Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE @ DIRECT®

Computational Statistics & Data Analysis 44 (2003) 253–271

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

# Pseudo $R$-squared measures for Poisson regression models with over- or underdispersion

Harald Heinzl, Martina Mittlböck*

*Department of Medical Computer Sciences, University of Vienna, Spitalgasse 23, A-1090, Vienna, Austria*

Received 14 August 2002; received in revised form 15 March 2003

**Abstract**

The Poisson regression model is frequently used to analyze count data. Pseudo $R$-squared measures for Poisson regression models have recently been proposed and bias adjustments recommended in the presence of small samples and/or a large number of covariates. In practice, however, data are often over- or sometimes even underdispersed as compared to the standard Poisson model. The definition of Poisson $R$-squared measures can be applied in these situations as well, albeit with bias adjustments accordingly adapted. These adjustments are motivated by arguments of quasi-likelihood theory. Properties of unadjusted and adjusted $R$-squared measures are studied by simulation under standard Poisson; over- and underdispersed Poisson regression models and their use is exemplified and discussed with popcorn data.


*Keywords:* Adjusted $R$-squared measure; Poisson regression; Overdispersion; Underdispersion; Extra-Poisson variation; Deviance residuals; Likelihood ratio; Shrinkage; Quasi-likelihood

## 1. Introduction

Regression methods for handling count data based on Poisson distribution are in frequent use. Pseudo $R$-squared measures—also called measures of explained variation or coefficients of determination—for Poisson regression have recently been defined (Cameron and Windmeijer, 1996; Waldhör et al., 1998), corresponding bias adjustments for small samples and/or many covariates have been studied (Mittlböck and Waldhör,

* Corresponding author. Tel.: +43-1-40400-2276; fax: +43-1-40400-2278.
*E-mail addresses:* harald.heinzl@univie.ac.at (H. Heinzl), martina.mittlboeck@akh-wien.ac.at (M. Mittlböck).

0167-9473/$ - see front matter 
doi:10.1016/S0167-9473(03)00062-8

2000; Mittlböck, 2002), and distinctive features have been clarified (Mittlböck and Heinzl, 2001; Waldhör et al., 2001). In practice, however, data are often over- and sometimes even underdispersed as compared to the standard Poisson model. Overdispersion (also called extra-Poisson variation) occurs if the dispersion of the observed outcome $Y$ is greater than predicted by the Poisson model, i.e. $\mathrm{Var}(Y) > \mu$, where $\mu = E(Y)$. By analogy, in the case of underdispersion, $\mathrm{Var}(Y) < \mu$. Possible reasons for overdispersion are described e.g. in McCullagh and Nelder (1989).

When applying the standard Poisson model to over- or underdispersed data, the efficiency of parameter estimates remains reasonably high, yet their standard errors are under- or overestimated, respectively. Hence, coverage probabilities of confidence intervals and significance levels of tests are no longer valid and can result in highly misleading outcome (e.g. see Paul and Plackett, 1978, or Cox, 1983).

If the precise mechanism that produces the over- or underdispersion is known, specific methods may be applied to model the data, see e.g. McCullagh and Nelder (1989). In the absence of such knowledge, it is convenient to assume approximatively that $\mathrm{Var}(Y) = \phi\mu$ for some positive constant $\phi$. In the case of overdispersion $\phi > 1$, and in the case of underdispersion $\phi < 1$. This is a rather robust approach to tackle the problem, since even quite substantial deviations in the assumed simple linear functional form, $\mathrm{Var}(Y) = \phi\mu$, generally have a merely minor effect on the conclusions related to standard errors, confidence intervals and $p$-values (McCullagh and Nelder, 1989). Now, fitting this model is a three-step procedure. Firstly, regression coefficients and their standard errors are estimated by maximizing the likelihood of a standard Poisson regression model. Secondly, the dispersion parameter $\phi$ is estimated separately. And thirdly, the standard errors are adjusted for the estimated dispersion parameter so that proper confidence intervals and test statistics can be obtained. Note that the resulting likelihood-ratio tests are based on quasi-likelihood theory (see McCullagh and Nelder, 1989).

The objective of this paper is to study Poisson pseudo $R$-squared measures for over- and underdispersion. We generalize the adjusted $R$-squared measure of Mittlböck and Waldhör (2000) for over- and underdispersed Poisson regression models in an attempt to adjust for bias in the presence of small samples and/or many covariates. We investigate the performance of $R$-squared measures for Poisson regression models under over- and underdispersion when ignoring the fact of over-/underdispersion and compare these results with the performance of our proposed adjustment.

Section 2 gives the deviance-based $R$-squared measure and the adjustment of Mittlböck and Waldhör (2000), and an adjustment for $R$-squared measures in over- and underdispersed Poisson models is introduced in Section 3. A simulation study in Section 4 compares the performance of the $R$-squared measures under standard and over- and underdispersed Poisson models. A real data example is presented in Section 5 and a discussion given in Section 6.

## 2. Unadjusted and adjusted deviance-based $R$-squared measures

With regard to Poisson regression models, Cameron and Windmeijer (1996) and Waldhör et al. (1998) gave preference to the $R$-squared measure based on deviances denoted by $R^2_{\mathrm{D}}$,

$$R^2_{\mathrm{D}} = 1 - \frac{D(\mathbf{y};\hat{\mu})}{D(\mathbf{y};\bar{y})} = 1 - \frac{\log L(\mathbf{y}) - \log L(\hat{\mu})}{\log L(\mathbf{y}) - \log L(\bar{y})}$$

$$= 1 - \frac{\sum_i [y_i \log(y_i/\hat{\mu}_i) - (y_i - \hat{\mu}_i)]}{\sum_i [y_i \log(y_i/\bar{y}) - (y_i - \bar{y})]},$$

where $\log L(\mathbf{y})$ is the log-likelihood of the saturated model; $\log L(\hat{\mu})$ and $D(\mathbf{y};\hat{\mu}) = 2[\log L(\mathbf{y}) - \log L(\hat{\mu})]$ are the log-likelihood and the deviance of the full model which contains the covariates of interest; $\log L(\bar{y})$ and $D(\mathbf{y};\bar{y}) = 2[\log L(\mathbf{y}) - \log L(\bar{y})]$ are the log-likelihood and the deviance of the intercept-only model; $\mathbf{y} = (y_1,\ldots,y_n)'$ are the observed values of the dependent variable; $\hat{\mu} = (\hat{\mu}_1,\ldots,\hat{\mu}_n)'$ are the corresponding predicted values under the full model, respectively. To simplify matters, no offset is assumed throughout the paper so that the predicted values under the intercept-only model are all equal to $\bar{y} = n^{-1}\sum_{i=1}^{n} y_i$.

$R^2_{\mathrm{D}}$ can be interpreted as the relative reduction in deviance due to the covariates in the model. An alternative definition of an $R$-squared measure could also be based on sums-of-squares, but this has come to be seen as suboptimal for Poisson regression models (Cameron and Windmeijer, 1996; Waldhör et al., 1998).

$R^2_{\mathrm{D}}$ for Poisson models shares the characteristics of the $R$-squared for the common linear model: It is restricted to values between zero and one and does not decrease with an increasing number of covariates. However, it is not desirable that values of an $R$-squared measure increase when covariates that lack association with the dependent variable are added to the model. In cases with small samples and/or many covariates, values of the $R$-squared measure may be seriously inflated, as part of its value arises from modeling pure chance.

Mittlböck and Waldhör (2000) proposed an adjustment of the $R^2_{\mathrm{D}}$ bias for Poisson regression models, defined by

$$R^2_{\mathrm{D},\gamma} = 1 - \frac{D(\mathbf{y};\hat{\mu}) + k}{D(\mathbf{y};\bar{y})} = 1 - \frac{\log L(\mathbf{y}) - \log L(\hat{\mu}) + k/2}{\log L(\mathbf{y}) - \log L(\bar{y})}$$

$$= 1 - \frac{\sum_i [y_i \log(y_i/\hat{\mu}_i) - (y_i - \hat{\mu}_i)] + k/2}{\sum_i [y_i \log(y_i/\bar{y}) - (y_i - \bar{y})]},$$

where $k$ is the number of covariates fitted (without intercept) in the full model.

The notation "$\gamma$" in the subscript refers to the fact that $R^2_{\mathrm{D},\gamma}$ is equal to the shrunk $R$-squared measure, that is $R^2_{\mathrm{D},\gamma} = \gamma R^2_{\mathrm{D}}$, where $\gamma$ is the shrinkage factor (see Copas, 1983, 1987, 1997; Van Houwelingen and Le Cessie, 1990). It is estimated by

$$\hat{\gamma} = \frac{\hat{G} - k}{\hat{G}} = \frac{D^*(\mathbf{y};\bar{y}) - D^*(\mathbf{y};\hat{\mu}) - k}{D^*(\mathbf{y};\bar{y}) - D^*(\mathbf{y};\hat{\mu})},$$

where $\hat{G} = D^*(\mathbf{y};\bar{y}) - D^*(\mathbf{y};\hat{\mu})$ is the log-likelihood ratio $\chi^2$-statistic for testing whether any of the $k$ fitted covariates are associated with the response. The expression $D^*(\mathbf{y};\mu) = D(\mathbf{y};\mu)/\phi$ is the so-called scaled deviance. The deviance equals the scaled deviance in standard Poisson models where $\phi = 1$ by definition. Note that $E(\hat{G})$ is approximately $k$ under the null hypothesis of no covariate effects. The estimated shrinkage factor $\hat{\gamma}$ reaches its maximum value of nearly 1, indicating no shrinkage if $\hat{G}$ achieves a high value. The model will be statistically significant in such a case. $\hat{\gamma}$ approximates zero with decreasing $\hat{G}$, indicating that a major part of $R^2_{\mathrm{D}}$ is due to pure chance alone. In the case of negative shrinkage, $\hat{G} < k$, the model explains less than expected under the null hypothesis of no covariate effects. This is identical to no explained variation at all, and a value of zero will be commonly reported for the adjusted $R$-squared measure.

