA ANALYSIS

#### 5.1. CORN YIELD WITH VARIABLE NITROGEN

A total of 1,705 observations were collected in 2001 from four different topographic regions on corn yield with five different nitrogen treatments in Argentina (Edmondson 2014; Lambert et al. 2004). Recorded with each observation is a brightness value, proxy for low organic matter content, which varies across different topographic regions. With possible fixed effects predictors nitrogen treatments (N) and brightness value (B), we consider different linear mixed models, all having the topographic region as a factor with random effects. Since there is high association between brightness value and topographic region, we also regressed brightness value against topographic region and include its residual (B_r), instead of itself, in the models, see Table 2. As references, classical and adjusted coefficients of determination (\( R^2 \)) are also calculated for the corresponding models but including the topographic region with fixed effects.

When including original brightness value in the model, the method by Nakagawa et al. (2017) estimated \( \rho_M^2 \) with slightly lower values than \( R_M^2 \) except that, for the model with only N having fixed effects, the method by Nakagawa et al. (2017) reported its largest value at 0.774. Note that the method by Nakagawa et al. (2017) reported a larger value for the model with B only than the model with B and \( B^2 \), and also decreasing values for the three increasing models: (i) the model with N only; (ii) the model with N and B; and (iii) the model with N, B, and \( B^2 \). Therefore, the method by Nakagawa et al. (2017) may not necessarily report larger values for larger models, unlike other unadjusted coefficients of determinations which usually report larger values for larger models. The method by Nakagawa et al. (2017) also estimated \( \rho_F^2 \) with much smaller values than \( R_F^2 \) except for the model with N only.
Table 2. Coefficients of determination in analysis of the corn yield data

| Fixed effects | $\rho_M^2$ | NJS$^a$ | Xu$^b$ | $\rho_F^2$ | AIC | BIC | $R^2$ |
|---------------|-----------|---------|--------|-----------|-----|-----|-------|
| N, B, B$^2$   | 0.779 (0.778) | 0.754 | 0.768 | 0.303 | 13,444 | 13,499 | 0.769 (0.767) |
| N, B          | 0.778 (0.778) | 0.758 | 0.768 | 0.284 | 13,446 | 13,494 | 0.768 (0.767) |
| N             | 0.730 (0.729) | 0.774 | 0.719 | 0.005 | 13,770 | 13,814 | 0.719 (0.717) |
| B, B$^2$      | 0.774 (0.774) | 0.748 | 0.763 | 0.297 | 13,460 | 13,498 | 0.763 (0.762) |
| B             | 0.774 (0.773) | 0.753 | 0.762 | 0.279 | 13,485 | 13,507 | 0.762 (0.762) |
| N, B$^r$, B$^2_r$ | 0.776 (0.775) | 0.813 | 0.770 | 0.069 | 13,433 | 13,488 | 0.770 (0.769) |
| N, B$^r$      | 0.776 (0.775) | 0.813 | 0.768 | 0.054 | 13,447 | 13,496 | 0.768 (0.767) |
| B$^r$, B$^2_r$ | 0.773 (0.773) | 0.809 | 0.764 | 0.063 | 13,473 | 13,501 | 0.764 (0.764) |
| B$^r$         | 0.771 (0.771) | 0.809 | 0.764 | 0.048 | 13,486 | 13,508 | 0.762 (0.762) |

Bold values indicate the highest value in each column
With adjusted values in parentheses

$^a$The method by Nakagawa et al. (2017)
$^b$The method by Xu (2003)

When the brightness values were replaced by their residuals, the method by Nakagawa et al. (2017) estimated $\rho_M^2$ with larger values than $R_M^2$, and reported slightly smaller estimates of $\rho_F^2$ than $R_F^2$. The method by Nakagawa et al. (2017) reported the largest estimate of $\rho_M^2$ at 0.813 for two sets of models including N and B$^r$. Note that the model with N, B$^r$, and B$^2_r$ is selected by AIC, BIC, adjusted $R^2$, the methods by Xu (2003) and Nakagawa et al. (2017), while adjusted $R_M^2$ reported this model with the highest value among all models replacing the brightness value with its residual. However, when selecting models by considering the original brightness values, the method by Nakagawa et al. (2017) selected the model with N only, while BIC selected the model with N and B, and all other methods selected the model with N, B, and B$^2$.