## 3. Adjusted *R*-squared measures for over- and underdispersed Poisson data

In standard Poisson regression models, the dispersion parameter $\phi$ is by definition constrained to a value of one and $\mathrm{Var}(Y) = \mu$. An unconstrained dispersion parameter has been suggested to tackle the problem of over- and underdispersed Poisson data. That is, $\mathrm{Var}(Y) = \phi\mu$, with $\phi > 1$ and $\phi < 1$ for over- and underdispersion, respectively. Note that $\sigma^2$ will often be used instead of $\phi$ to denote the dispersion parameter in Poisson models.

It is now straightforward to define a shrinkage factor for over- or underdispersed Poisson data, generically denoted by $\gamma_{\mathrm{ov}}$,

$$\gamma_{\mathrm{ov}} = \frac{D^*(\mathbf{y}; \bar{y}) - D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) - k}{D^*(\mathbf{y}; \bar{y}) - D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})} = \frac{D(\mathbf{y}; \bar{y}) - D(\mathbf{y}; \hat{\boldsymbol{\mu}}) - k\phi}{D(\mathbf{y}; \bar{y}) - D(\mathbf{y}; \hat{\boldsymbol{\mu}})}.$$

This shrinkage factor is no different from a common likelihood-based shrinkage factor in generalized linear models with an unconstrained dispersion parameter, e.g. the normal, the gamma or the inverse Gaussian model (Mittlböck and Heinzl, 2002; Heinzl and Mittlböck, 2002). However, the likelihood function of a Poisson distribution with an unconstrained dispersion parameter is a quasi-likelihood function, and no longer a legitimate likelihood function. Nevertheless, most of the asymptotic theory for log-likelihoods also applies to quasi-likelihoods (see McCullagh and Nelder, 1989, Chapter 9).

There are two common ways to estimate the dispersion parameter $\phi$ for over-/underdispersed Poisson regression, see e.g. McCullagh and Nelder (1989). Firstly, $\phi$ can be estimated by the deviance divided by the degrees of freedom,

$$\hat{\phi}_{\mathrm{D}} = D(\mathbf{y}; \hat{\boldsymbol{\mu}})/(n - k - 1),$$

and secondly, $\phi$ can be estimated by the generalized Pearson statistic, $X^2$, divided by the degrees of freedom, which is

$$\hat{\phi}_{\mathrm{P}} = X^2/(n - k - 1) = \sum_i \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i} \Big/ (n - k - 1)$$

in the case of the Poisson model.

Accordingly, there are two estimates for the shrinkage factor $\gamma_{\mathrm{ov}}$, i.e. $\hat{\gamma}_{\mathrm{D}}$ and $\hat{\gamma}_{\mathrm{P}}$, respectively. Now, adjustments of $R^2_{\mathrm{D}}$ that account for over- and underdispersion are either

$$R^2_{D,\gamma_D} = \hat{\gamma}_D R^2_D = 1 - \frac{D(\mathbf{y};\hat{\mu}) + k\hat{\phi}_D}{D(\mathbf{y};\bar{y})}$$

$$= 1 - \frac{D(\mathbf{y};\hat{\mu})/(n-k-1)}{D(\mathbf{y};\bar{y})/(n-1)}$$

$$= 1 - \frac{(n-k-1)^{-1}\sum_i [y_i \log(y_i/\hat{\mu}_i) - (y_i - \hat{\mu}_i)]}{(n-1)^{-1}\sum_i [y_i \log(y_i/\bar{y}) - (y_i - \bar{y})]}$$

or

$$R^2_{D,\gamma_P} = \hat{\gamma}_P R^2_D = 1 - \frac{D(\mathbf{y};\hat{\mu}) + k\hat{\phi}_P}{D(\mathbf{y};\bar{y})}.$$

The necessity to consider over-/underdispersion in the bias adjustment for $R^2_D$ increases the more the dispersion parameter $\phi$ differs from one. In the case of overdispersion, the shrinkage factors $\hat{\gamma}_D$ and $\hat{\gamma}_P$ will be smaller than the standard Poisson shrinkage factor $\hat{\gamma}$, where a dispersion parameter equal to one is assumed. Consequently, the naive adjusted $R$-squared measure, $R^2_{D,\gamma}$, will be too large as compared to either $R^2_{D,\gamma_D}$ or $R^2_{D,\gamma_P}$. By analogy, $\hat{\gamma}_D$ and $\hat{\gamma}_P$ will be larger than $\hat{\gamma}$ in underdispersed Poisson models, since less shrinkage will occur. Of course, both the definitions of $R^2_{D,\gamma_D}$ and $R^2_{D,\gamma_P}$ will simplify to $R^2_{D,\gamma}$ in the case of $\phi = 1$.

Note that $R^2_{D,\gamma_D}$ is equivalent to an adjusted $R$-squared measure (usually denoted by $R^2_{D,df}$), which has been studied for the standard Poisson regression model (Cameron and Windmeijer, 1996; Waldhör et al., 1998; Mittlböck and Waldhör, 2000). These authors were guided by a rather heuristic motivation in emphasizing the analogy with the common definition of the adjusted linear model $R$-squared measure. However, the linear regression model with normal errors has an unrestricted dispersion parameter, whereas the standard Poisson regression model has a restricted dispersion parameter. In other words, simply reasoning by analogy from the linear model to the standard Poisson model has unintentionally resulted in an adjusted $R$-squared measure for over- and underdispersed Poisson models. (Note that the same argument would apply for the logistic regression model as well.)

It is interesting to note that in the classical linear regression model with homoscedastic normal errors, $\hat{\phi}_D$ equals $\hat{\phi}_P$ by definition so that a unique definition of a shrunk $R$-squared measure becomes available. For the linear regression model, Harrell et al. (1996) have shown that the shrunk $R$-squared measure is equal to the well known adjusted $R$-squared measure, which is the regression mean square divided by corrected total mean square, calculated routinely in nearly every statistical program package.

Finally, one question remains to be resolved: The shrinkage factor $\gamma_{ov}$ is generated from the scaled deviance of the full model, $D^*(\mathbf{y};\hat{\mu}) = D(\mathbf{y};\hat{\mu})/\phi$, and the scaled deviance of the intercept-only model, $D^*(\mathbf{y};\bar{y}) = D(\mathbf{y};\bar{y})/\phi$, with the full-model dispersion parameter $\phi$ used for scaling purposes in both cases. It is crucial to use the full-model dispersion parameter for both scaled deviances, as the regressor-dependent fit is to be measured, and not that depending on the dispersion parameter (Cameron and Windmeijer, 1996; Mittlböck and Heinzl, 2002; Heinzl and Mittlböck, 2002). It would now seem instructive again to consider the classical linear regression model with homoscedastic normal errors. If the dispersion parameter $\phi = \sigma^2$ is known, the scaled deviance is $\sum_i (y_i - \hat{\mu}_i)^2/\sigma^2$ for the full model and $\sum_i (y_i - \bar{y})^2/\sigma^2$ for the intercept-only model. The same estimate $\hat{\sigma}^2 = (n-k-1)^{-1} \sum_i (y_i - \hat{\mu}_i)^2$ will be applied for both models if $\sigma^2$ is unknown. Provided that we would use $\hat{\sigma}^2$ for the full model and $\hat{\sigma}_0^2 = (n-1)^{-1} \sum_i (y_i - \bar{y})^2$ for the intercept-only model, then the shrinkage factor would always equal zero. In addition, note that the same argument applies to why it is irrelevant whether $R_D^2$ is defined using unscaled or scaled deviances. In the latter case, the same dispersion parameter has to be used in the nominator and the denominator, so that it will cancel out. To make the point clear, let us again consider the classical homoscedastic error regression model: Applying the dispersion parameters $\hat{\sigma}^2$ and $\hat{\sigma}_0^2$ to scale the full model and intercept-only model deviances, respectively, the resulting scaled deviance-based $R_D^2$ would always equal $k/(n-1)$ (Cameron and Windmeijer, 1996).

## 4. Simulation study

The performance of the bias adjustments for the pseudo $R$-squared measures was compared, under various conditions, for over- and underdispersed Poisson data. Using the FACTEX procedure of SAS (SAS Institute Inc., SAS®, Version 8, Cary, NC), a factorial design was produced for sample sizes $n=16$, 32, 64 and 1048576 [$=2^{20}$] with $k=1$, 3 and 5 completely balanced dichotomous covariates $x_1, \dots, x_k$ with values 0 and 1. A fractional factorial design only was produced for simulations with $k=5$ covariates and a sample size of $n=16$, as the number of observations in this case was insufficient for a factorial design. Over- and underdispersed Poisson random numbers were generated as follows: With the RANPOI function of SAS, Poisson-distributed random values were generated with mean $\mu/\phi$ and multiplied with $\phi$, so that the resulting random variable $Y$ is over- or underdispersed Poisson-distributed with mean $\mu$ and variance $\mu\phi$. The mean $\mu$ is usually expressed as $\mu = \exp(\beta^0 + \beta^1 x_1 + \cdots + \beta^k x_k)$ and consequently, the mean value used in the RANPOI function was $\mu/\phi = \exp(\beta^0 - \log(\phi) + \beta^1 x_1 + \cdots \beta^k x_k)$. Only the first covariate $x_1$ was assumed to influence the mean; the prognostic effect of all other covariates was eliminated by setting $\beta^2 = \cdots = \beta^k = 0$, and $\beta^0$ was set to an arbitrary value of 2. An iterative numerical search was employed to identify the values of $\beta^1$ that yield values for $R_D^2$ of 20, 40, 60 and 80 percent in a large sample ($n=1048576$) with one covariate ($k=1$) and different dispersion parameters ($\sqrt{\phi}=0.5$, 1, 2, 3), see Table 1. Since the inflation of $R_D^2$ is numerically irrelevant when fitting one covariate to a large sample, these results were taken to be the true $R$-squared values denoted by $R_{D,1}^2$. Note that in such large samples, the estimated dispersion parameters also reproduce the true values with considerable numerical accuracy. The GENMOD procedure of SAS was used to fit Poisson regressions to the data with a logarithmic link function and a SAS macro was written to do all subsequent calculations. The number of repetitions was 1000 for sample sizes $n = 16$, 32 and 64; only one single computation was performed for the large sample ($n = 1048576$).

Table 1
Chosen values for the regression coefficient $\beta^1$ in the simulations studies

| Poisson model | $\sqrt{\phi} \setminus R^2_{D,1}$ | 0% | 20% | 40% | 60% | 80% |
|---|---|---|---|---|---|---|
| Underdispersed | 0.5 | 0 | 0.17625 | 0.28045 | 0.40750 | 0.6295 |
| Standard | 1 | 0 | 0.34095 | 0.5307 | 0.7526 | 1.1199 |
| Overdispersed | 2 | 0 | 0.659 | 0.985 | 1.3434 | 1.8945 |
| Overdispersed | 3 | 0 | 0.9303 | 1.351 | 1.793 | 2.4465 |

The unadjusted and adjusted $R$-squared measures were estimated for the various combinations of covariate numbers $k$, sample sizes $n$ and true $R$-squared values $R^2_{D,1}$. The mean values of 1000 replications are listed in Tables 2–5 for underdispersed ($\sqrt{\phi}=0.5$), standard ($\phi=1$) and overdispersed Poisson models ($\sqrt{\phi}=2$ and 3), respectively. As to the case of the largest number of covariates, $k=5$, and the smallest sample size, $n=16$, boxplots in Figs. 1–5 display the effects of over- and underdispersion on the distributions of $R$-squared values for $R_{D,1}=0$, 20, 40, 60 and 80 percent, respectively.

The mean squared error, mse, was estimated by taking the mean of the squared differences between the estimated $R$-squared value and $R^2_{D,1}$ over all 1000 replications. The mse of an estimator is defined as bias-squared plus variance. Reducing the bias of an estimator may thus increase its variance, and the mse can be seen as cost-benefit calculation. For the sake of convenient comparison, Tables 2–5 present mse ratios between adjusted and unadjusted $R$-squared measures. Throughout the paper, mse($R^2_D$) has been put in the denominator of the mse ratios and an adjusted $R$-squared measure in the nominator. An mse ratio smaller than one indicates that the adjusted $R$-squared measure (in the nominator) will perform better than its unadjusted counterpart $R^2_D$ (in the denominator), and vice versa. If the mse ratio equals one, then both estimators will perform equally well or poorly in terms of the mse. Of course, the mse ratios can be used to mutually compare the various adjusted $R$-squared measure definitions in terms of the mse.

Since it is common practice to report a value of zero in the presence of a negative-computed adjusted $R$-squared value, a so-called truncated mean squared error, tmse, has been defined. The tmse is computed in exactly the same way as the mse, apart from the adjusted $R$-squared values below zero being set to a value equal to zero before the computations. Note that mse($R^2_D$) = tmse($R^2_D$) since $0 \leqslant R^2_D \leqslant 1$. Table 6 presents tmse ratios for $R^2_{D,1}$ of 0 and 20 percent. Negative values of the adjusted $R$-squared measures are quite common for such low values of $R^2_{D,1}$ (see also Figs. 1 and 2), and mse and tsme may differ considerably.

Tables 2–5 show that the bias of $R^2_D$ increases with the number of covariates $k$, decreasing sample size $n$, and decreasing true value $R^2_{D,1}$. The bias of $R^2_D$ does not depend on $\sqrt{\phi}$, i.e. it is not affected by under- or overdispersion.

When adjusted $R$-squared measures are used to reduce the bias of $R^2_D$, both $R^2_{D,\gamma_D}$ and $R^2_{D,\gamma_P}$ perform well, the latter even slightly better than the former. In contrast, $R^2_{D,\gamma}$ is an improper bias adjustment for under- or overdispersed Poisson models. The amount

Table 2
Simulated results for underdispersed Poisson model with $\sqrt{\phi} = 0.5$, 1000 replications. Mean values (in percent) and mean squared error ratios of pseudo $R$-squared measures are given

| $R^2_{D,1}$ | $n$ | $k$ | mean $R^2_D$ | mean $R^2_{D,\gamma}$ | mean $R^2_{D,\gamma_D}$ | mean $R^2_{D,\gamma_P}$ | $\frac{\mathrm{mse}(R^2_{D,\gamma})}{\mathrm{mse}(R^2_D)}$ | $\frac{\mathrm{mse}(R^2_{D,\gamma_D})}{\mathrm{mse}(R^2_D)}$ | $\frac{\mathrm{mse}(R^2_{D,\gamma_P})}{\mathrm{mse}(R^2_D)}$ |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 16 | 1 | 6 | −24 | 0 | 0 | 7.3 | 0.7 | 0.7 |
| | | 3 | 19 | −71 | −1 | −1 | 12.0 | 0.5 | 0.5 |
| | | 5 | 33 | −118 | 0 | 0 | 13.3 | 0.4 | 0.4 |
| | 32 | 1 | 3 | −11 | 0 | 0 | 6.4 | 0.7 | 0.7 |
| | | 3 | 9 | −32 | −1 | −1 | 9.8 | 0.4 | 0.4 |
| | | 5 | 16 | −53 | −1 | 0 | 10.2 | 0.3 | 0.3 |
| | 64 | 1 | 2 | −5 | 0 | 0 | 4.5 | 0.7 | 0.7 |
| | | 3 | 5 | −15 | 0 | 0 | 7.6 | 0.4 | 0.4 |
| | | 5 | 8 | −25 | 0 | 0 | 8.4 | 0.3 | 0.3 |
| 20 | 16 | 1 | 24 | 0 | 18 | 18 | 3.2 | 1.1 | 1.1 |
| | | 3 | 34 | −38 | 18 | 18 | 10.0 | 0.9 | 0.9 |
| | | 5 | 45 | −75 | 18 | 18 | 13.1 | 0.7 | 0.7 |
| | 32 | 1 | 22 | 11 | 20 | 20 | 1.7 | 1.0 | 1.0 |
| | | 3 | 27 | −6 | 19 | 19 | 4.9 | 0.9 | 0.9 |
| | | 5 | 32 | −22 | 19 | 19 | 7.5 | 0.7 | 0.7 |
| | 64 | 1 | 21 | 15 | 19 | 19 | 1.4 | 1.0 | 1.0 |
| | | 3 | 23 | 7 | 19 | 19 | 3.1 | 1.0 | 1.0 |
| | | 5 | 26 | 0 | 19 | 19 | 5.2 | 0.8 | 0.8 |
| 40 | 16 | 1 | 43 | 25 | 39 | 39 | 2.2 | 1.1 | 1.1 |
| | | 3 | 51 | −2 | 38 | 38 | 7.2 | 1.1 | 1.1 |
| | | 5 | 59 | −29 | 39 | 39 | 10.8 | 0.9 | 0.8 |
| | 32 | 1 | 42 | 34 | 40 | 40 | 1.4 | 1.0 | 1.0 |
| | | 3 | 45 | 21 | 40 | 40 | 3.5 | 1.0 | 1.0 |
| | | 5 | 49 | 9 | 40 | 40 | 5.8 | 0.8 | 0.8 |
| | 64 | 1 | 40 | 37 | 39 | 39 | 1.2 | 1.0 | 1.0 |
| | | 3 | 42 | 31 | 39 | 39 | 2.4 | 1.0 | 1.0 |
| | | 5 | 44 | 25 | 39 | 39 | 3.9 | 0.9 | 0.9 |
| 60 | 16 | 1 | 62 | 51 | 60 | 60 | 1.7 | 1.1 | 1.1 |
| | | 3 | 68 | 34 | 59 | 59 | 5.3 | 1.1 | 1.1 |
| | | 5 | 73 | 17 | 60 | 60 | 8.3 | 0.9 | 0.9 |
| | 32 | 1 | 61 | 56 | 60 | 60 | 1.2 | 1.0 | 1.0 |
| | | 3 | 64 | 48 | 60 | 60 | 2.7 | 1.0 | 1.0 |
| | | 5 | 66 | 40 | 60 | 60 | 4.7 | 0.9 | 0.9 |
| | 64 | 1 | 60 | 58 | 60 | 60 | 1.2 | 1.0 | 1.0 |
| | | 3 | 62 | 54 | 60 | 60 | 2.0 | 1.0 | 1.0 |
| | | 5 | 63 | 50 | 60 | 60 | 3.2 | 1.0 | 1.0 |
| 80 | 16 | 1 | 82 | 76 | 80 | 80 | 1.4 | 1.1 | 1.1 |
| | | 3 | 84 | 68 | 80 | 80 | 3.7 | 1.0 | 1.0 |
| | | 5 | 87 | 60 | 80 | 80 | 6.0 | 0.9 | 0.9 |
| | 32 | 1 | 81 | 78 | 80 | 80 | 1.1 | 1.0 | 1.0 |
| | | 3 | 82 | 74 | 80 | 80 | 2.2 | 1.0 | 1.0 |
| | | 5 | 83 | 71 | 80 | 80 | 3.7 | 0.9 | 0.9 |
| | 64 | 1 | 80 | 79 | 80 | 80 | 1.1 | 1.0 | 1.0 |
| | | 3 | 81 | 77 | 80 | 80 | 1.7 | 1.0 | 1.0 |
| | | 5 | 82 | 75 | 80 | 80 | 2.7 | 0.9 | 0.9 |