### 5.2. Distribution of Elaphostrongylus cervi in Red Deers

Vicente et al. (2006) studied the distribution of the first-stage larvae of *Elaphostrongylus cervi* in the red deer across Spain. A total of 826 deers across 24 farms were surveyed for the incidence (Zuur et al. 2009). We considered different logistic mixed models to investigate how the length of a deer affected its chance to get the first-stage larvae of *Elaphostrongylus cervi*, and therefore include both the standardized length (L) and its quadratic term (L$^2$), as well as deer sex (S), as possible predictors. All mixed-effects models included farm as a factor with random effects, see Table 3. As references, $R_V^2$ proposed by Zhang (2017) and $R_{KL}^2$ proposed by Cameron and Windmeijer (1997) were calculated for corresponding models but including farm as a factor with fixed effects instead.

In general, the method by Nakagawa et al. (2017) provided slightly larger estimates of $\rho_F^2$ than $R_F^2$ (except two models with similar values), but much larger estimates of $\rho_M^2$ than $R_M^2$ in all models. On the other hand, $R_M^2$ reported very similar values to $R_F^2$ while $R_{KL}^2$ reported slightly smaller values than both of them. However, adjusted versions of $R_M^2$, $R_V^2$, and $R_{KL}^2$ all selected the largest model as the best, while both AIC and BIC selected the
Table 3. Coefficients of determination in analysis of the red deers data

| Fixed effects | $\rho^2_M$ | $\rho^2_F$ | AIC | BIC | $R^2_V$ | $R^2_{KL}$ |
|---------------|------------|------------|-----|-----|---------|------------|
| $S \times L$, $S \times L^2$ | 0.352 (0.347) 0.507 | 0.103 0.156 | 832 | 865 | 0.357 (0.334) | 0.308 (0.284) |
| $S \times L$ | 0.348 (0.345) 0.504 | 0.087 0.143 | 833 | 856 | 0.354 (0.333) | 0.305 (0.283) |
| $S$, $L$, $L^2$ | 0.349 (0.346) 0.509 | 0.102 0.099 | 831 | 854 | 0.354 (0.333) | 0.306 (0.283) |
| $S$, $L$ | 0.340 (0.337) 0.501 | 0.083 0.131 | 841 | 860 | 0.345 (0.324) | 0.296 (0.274) |
| $L$, $L^2$ | 0.347 (0.345) 0.507 | 0.102 0.159 | 830 | 849 | 0.353 (0.332) | 0.304 (0.282) |
| $L$ | 0.334 (0.333) 0.495 | 0.083 0.128 | 844 | 858 | 0.340 (0.320) | 0.291 (0.270) |
| $S$ | 0.239 (0.237) 0.344 | 0.012 0.005 | 931 | 945 | 0.247 (0.225) | 0.207 (0.184) |

Bold values indicate the highest value in each column
With adjusted values in parentheses

*The method by Nakagawa et al. (2017)*

Table 4. Coefficients of determination in analysis of the owl data

| Fixed effects | $\rho^2_M$ | $\rho^2_F$ | AIC | BIC | $R^2_V$ | $R^2_{KL}$ |
|---------------|------------|------------|-----|-----|---------|------------|
| $S \times F$, $S \times T$ | 0.277 (0.269) 0.684 | 0.174 0.278 | 5009 | 5040 | 0.276 (0.237) | 0.226 (0.184) |
| $S \times F$, $T$ | 0.275 (0.269) 0.684 | 0.174 0.279 | 5008 | 5034 | 0.275 (0.237) | 0.226 (0.185) |
| $S$, $F$, $T$ | 0.275 (0.270) 0.682 | 0.174 0.277 | 5010 | 5032 | 0.275 (0.238) | 0.225 (0.186) |
| $F$, $T$ | 0.275 (0.271) 0.682 | 0.172 0.276 | 5009 | 5027 | 0.274 (0.239) | 0.225 (0.187) |
| $T$ | 0.217 (0.214) 0.610 | 0.096 0.139 | 5288 | 5301 | 0.218 (0.181) | 0.157 (0.117) |
| $F$ | 0.201 (0.199) 0.639 | 0.106 0.184 | 5212 | 5226 | 0.201 (0.163) | 0.175 (0.136) |

Bold values indicate the highest value in each column
With adjusted values in parentheses

*The method by Nakagawa et al. (2017)*

model with L and $L^2$, and the method by Nakagawa et al. (2017) selected the model with $S$, L, and $L^2$ instead.