Table 3
Simulated results for standard Poisson model with $\phi = 1$, 1000 replications. Mean values (in percent) and mean squared error ratios of pseudo $R$-squared measures are given

| $R^2_{D,1}$ | $n$ | $k$ | mean $R^2_D$ | mean $R^2_{D,\gamma}$ | mean $R^2_{D,\gamma_D}$ | mean $R^2_{D,\gamma_P}$ | $\frac{mse(R^2_{D,\gamma})}{mse(R^2_D)}$ | $\frac{mse(R^2_{D,\gamma_D})}{mse(R^2_D)}$ | $\frac{mse(R^2_{D,\gamma_P})}{mse(R^2_D)}$ |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 16 | 1 | 6 | −1 | −1 | 0 | 0.7 | 0.7 | 0.7 |
| | | 3 | 19 | −3 | −1 | 0 | 0.5 | 0.5 | 0.5 |
| | | 5 | 32 | −5 | −2 | −1 | 0.4 | 0.5 | 0.4 |
| | 32 | 1 | 3 | 0 | 0 | 0 | 0.7 | 0.7 | 0.7 |
| | | 3 | 9 | −1 | −1 | 0 | 0.4 | 0.4 | 0.4 |
| | | 5 | 15 | −1 | −1 | 0 | 0.3 | 0.3 | 0.3 |
| | 64 | 1 | 1 | 0 | 0 | 0 | 0.7 | 0.7 | 0.7 |
| | | 3 | 4 | 0 | 0 | 0 | 0.4 | 0.4 | 0.4 |
| | | 5 | 8 | 0 | 0 | 0 | 0.3 | 0.3 | 0.3 |
| 20 | 16 | 1 | 24 | 18 | 19 | 19 | 1.0 | 1.1 | 1.1 |
| | | 3 | 34 | 17 | 18 | 18 | 0.8 | 0.9 | 0.9 |
| | | 5 | 45 | 16 | 17 | 18 | 0.6 | 0.7 | 0.7 |
| | 32 | 1 | 22 | 19 | 20 | 20 | 1.0 | 1.0 | 1.0 |
| | | 3 | 27 | 19 | 19 | 19 | 0.8 | 0.9 | 0.9 |
| | | 5 | 32 | 19 | 19 | 19 | 0.6 | 0.7 | 0.7 |
| | 64 | 1 | 21 | 20 | 20 | 20 | 1.0 | 1.0 | 1.0 |
| | | 3 | 23 | 20 | 20 | 20 | 0.9 | 0.9 | 0.9 |
| | | 5 | 26 | 20 | 20 | 20 | 0.7 | 0.8 | 0.8 |
| 40 | 16 | 1 | 43 | 39 | 39 | 39 | 1.0 | 1.1 | 1.1 |
| | | 3 | 51 | 38 | 38 | 39 | 0.9 | 1.1 | 1.1 |
| | | 5 | 59 | 37 | 38 | 38 | 0.7 | 0.9 | 0.9 |
| | 32 | 1 | 42 | 40 | 40 | 40 | 1.0 | 1.0 | 1.0 |
| | | 3 | 45 | 39 | 40 | 40 | 0.9 | 1.0 | 1.0 |
| | | 5 | 49 | 39 | 40 | 40 | 0.7 | 0.8 | 0.8 |
| | 64 | 1 | 41 | 40 | 40 | 40 | 1.0 | 1.0 | 1.0 |
| | | 3 | 43 | 40 | 40 | 40 | 0.9 | 1.0 | 1.0 |
| | | 5 | 45 | 40 | 40 | 40 | 0.8 | 0.9 | 0.9 |
| 60 | 16 | 1 | 62 | 60 | 60 | 60 | 1.0 | 1.1 | 1.1 |
| | | 3 | 67 | 59 | 59 | 60 | 0.9 | 1.1 | 1.1 |
| | | 5 | 73 | 59 | 59 | 59 | 0.6 | 1.0 | 1.0 |
| | 32 | 1 | 62 | 60 | 60 | 60 | 1.0 | 1.0 | 1.0 |
| | | 3 | 64 | 60 | 60 | 60 | 0.9 | 1.0 | 1.0 |
| | | 5 | 66 | 60 | 60 | 60 | 0.7 | 0.9 | 0.9 |
| | 64 | 1 | 61 | 60 | 60 | 60 | 1.0 | 1.0 | 1.0 |
| | | 3 | 62 | 60 | 60 | 60 | 0.9 | 1.0 | 1.0 |
| | | 5 | 63 | 60 | 60 | 60 | 0.8 | 0.9 | 0.9 |
| 80 | 16 | 1 | 82 | 80 | 80 | 80 | 1.0 | 1.1 | 1.1 |
| | | 3 | 84 | 80 | 80 | 80 | 0.7 | 1.1 | 1.1 |
| | | 5 | 87 | 80 | 80 | 80 | 0.5 | 1.0 | 0.9 |
| | 32 | 1 | 81 | 80 | 80 | 80 | 1.0 | 1.0 | 1.0 |
| | | 3 | 82 | 80 | 80 | 80 | 0.8 | 1.0 | 1.0 |
| | | 5 | 83 | 80 | 80 | 80 | 0.7 | 0.9 | 0.9 |
| | 64 | 1 | 81 | 80 | 80 | 80 | 1.0 | 1.0 | 1.0 |
| | | 3 | 81 | 80 | 80 | 80 | 0.9 | 1.0 | 1.0 |
| | | 5 | 82 | 80 | 80 | 80 | 0.8 | 0.9 | 0.9 |

Table 4
Simulated results for overdispersed Poisson model with $\sqrt{\phi}=2$, 1000 replications. Mean values (in percent) and mean squared error ratios of pseudo $R$-squared measures are given