While all of $R^2_M$, $R^2_V$, and $R^2_{KL}$ reported larger values for larger models, the method by Nakagawa et al. (2017) also reported larger values for larger models except the model with $S$, L, and $L^2$ for which it reported the largest value at .509, higher than the value for the largest model. However, both $R^2_F$ and the method by Nakagawa et al. (2017) estimated $\rho^2_F$ with larger values for larger models.

5.3. THE BEGGING BEHAVIOR OF NESTLING BARN OWLS

Roulin and Bersier (2007) studied vocal begging behavior of nestling barn owls when the parents brought prey across 27 nests, with 2 to 7 nestlings per nest. A total of 599 observations collected with the response variable sibling negotiation, which is defined as the number of calls just before arrival of a parent at a nest, as well as the number of siblings per nest, the parent’s sex (S), arrival time at the nest (T), and whether the nestlings were food satiated or deprived (F) (Zuur et al. 2009). We considered loglinear mixed models with
the number of calls offset by the number of siblings per nest. All models included nest as a factor with random effects, see Table 4. As references, \( R^2_V \) proposed by Zhang (2017) and \( R^2_{KL} \) proposed by Cameron and Windmeijer (1997) were calculated for corresponding models but including nest as a factor with fixed effects instead.

As shown in Table 4, the method by Nakagawa et al. (2017) provided much larger estimates of \( \rho^2_M \) than \( R^2_M \) as well as much larger estimates of \( \rho^2_F \) than \( R^2_F \) in all models. On the other hand, \( R^2_M \) reported very similar values to \( R^2_V \) as \( R^2_{KL} \) reported slightly smaller values than both of them. However, adjusted versions of \( R^2_M, R^2_V, \) and \( R^2_{KL} \) as well as BIC all selected the model with \( F \) and \( T \) as the best, while AIC selected the model with \( S \times F \) and \( T \). The method by Nakagawa et al. (2017) instead have two models including \( S \times F \) and \( T \) with the highest estimate of \( \rho^2_M \) at .684. In general, all coefficients of determination reported larger values for larger models.

6. DISCUSSION

Unlike \( p \)-values that rely on sample sizes to signal the variable significance, the coefficient of determination, a.k.a. \( R^2 \), measures the proportion of the variation in the response variable explained by a set of predictors. It plays an important role in agricultural, biological, and ecological research, for example, quantifying the heritability of a trait in molecular biology (Visscher et al. 2008). The popularly used mixed-effects models in such studies demand appropriate extension of \( R^2 \). Our coefficients of determination are well defined as long as the link and variance functions are specified, such as quasi-models that may rely on quasi-likelihood functions, other than likelihood functions, to obtain parameter estimates (McCullagh, 1983). When the first-order derivative of the variance function is constant as for normal and Poisson distributions, our defined \( R^2 \) for generalized linear mixed models will be reduced to those defined for linear mixed models following the law of total variance. Together with Zhang (2017), our definitions unify linear models, generalized linear models, linear mixed models, and generalized linear mixed models.

Practice in science may demand further extension or improvement on the proposed measures. Firstly, it may be of interest to measure the utility of a set of predictors beyond others in modeling a response variable. We may define the coefficient of partial determination as shown in Zhang (2017) to measure the proportion of variation in the response variable unexplained by a set of predictors that can be explained by the additional set of predictors. Secondly, calculating aforementioned heritability may demand evaluation of the contribution due to certain random effects while controlling the fixed effects of other factors. In this case, we may directly define \( \rho^2_R \) based on the law of total variance, though it may be a challenging task, in particular for generalized linear mixed models. Thirdly, our coefficients of determinations are defined for mixed-effects models assuming that the observations are randomly sampled within groups and the groups are randomly sampled within the underlying population. However, the same number of observations may be collected within each group although the subpopulation may be of different size, and sometimes it is the other way around. Therefore the coefficients may be modified using different weights on individual variations from different groups.
All of our defined measures for linear mixed model and generalized linear mixed model are implemented in the R package rsq, which is publicly available via the Comprehensive R Archive Network (CRAN).

ACKNOWLEDGEMENTS

The author thanks the editor, Professor Brian Reich, an associate editor, and two anonymous reviewers for their useful comments that led to this improved version of the article. The author’s research was partially supported by the National Cancer Institute under award number R03CA235363.

[Received August 2021. Revised July 2022. Accepted July 2022. Published Online August 2022.]

REFERENCES

Cameron AC, Windmeijer AG (1997) An R-squared measure of goodness of fit for some common nonlinear regression models. J Econometr 77:329–342
Cox DR, Snell EJ (1989) The analysis of binary data, 2nd edn. Chapman and Hall, London
Dempster ER, Lerner IM (1950) Heritability of threshold characters. Genetics 35:212–236
Edmondson RN (2014) Agridat. J Agric Sci 152:2
Gbur EE, Stroup WW, McCarter KS, Durham S, Yong LJ, Christman M, West M, Kramer M (2012) Analysis of generalized linear mixed models in the agricultural and natural resources sciences. American Society of Agronomy
Ives AR (2019) $R^2$s for correlated data: phylogenetic models, LMMs, and GLMMs. Syst Biol 68:234–251
Jaeger BC, Edwards LJ, Das K, Sen PK (2017) An $R^2$ statistic for fixed effects in the generalized linear mixed model. J Appl Stat 44:1086–1105
Jaeger BC, Edwards LJ, Gurka MJ (2019) An $R^2$ statistic for covariance model selection in the linear mixed model. J Appl Stat 46:164–184
Laird NM, Ware JH (1982) Random-effects models for longitudinal data. Biometrics 38:963–974
Lambert DM, Lowenberg-Deboer J, Bongiovanni R (2004) A comparison of four spatial regression models for yield monitor data: a case study from Argentina. Precis Agric 5:579–600
Maddala GS (1983) Limited-dependent and qualitative variables in econometrics. Cambridge University
Magee L (1990) $R^2$ measures based on Wald and likelihood ratio joint significance tests. Am Stat 44:250–253
McCullagh P (1983) Quasi-likelihood functions. Ann Stat 11:59–67
McCullagh P, Nelder JA (1989) Generalized linear models. Chapman and Hall/CRC
McCulloch CE, Searle SR, Neuhaus JM (2008) Generalized, linear, and mixed models, 2nd edn. Wiley, New York
Nagelkerke NJD (1991) A note on a general definition of the coefficient of determination. Biometrika 78:691–692
Nakagawa S, Johnson PCD, Schielzeth H (2017) The coefficient of determination $R^2$ and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. J R Soc Interface 14:20170213
Nakagawa S, Schielzeth H (2013) A general and simple method for obtaining $R^2$ from generalized linear mixed models. Methods Ecol Evol 4:133–142
Roulin A, Bersier LF (1998) Nestling barn owls beg more intensely in the presence of their mother than their father. Anim Behav 74:1099–1106
Vicente J, de Mera IGF, Gortazar J (2006) Epidemiology and risk factors analysis of elaphostrongylosis in red deer (Cervus elaphus from Spain. Parasitol Res 98:77–85
Visscher PM, Hill WG, Wray NR (2008) Heritability in the genomics era-concepts and misconceptions. Nat Rev Genet 9:255–266
Xu X (2003) Measuring explained variation in linear mixed effects models. Stat Med 22:3527–3541
Zhang D (2017) A coefficient of determination for generalized linear models. Am Stat 71:310–316
Zuur AF, Ieno EN, Walker N, Saveliev AA, Smith GM (2009) Mixed effects models and extensions in ecology with R. Springer

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.