| $R^2_{D,1}$ | $n$ | $k$ | mean $R^2_D$ | mean $R^2_{D,\gamma}$ | mean $R^2_{D,\gamma_D}$ | mean $R^2_{D,\gamma_P}$ | $\frac{mse(R^2_{D,\gamma})}{mse(R^2_D)}$ | $\frac{mse(R^2_{D,\gamma_D})}{mse(R^2_D)}$ | $\frac{mse(R^2_{D,\gamma_P})}{mse(R^2_D)}$ |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 16 | 1 | 5 | 4 | −1 | 0 | 0.8 | 0.7 | 0.7 |
| | | 3 | 17 | 12 | −4 | 0 | 0.7 | 0.6 | 0.5 |
| | | 5 | 29 | 21 | −6 | −1 | 0.6 | 0.5 | 0.4 |
| | 32 | 1 | 3 | 2 | −1 | 0 | 0.8 | 0.7 | 0.7 |
| | | 3 | 8 | 6 | −1 | 0 | 0.7 | 0.5 | 0.4 |
| | | 5 | 14 | 10 | −2 | 0 | 0.7 | 0.4 | 0.3 |
| | 64 | 1 | 1 | 1 | 0 | 0 | 0.8 | 0.7 | 0.7 |
| | | 3 | 4 | 3 | −1 | 0 | 0.7 | 0.4 | 0.4 |
| | | 5 | 7 | 5 | −1 | 0 | 0.7 | 0.4 | 0.3 |
| 20 | 16 | 1 | 24 | 23 | 19 | 20 | 1.0 | 1.1 | 1.0 |
| | | 3 | 34 | 30 | 18 | 20 | 0.8 | 0.9 | 0.9 |
| | | 5 | 44 | 37 | 16 | 19 | 0.7 | 0.8 | 0.7 |
| | 32 | 1 | 22 | 21 | 19 | 20 | 1.0 | 1.0 | 1.0 |
| | | 3 | 27 | 25 | 19 | 20 | 0.9 | 0.9 | 0.9 |
| | | 5 | 31 | 28 | 18 | 20 | 0.8 | 0.7 | 0.7 |
| | 64 | 1 | 21 | 21 | 20 | 20 | 1.0 | 1.0 | 1.0 |
| | | 3 | 23 | 23 | 20 | 20 | 0.9 | 0.9 | 0.9 |
| | | 5 | 26 | 24 | 19 | 20 | 0.9 | 0.8 | 0.8 |
| 40 | 16 | 1 | 44 | 43 | 40 | 40 | 1.0 | 1.1 | 1.1 |
| | | 3 | 51 | 48 | 39 | 40 | 0.9 | 1.0 | 1.0 |
| | | 5 | 58 | 54 | 38 | 40 | 0.7 | 0.9 | 0.9 |
| | 32 | 1 | 42 | 41 | 40 | 40 | 1.0 | 1.0 | 1.0 |
| | | 3 | 45 | 44 | 39 | 40 | 0.9 | 1.0 | 1.0 |
| | | 5 | 49 | 47 | 39 | 40 | 0.8 | 0.8 | 0.8 |
| | 64 | 1 | 41 | 41 | 40 | 40 | 1.0 | 1.0 | 1.0 |
| | | 3 | 43 | 42 | 40 | 40 | 1.0 | 1.0 | 1.0 |
| | | 5 | 44 | 43 | 40 | 40 | 0.9 | 0.9 | 0.9 |
| 60 | 16 | 1 | 63 | 62 | 60 | 61 | 1.0 | 1.1 | 1.1 |
| | | 3 | 68 | 66 | 60 | 61 | 0.9 | 1.1 | 1.0 |
| | | 5 | 73 | 70 | 59 | 61 | 0.8 | 1.0 | 0.9 |
| | 32 | 1 | 62 | 61 | 60 | 60 | 1.0 | 1.0 | 1.0 |
| | | 3 | 64 | 63 | 60 | 60 | 0.9 | 1.0 | 1.0 |
| | | 5 | 66 | 65 | 60 | 60 | 0.8 | 0.9 | 0.8 |
| | 64 | 1 | 61 | 61 | 60 | 60 | 1.0 | 1.0 | 1.0 |
| | | 3 | 62 | 62 | 60 | 60 | 1.0 | 1.0 | 1.0 |
| | | 5 | 63 | 63 | 60 | 60 | 0.9 | 0.9 | 0.9 |
| 80 | 16 | 1 | 82 | 81 | 80 | 80 | 1.0 | 1.1 | 1.1 |
| | | 3 | 84 | 83 | 80 | 80 | 0.9 | 1.1 | 1.0 |
| | | 5 | 86 | 85 | 79 | 80 | 0.8 | 1.0 | 0.9 |
| | 32 | 1 | 81 | 81 | 80 | 80 | 1.0 | 1.0 | 1.0 |
| | | 3 | 82 | 82 | 80 | 80 | 0.9 | 1.0 | 1.0 |
| | | 5 | 83 | 82 | 80 | 80 | 0.9 | 0.9 | 0.9 |
| | 64 | 1 | 80 | 80 | 80 | 80 | 1.0 | 1.0 | 1.0 |
| | | 3 | 81 | 81 | 80 | 80 | 1.0 | 1.0 | 1.0 |
| | | 5 | 82 | 81 | 80 | 80 | 0.9 | 0.9 | 0.9 |

Table 5
Simulated results for overdispersed Poisson model with $\sqrt{\phi}=3$, 1000 replications. Mean values (in percent) and mean squared error ratios of pseudo $R$-squared measures are given

| $R^2_{D,1}$ | $n$ | $k$ | mean $R^2_D$ | mean $R^2_{D,\gamma}$ | mean $R^2_{D,\gamma_D}$ | mean $R^2_{D,\gamma_P}$ | $\frac{mse(R^2_{D,\gamma})}{mse(R^2_D)}$ | $\frac{mse(R^2_{D,\gamma_D})}{mse(R^2_D)}$ | $\frac{mse(R^2_{D,\gamma_P})}{mse(R^2_D)}$ |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 16 | 1 | 6 | 5 | −1 | 0 | 0.9 | 0.8 | 0.7 |
| | | 3 | 18 | 16 | −2 | 0 | 0.9 | 0.6 | 0.5 |
| | | 5 | 30 | 27 | −4 | −1 | 0.8 | 0.5 | 0.4 |
| | 32 | 1 | 3 | 2 | −1 | 0 | 0.9 | 0.7 | 0.7 |
| | | 3 | 9 | 8 | −1 | 0 | 0.9 | 0.5 | 0.5 |
| | | 5 | 14 | 13 | −2 | 0 | 0.8 | 0.4 | 0.3 |
| | 64 | 1 | 1 | 1 | 0 | 0 | 0.9 | 0.7 | 0.7 |
| | | 3 | 4 | 4 | −1 | 0 | 0.9 | 0.5 | 0.4 |
| | | 5 | 7 | 6 | −1 | 0 | 0.9 | 0.4 | 0.3 |
| 20 | 16 | 1 | 24 | 23 | 19 | 19 | 1.0 | 1.1 | 1.0 |
| | | 3 | 34 | 32 | 17 | 19 | 0.9 | 0.9 | 0.8 |
| | | 5 | 43 | 40 | 15 | 19 | 0.9 | 0.8 | 0.7 |
| | 32 | 1 | 22 | 22 | 19 | 20 | 1.0 | 1.0 | 1.0 |
| | | 3 | 26 | 26 | 18 | 20 | 0.9 | 0.9 | 0.9 |
| | | 5 | 31 | 30 | 18 | 19 | 0.9 | 0.8 | 0.7 |
| | 64 | 1 | 21 | 21 | 20 | 20 | 1.0 | 1.0 | 1.0 |
| | | 3 | 23 | 23 | 19 | 20 | 1.0 | 1.0 | 0.9 |
| | | 5 | 25 | 25 | 19 | 20 | 0.9 | 0.8 | 0.8 |
| 40 | 16 | 1 | 43 | 43 | 39 | 39 | 1.0 | 1.1 | 1.1 |
| | | 3 | 50 | 49 | 38 | 39 | 0.9 | 1.1 | 1.0 |
| | | 5 | 58 | 56 | 36 | 39 | 0.9 | 0.9 | 0.8 |
| | 32 | 1 | 41 | 41 | 39 | 40 | 1.0 | 1.1 | 1.0 |
| | | 3 | 45 | 44 | 39 | 40 | 1.0 | 1.0 | 1.0 |
| | | 5 | 48 | 47 | 39 | 40 | 0.9 | 0.9 | 0.8 |
| | 64 | 1 | 41 | 41 | 40 | 40 | 1.0 | 1.0 | 1.0 |
| | | 3 | 43 | 42 | 40 | 40 | 1.0 | 1.0 | 1.0 |
| | | 5 | 44 | 44 | 40 | 40 | 1.0 | 0.9 | 0.9 |
| 60 | 16 | 1 | 62 | 62 | 60 | 60 | 1.0 | 1.1 | 1.1 |
| | | 3 | 67 | 66 | 59 | 60 | 0.9 | 1.1 | 1.0 |
| | | 5 | 72 | 71 | 58 | 60 | 0.9 | 1.0 | 0.9 |
| | 32 | 1 | 61 | 61 | 60 | 60 | 1.0 | 1.0 | 1.0 |
| | | 3 | 64 | 63 | 60 | 60 | 1.0 | 1.0 | 1.0 |
| | | 5 | 66 | 65 | 59 | 60 | 0.9 | 0.9 | 0.8 |
| | 64 | 1 | 61 | 61 | 60 | 60 | 1.0 | 1.0 | 1.0 |
| | | 3 | 62 | 62 | 60 | 60 | 1.0 | 1.0 | 1.0 |
| | | 5 | 63 | 63 | 60 | 60 | 1.0 | 0.9 | 0.9 |
| 80 | 16 | 1 | 82 | 81 | 80 | 80 | 1.0 | 1.1 | 1.1 |
| | | 3 | 84 | 84 | 80 | 80 | 0.9 | 1.0 | 1.0 |
| | | 5 | 86 | 86 | 80 | 80 | 0.9 | 0.8 | 0.8 |
| | 32 | 1 | 81 | 81 | 80 | 80 | 1.0 | 1.0 | 1.0 |
| | | 3 | 82 | 82 | 80 | 80 | 1.0 | 1.0 | 1.0 |
| | | 5 | 83 | 83 | 80 | 80 | 0.9 | 0.8 | 0.8 |
| | 64 | 1 | 80 | 80 | 80 | 80 | 1.0 | 1.0 | 1.0 |
| | | 3 | 81 | 81 | 80 | 80 | 1.0 | 1.0 | 1.0 |
| | | 5 | 82 | 81 | 80 | 80 | 1.0 | 0.9 | 0.9 |

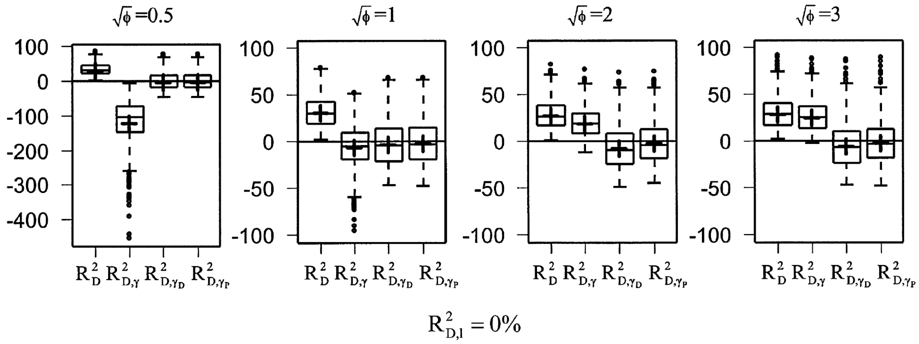


Fig. 1. Boxplots of the distributions of 1000 simulated samples for unadjusted and adjusted pseudo $R$-squared values (in percent) for $n = 16$, $k = 5$, $\sqrt{\phi} = 0.5$, 1, 2 and 3, respectively. Horizontal lines are drawn for true population value $R^2_{\mathrm{D,I}} = 0$ percent.

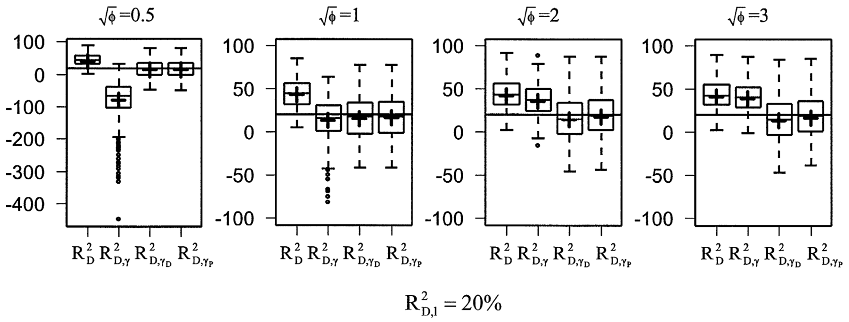


Fig. 2. Boxplots of the distributions of 1000 simulated samples for unadjusted and adjusted pseudo $R$-squared values (in percent) for $n = 16$, $k = 5$, $\sqrt{\phi} = 0.5$, 1, 2 and 3, respectively. Horizontal lines are drawn for true population value $R_{\mathrm{D,I}} = 20$ percent.

of the failure of $R^2_{\mathrm{D},\gamma}$ increases with the number of covariates $k$, decreasing sample size $n$ and decreasing true value $R^2_{\mathrm{D,I}}$. The simulation results confirm the considerations in Section 3: In the case of overdispersion, the shrinkage of $R^2_{\mathrm{D},\gamma}$ is insufficient, whereas extensive shrinkage of $R^2_{\mathrm{D},\gamma}$ will be observed in the case of underdispersion, see also Figs. 1–5.

In the case of standard Poisson models ($\phi = 1$), all three adjusted $R$-squared measures may reduce bias in a satisfying manner (Table 3). With a focus on bias adjustment, the impression could arise that both $R^2_{\mathrm{D},\gamma_{\mathrm{D}}}$ and $R^2_{\mathrm{D},\gamma_{\mathrm{P}}}$ are able to outperform the standard adjusted $R$-squared measure, $R^2_{\mathrm{D},\gamma}$, even in the standard Poisson model. However, the mean squared error ratios neutralize this impression. In particular, comparing the mse ratios of adjusted $R$-squared measures for the small samples ($n = 16$) in Table 3 shows in a focused manner that overfitting the model by adding a parameter without necessity may pay off only slightly for $R^2_{\mathrm{D,I}} = 0$, but becomes increasingly disadvantageous for

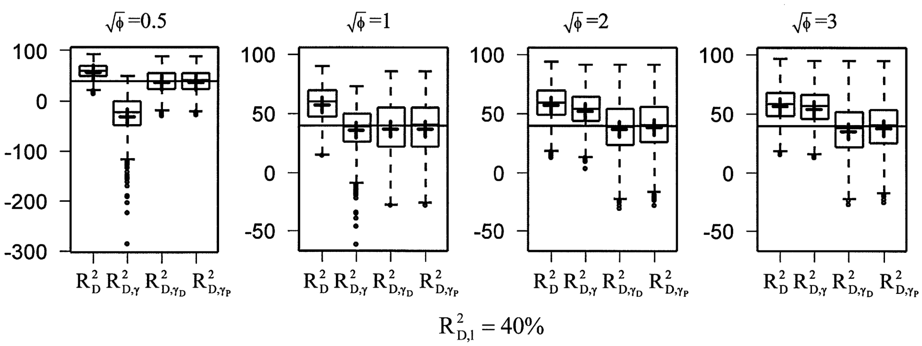


Fig. 3. Boxplots of the distributions of 1000 simulated samples for unadjusted and adjusted pseudo $R$-squared values (in percent) for $n = 16$, $k = 5$, $\sqrt{\phi} = 0.5$, 1, 2 and 3, respectively. Horizontal lines are drawn for true population value $R^2_{D,1} = 40$ percent.

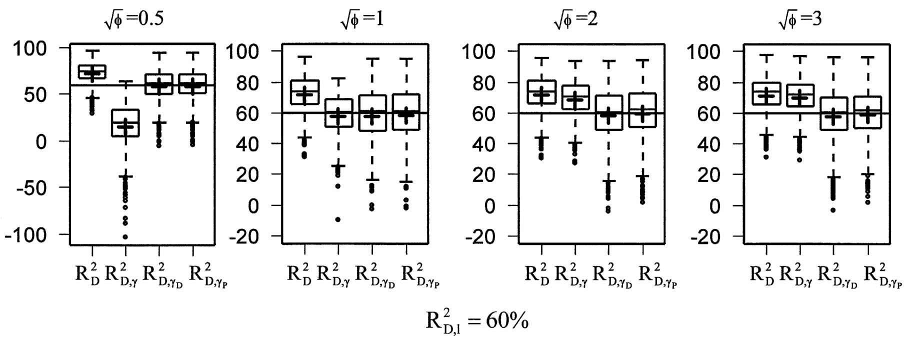


Fig. 4. Boxplots of the distributions of 1000 simulated samples for unadjusted and adjusted pseudo $R$-squared values (in percent) for $n = 16$, $k = 5$, $\sqrt{\phi} = 0.5$, 1, 2 and 3, respectively. Horizontal lines are drawn for true population value $R^2_{D,1} = 60$ percent.

increasing values of $R^2_{D,1}$ (see also Figs. 1–5). If the mse ratio values for $R^2_{D,\gamma_D}$ and $R^2_{D,\gamma_P}$ are around one, then these adjusted measures will have no beneficial effect in terms of mse as compared to the unadjusted measure $R^2_D$.

But even in the cases of under- or overdispersion (Tables 2, 4 and 5), mse($R^2_{D,\gamma_D}$) and mse($R^2_{D,\gamma_P}$) are only smaller than mse($R^2_D$) for small values of $R^2_{D,1}$ or increased numbers of covariates $k$ (note that one covariate at the most is designed to influence the outcome, the others are nuisance variables). That is, reducing smaller amounts of bias does not pay off since the inevitable inflation of the variance will absorb the positive effects. Figs. 1–5 confirm these findings graphically.

Table 6 shows the results for the tmse. The arising situation is noteworthy: Given a common-sense adjustment is applied to improve a poorly performing estimator, then

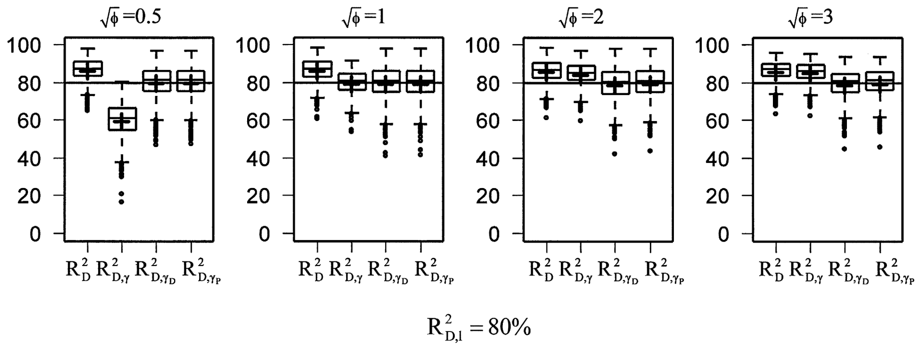


Fig. 5. Boxplots of the distributions of 1000 simulated samples for unadjusted and adjusted pseudo $R$-squared values (in percent) for $n = 16$, $k = 5$, $\sqrt{\phi} = 0.5$, 1, 2 and 3, respectively. Horizontal lines are drawn for true population value $R^2_{D,l} = 80$ percent.

the worse the estimator will perform before the adjustment, the better it may perform afterwards. Although $R^2_{D,\gamma}$ performs unacceptably in terms of mse, it performs best in terms of tmse in the case of underdispersion ($\sqrt{\phi} = 0.5$) and $R^2_{D,l} = 0$ percent. For the case of $n = 16$ and $k = 5$, in particular, the tmse ($R^2_{D,\gamma}$) precisely equals zero, since all estimated values for $R^2_{D,\gamma}$ of the 1000 samples are lesser than zero. The ordinary setting is already restored with a true $R$-squared value $R^2_{D,l}$ increased to a value of 20 percent (Table 6). It is not feasible to recommend the use of the truncated measure $R^2_{D,\gamma}$ for true values $R^2_{D,l}$ around 0 percent, since we do not know the true $R$-squared values, but rather have biased estimates $R^2_D$. The effect of truncating $R^2_{D,\gamma}$ at zero in the case of overdispersed Poisson data is of no bearing since the naive shrinkage factor $\gamma$ will prove too large, such that insufficient shrinkage will occur and negative values of $R^2_{D,\gamma}$ will prove to be infrequent (see Section 3, Table 6 and Figs. 1 and 2). On the other hand, the adjusted $R$-squared measures that account for under- or overdispersion, $R^2_{D,\gamma_D}$ and $R^2_{D,\gamma_P}$, take uniform advantage of the truncation over the range of $\sqrt{\phi}$ for true values $R^2_{D,l}$ around zero (Table 6).

## 5. Example

We draw on the popcorn example presented by Myers et al. (2002) to illustrate our approach. The impact of popping temperature, amount of oil, and the popping time on the number of inedible kernels of popcorn was studied in 15 single trials. Table 7 gives estimated results for three Poisson regression models (the common significance level $\alpha = 0.05$ is applied in the following). Firstly, only main effects are considered, secondly, a statistically significant interaction term is added, and thirdly, two pairwise interaction terms are added. Two situations are considered for all models: (a) obvious overdispersion is ignored and the naive standard error estimates and $p$-values for the parameter estimates are given, and (b) standard errors and $p$-values are adjusted for

Table 6
Simulated results for all models with $R^2_{D,1} = 0$ and 20 percent, 1000 replications. Ratios of truncated mean squared errors of pseudo $R$-squared measures are given

| | | | $R^2_{D,1} = 0\%$ | | | $R^2_{D,1} = 20\%$ | | |
|---|---|---|---|---|---|---|---|---|
| $\sqrt{\phi}$ | $n$ | $k$ | $\frac{\text{tmse}(R^2_{D,\gamma})}{\text{tmse}(R^2_D)}$ | $\frac{\text{tmse}(R^2_{D,\gamma_D})}{\text{tmse}(R^2_D)}$ | $\frac{\text{tmse}(R^2_{D,\gamma_P})}{\text{tmse}(R^2_D)}$ | $\frac{\text{tmse}(R^2_{D,\gamma})}{\text{tmse}(R^2_D)}$ | $\frac{\text{tmse}(R^2_{D,\gamma_D})}{\text{tmse}(R^2_D)}$ | $\frac{\text{tmse}(R^2_{D,\gamma_P})}{\text{tmse}(R^2_D)}$ |
| 0.5 | 16 | 1 | 0.1 | 0.5 | 0.5 | 1.1 | 1.0 | 1.0 |
| | | 3 | 0.0 | 0.3 | 0.3 | 0.8 | 0.7 | 0.7 |
| | | 5 | 0.0 | 0.2 | 0.2 | 0.4 | 0.5 | 0.5 |
| | 32 | 1 | 0.0 | 0.5 | 0.5 | 1.3 | 1.0 | 1.0 |
| | | 3 | 0.0 | 0.2 | 0.2 | 1.7 | 0.9 | 0.9 |
| | | 5 | 0.0 | 0.2 | 0.2 | 1.3 | 0.6 | 0.6 |
| | 64 | 1 | 0.0 | 0.5 | 0.5 | 1.4 | 1.0 | 1.0 |
| | | 3 | 0.0 | 0.2 | 0.3 | 2.4 | 1.0 | 1.0 |
| | | 5 | 0.0 | 0.2 | 0.2 | 2.8 | 0.8 | 0.8 |
| 1 | 16 | 1 | 0.4 | 0.5 | 0.5 | 0.9 | 1.0 | 1.0 |
| | | 3 | 0.2 | 0.3 | 0.3 | 0.6 | 0.7 | 0.7 |
| | | 5 | 0.1 | 0.2 | 0.2 | 0.3 | 0.5 | 0.5 |
| | 32 | 1 | 0.4 | 0.5 | 0.5 | 1.0 | 1.0 | 1.0 |
| | | 3 | 0.2 | 0.3 | 0.3 | 0.8 | 0.9 | 0.9 |
| | | 5 | 0.1 | 0.2 | 0.2 | 0.5 | 0.6 | 0.6 |
| | 64 | 1 | 0.5 | 0.5 | 0.5 | 1.0 | 1.0 | 1.0 |
| | | 3 | 0.2 | 0.3 | 0.3 | 0.9 | 0.9 | 0.9 |
| | | 5 | 0.2 | 0.2 | 0.2 | 0.7 | 0.8 | 0.8 |
| 2 | 16 | 1 | 0.8 | 0.5 | 0.5 | 1.0 | 1.0 | 1.0 |
| | | 3 | 0.7 | 0.3 | 0.3 | 0.8 | 0.7 | 0.7 |
| | | 5 | 0.6 | 0.2 | 0.2 | 0.7 | 0.5 | 0.5 |
| | 32 | 1 | 0.8 | 0.5 | 0.5 | 1.0 | 1.0 | 1.0 |
| | | 3 | 0.7 | 0.2 | 0.3 | 0.9 | 0.9 | 0.9 |
| | | 5 | 0.7 | 0.2 | 0.2 | 0.8 | 0.7 | 0.6 |
| | 64 | 1 | 0.8 | 0.5 | 0.5 | 1.0 | 1.0 | 1.0 |
| | | 3 | 0.7 | 0.2 | 0.3 | 0.9 | 0.9 | 0.9 |
| | | 5 | 0.7 | 0.2 | 0.2 | 0.9 | 0.8 | 0.8 |
| 3 | 16 | 1 | 0.9 | 0.5 | 0.5 | 1.0 | 1.0 | 1.0 |
| | | 3 | 0.9 | 0.3 | 0.3 | 0.9 | 0.7 | 0.7 |
| | | 5 | 0.8 | 0.2 | 0.2 | 0.9 | 0.5 | 0.5 |
| | 32 | 1 | 0.9 | 0.5 | 0.5 | 1.0 | 1.0 | 1.0 |
| | | 3 | 0.9 | 0.3 | 0.3 | 0.9 | 0.9 | 0.9 |
| | | 5 | 0.8 | 0.2 | 0.2 | 0.9 | 0.7 | 0.6 |
| | 64 | 1 | 0.9 | 0.5 | 0.5 | 1.0 | 1.0 | 1.0 |
| | | 3 | 0.9 | 0.3 | 0.3 | 1.0 | 1.0 | 0.9 |
| | | 5 | 0.9 | 0.2 | 0.2 | 0.9 | 0.8 | 0.8 |

overdispersion by using the dispersion parameter estimate $\hat{\phi}_P$ which is based on generalized Pearson statistic. The differences between the dispersion parameter estimates $\sqrt{\hat{\phi}_P} = 3.6570$, 2.3809 and 2.2042, and $\sqrt{\hat{\phi}_D} = 3.4822$, 2.3653 and 2.1423, are rather

Table 7
Results of Poisson regression models for the popcorn experiment of Myers et al. (2002): (a) overdispersion is ignored, and (b) overdispersion is taken into account and the dispersion parameter estimate $\hat{\phi}_{\mathrm{P}}$ is used which is based on generalized Pearson statistic (the results for the deviance-based estimate $\hat{\phi}_{\mathrm{D}}$ are similar and therefore not shown)

| Parameter | df | Estimate | Standard error | $p$-value | Standard error | $p$-value |
|---|---|---|---|---|---|---|
| *Model 1* | | | (a) $\phi = 1$ (fixed) | | (b) $\sqrt{\hat{\phi}_{\mathrm{P}}} = 3.6570$ (estimated) | |
| Intercept | 1 | 8.2439 | 0.5310 | — | 1.9418 | — |
| Temperature | 1 | −0.3462 | 0.0605 | <0.0001 | 0.2211 | 0.1173 |
| Oil | 1 | −0.0885 | 0.0595 | 0.1372 | 0.2178 | 0.6844 |
| Time | 1 | −0.0267 | 0.0041 | <0.0001 | 0.0148 | 0.0712 |
| *Model 2* | | | (a) $\phi = 1$ (fixed) | | (b) $\sqrt{\hat{\phi}_{\mathrm{P}}} = 2.3809$ (estimated) | |
| Intercept | 1 | 33.9112 | 2.9387 | — | 6.9968 | — |
| Temperature | 1 | −4.7314 | 0.4981 | <0.0001 | 1.1858 | <0.0001 |
| Oil | 1 | −0.0967 | 0.0623 | 0.1204 | 0.1482 | 0.5142 |
| Time | 1 | −0.3207 | 0.0334 | <0.0001 | 0.0796 | <0.0001 |
| Temperature*Time | 1 | 0.0501 | 0.0056 | <0.0001 | 0.0134 | 0.0002 |
| *Model 3* | | | (a) $\phi = 1$ (fixed) | | (b) $\sqrt{\hat{\phi}_{\mathrm{P}}} = 2.2042$ (estimated) | |
| Intercept | 1 | 24.9964 | 3.6530 | — | 8.0520 | — |
| Temperature | 1 | −4.2817 | 0.5679 | <0.0001 | 1.2519 | 0.0006 |
| Oil | 1 | 2.9898 | 0.7655 | <0.0001 | 1.6872 | 0.0764 |
| Time | 1 | −0.2517 | 0.0373 | <0.0001 | 0.0823 | 0.0022 |
| Temperature*Oil | 1 | −0.1634 | 0.0903 | 0.0705 | 0.1991 | 0.4119 |
| Temperature*Time | 1 | 0.0505 | 0.0057 | <0.0001 | 0.0126 | <0.0001 |
| Oil*Time | 1 | −0.0242 | 0.0060 | <0.0001 | 0.0131 | 0.0654 |

small for Models 1, 2 and 3, respectively; the results for the deviance-based estimate $\hat{\phi}_{\mathrm{D}}$ are thus not presented. On account of massive overdispersion, the standard error estimates and consequently the $p$-values are considerably higher in situations (b) than in situations (a).

Model 1 shows two significant covariates to become statistically non-significant after accounting for overdispersion, although both $p$-values of the standard Poisson regression are lesser than 0.0001 (Table 7).

In Model 2, the interaction term between temperature and time has been added. This factor is obviously crucial for the number of inedible kernels, since it remains statistically significant after overdispersion has been taken into account (Table 7).

The remaining two pairwise interaction terms have been added in Model 3. This leads us to a similar situation as shown in Model 1. In the standard Poisson regression, the interaction between oil and time has a p-value lesser than 0.0001, but this factor becomes statistically non-significant after accounting for overdispersion (Table 7).

Table 8 contains the estimates for the $R$-squared measures and shrinkage factors. The unadjusted $R$-squared values $R^2_{\mathrm{D}}$ suggest that 38 percent of the variation in the number of inedible kernels can be explained by temperature, oil and time. Adding temperature–time interaction, this percentage increases to 74 percent and then to 83 percent when adding the other interactions. The standard Poisson shrinkage factor $\hat{\gamma}$ is close to one for all three models, thus suggesting only small shrinkage.

Table 8
Estimated pseudo $R$-squared measures (in percent) and shrinkage factors for the popcorn experiment of Myers et al. (2002): (a) overdispersion is ignored, and (b) overdispersion is taken into account

| | | (a) | (b) | | (a) | (b) | |
|---|---|---|---|---|---|---|---|
| | $R^2_{\mathrm{D}}$ | $R^2_{\mathrm{D},\gamma}$ | $R^2_{\mathrm{D},\gamma_{\mathrm{D}}}$ | $R^2_{\mathrm{D},\gamma_{\mathrm{P}}}$ | $\hat{\gamma}$ | $\hat{\gamma}_{\mathrm{D}}$ | $\hat{\gamma}_{\mathrm{P}}$ |
| Model 1 | 38 | 36 | 20 | 19 | 0.96 | 0.55 | 0.50 |
| Model 2 | 74 | 72 | 63 | 63 | 0.97 | 0.86 | 0.86 |
| Model 3 | 83 | 80 | 70 | 69 | 0.97 | 0.84 | 0.84 |

When accounting for overdispersion, the estimated shrinkage factors $\hat{\gamma}_{\mathrm{D}}$ and $\hat{\gamma}_{\mathrm{P}}$ are smaller than $\hat{\gamma}$. $R^2_{\mathrm{D}}$ is consequently shrunken to $R^2_{\mathrm{D},\gamma_{\mathrm{P}}} = 19$, 63 and 69 percent for Models 1, 2 and 3, respectively. Similar values are obtained for $R^2_{\mathrm{D},\gamma_{\mathrm{D}}}$; i.e. the amount of explained variation nearly halves for Model 1. The reduction of explained variation is less pronounced for Models 2 and 3 that still contain statistically significant factors after accounting for overdispersion (Tables 7 and 8).

## 6. Discussion

Trivial reasons for the occurrence of over- or underdispersion in a Poisson regression model include incorrectly specified models, missing covariates and outliers in the data. Adequate adjustments for over- or underdispersion could be performed upon ruling out such causes after model checking. As demonstrated with the popcorn example (Section 5), these adjustments affect the standard errors of the parameter estimates, and consequently the width of the confidence intervals, the magnitude of the $p$-values and the bias adjustment for the pseudo $R$-squared measure $R^2_{\mathrm{D}}$.

We recommend the use of $R^2_{\mathrm{D},\gamma_{\mathrm{P}}}$ to measure the amount of explained variation in over- and underdispersed Poisson regression models. Based on an extensive simulation study (Section 4), and exemplified with popcorn data (Section 5), we have shown that the common pseudo $R$-squared measure $R^2_{\mathrm{D}}$ could be heavily biased, particularly in the presence of small sample sizes or various unimportant covariates. In contrast, $R^2_{\mathrm{D},\gamma_{\mathrm{P}}}$ is unbiased even in such extreme situations.

$R^2_{\mathrm{D},\gamma_{\mathrm{P}}}$ does not always perform best in terms of mean squared error. For instance, in larger samples with a strong predictor or in data sets with hardly any over- or underdispersion, the mean squared errors of the common unadjusted or adjusted $R$-squared measures, $\mathrm{mse}(R^2_{\mathrm{D}})$ and $\mathrm{mse}(R^2_{\mathrm{D},\gamma})$, may feature lesser values than $\mathrm{mse}(R^2_{\mathrm{D},\gamma_{\mathrm{P}}})$. However, even when $R^2_{\mathrm{D},\gamma_{\mathrm{P}}}$ is not the best choice, it will still be a reasonable choice in terms of the mean squared error. In contrast, $R^2_{\mathrm{D},\gamma}$ will completely break down in the case of underdispersion.

We have also investigated the performance of the adjusted $R$-squared definition $R^2_{\mathrm{D},\gamma_{\mathrm{D}}}$ that is similar to $R^2_{\mathrm{D},\gamma_{\mathrm{P}}}$. We do not recommend its use, however, since $R^2_{\mathrm{D},\gamma_{\mathrm{D}}}$ never performs better and at times slightly worse than $R^2_{\mathrm{D},\gamma_{\mathrm{P}}}$. The performance differences between $R^2_{D,\gamma_P}$ and $R^2_{D,\gamma_D}$ stem from the estimation of the dispersion parameter $\phi$. Note that our simulation results confirm McCullagh and Nelder (1989) who recommend the use of the moment estimator $\hat{\phi}_P$, which is based on the more robust generalized Pearson statistic.

Note here that the dispersion parameter estimates are based on the number of covariates $k$ in the full model. With respect to stepwise modeling, Harrell et al. (1996) suggest to correct for the number of covariates under investigation rather than for the number of covariates in the final model (the full model).

An adjusted $R$-squared measure estimate may occasionally result in a negative value. This applies when the unadjusted $R$-squared value is smaller than the bias adjustment, which is an expected value given the null hypothesis of no covariates effect. A value of zero should then be reported, indicating that no variation whatsoever is explained.

An $R$-squared measure allows quantifying the predictive power of a regression model. A large $R$-squared value may discourage investigators from seeking additional prognostic factors, so that the use of proper bias adjustments is recommended.

## Acknowledgements

We would like to thank two referees for their constructive remarks and useful comments.

## References

Cameron, A.C., Windmeijer, F.A.G., 1996. $R^2$ measures for count data regression models with applications to health-care utilization. J. Bus. Econom. Statist. 14, 209–220.

Copas, J.B., 1983. Regression, prediction and shrinkage (with discussion). J. Roy. Statist. Soc. B 45, 311–354.

Copas, J.B., 1987. Cross-validation shrinkage of regression predictors. J. Roy. Statist. Soc. B 49, 175–183.

Copas, J.B., 1997. Using regression models for prediction: shrinkage and regression to the mean. Statist. Methods Med. Res. 6, 165–183.

Cox, D.R., 1983. Some remarks on overdispersion. Biometrika 70, 269–274.

Harrell, F.E., Lee, K.L., Mark, D.B., 1996. Tutorial in biostatistics: multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statist. Med. 15, 361–387.

Heinzl, H., Mittlböck, M., 2002. $R$-squared measures for the inverse Gaussian regression model. Comput. Statist. 17, 525–544.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models. Chapman & Hall Ltd, London.

Mittlböck, M., 2002. Calculating adjusted $R^2$ measures for Poisson regression models. Comput. Methods Programs Biomed. 68, 205–214.

Mittlböck, M., Heinzl, H., 2001. A note on $R^2$ measures for Poisson and logistic regression models when both models are applicable. J. Clin. Epidemiol. 54, 99–103.

Mittlböck, M., Heinzl, H., 2002. Measures of explained variation in Gamma regression model. Comm. Statist. Simulation Comput. 31, 61–73.

Mittlböck, M., Waldhör, T., 2000. Adjustments for $R^2$-measures for Poisson regression models. Comput. Statist. Data Anal. 34, 461–472.

Myers, R.H., Montgomery, D.C., Vining, G.G., 2002. Generalized Linear Models—with Applications in Engineering and the Sciences. Wiley-Interscience, New York.

Paul, S.R., Plackett, R.L., 1978. Inference sensitivity for Poisson mixtures. Biometrika 65, 591–602.

Van Houwelingen, J.C., Le Cessie, S., 1990. Predictive value of statistical models. Statist. Med. 9, 1303–1325.

Waldhör, T., Haidinger, G., Schober, E., 1998. Comparison of $R^2$ measures for Poisson regression by simulation. J. Epidemiol. Biostatist. 3, 209–215.

Waldhör, T., Mittlböck, M., Haidinger, G., Zidek, T., Schober, E., 2001. Partial $R^2$-values based on deviance residuals in Poisson regression models. Inform. Biometrie Epidemiol. Med. Biol. 32, 341–347.