

# **R-Squared Measures for Count Data Regression Models With Applications to Health Care Utilization**

A. Colin Cameron  
Dept. of Economics  
University of California  
Davis CA 95616-8578  
USA

Frank A.G. Windmeijer  
Dept. of Economics  
University College London  
London WC1E 6BT  
UK

April 1995

Journal of Business and Economic Statistics (forthcoming)

## **Abstract**

R-squared measures of goodness of fit for count data are rarely, if ever, reported in empirical studies or by statistical packages. We propose several R-squared measures based on various definitions of residuals, for the basic Poisson regression model and for more general models such as negative binomial that accommodate overdispersed data. The preferred R-squared is based on the deviance residual. An application to data on health care service utilization measured in counts illustrates the performance and usefulness of the various R-squareds.

**KEY WORDS:** Goodness-of-fit, Poisson regression, negative binomial regression, deviance, deviance residual, Pearson residual.

## 1. INTRODUCTION

R-squared ( $R^2$ ) measures were originally developed for linear regression models with homoscedastic errors. Extensions to models with heteroscedastic errors with known variance were proposed by Buse (1973). Extensions to other models are rare, with the notable exceptions of logit and probit models, see Windmeijer (1994) and the references therein, and tobit models, surveyed by Veall and Zimmermann (1994).

In this paper we investigate  $R^2$  measures for Poisson and other related count data regression models. Surprisingly,  $R^2$  is rarely reported in empirical studies or by statistical packages for count data. (For Poisson, exceptions are Merkle and Zimmermann (1992) and the statistical package STATA. These are discussed in section 2.6). Instead the standard measures of goodness of fit for Poisson regression models are the deviance and Pearson's statistic. These two statistics are widely used by generalized linear model practitioners, see McCullagh and Nelder (1989), and seldom used in econometrics applications.

We propose several  $R^2$  measures based on various definitions of residuals. These measures are intended to measure goodness of fit within a particular type of count data model, e.g. Poisson, rather than across model types, e.g. Poisson versus negative binomial. We distinguish between various  $R^2$  on the following criteria:

- (1)  $0 \leq R^2 \leq 1$
- (2)  $R^2$  does not decrease as regressors are added (without degree of freedom correction)
- (3)  $R^2$  based on residual sum of squares coincides with  $R^2$  based on explained sum of squares
- (4) There is a correspondence between  $R^2$  and a significance test on all slope parameters, and between changes in  $R^2$  as regressors are added and significance tests
- (5)  $R^2$  has an interpretation in terms of information content of the data.

Criterion 3 is the Pythagorean relationship discussed by Efron (1978) for logit models with grouped data. Criterion 4 is used by Dhrymes (1986) for logit and probit models.

$R^2$  measures for the Poisson regression model are presented and discussed in detail in section 2. The preferred measure is one based on deviance residuals. Non-trivial extensions to negative binomial models are presented in section 3. The empirical performance of  $R^2$  measures is analyzed in section 4 in an application to the determinants of individual utilization of health care services recorded as counts. Conclusions are given in section 5.

## 2. R-SQUARED FOR POISSON MODEL

### 2.1 Poisson Model and Residuals

We begin with the Poisson regression model, see for example Cameron and Trivedi (1986). The dependent variable  $y_i$ ,  $i = 1, \dots, N$ , is independent Poisson distributed with log-density

$$l_i(\mu_i) = -\mu_i + y_i \log \mu_i - \log y_i!, \quad (2.1)$$

where for brevity the dependence of  $l_i$  on  $y_i$  is suppressed throughout, with conditional mean

$$E[y_i | X_i] = \mu_i = \mu(X_i, \beta), \quad (2.2)$$

where  $\mu(\cdot)$  is a specified function,  $X_i$  is a vector of exogenous regressors which throughout we assume includes a constant term, and  $\beta$  is a  $k \times 1$  parameter vector. For this model the conditional variance equals the conditional mean

$$\text{Var}(y_i | X_i) = \mu_i. \quad (2.3)$$

The fitted value of  $y_i$  is denoted  $\hat{\mu}_i = \mu(X_i, \hat{\beta})$  where  $\hat{\beta}$  is the maximum likelihood (ML) estimator of  $\beta$ . It is customary to specify the conditional mean as  $\mu_i = \exp(X_i' \beta)$ . Then, since  $X_i$  includes a constant, the ML first-order conditions imply

$$\sum_{i=1}^N (y_i - \hat{\mu}_i) = 0. \quad (2.4)$$

Formulae for several of the  $R^2$  measures below simplify if (2.4) holds. In the intercept-only Poisson model the individual predicted mean is  $\bar{y}$ , whatever conditional mean function is specified.

R-squared measures and other measures of goodness-of-fit will generally involve sums of squared residuals. The simplest choice of residual is the *unweighted (or raw) residual*

$$r_i = (y_i - \hat{\mu}_i) . \quad (2.5)$$

This residual is heteroscedastic from (2.3), and a standardized residual may be preferred. The two standard choices are Pearson and deviance residuals, with associated measures of goodness of fit being Pearson's statistic and the deviance.

The *Pearson residual* is the obvious standardized residual

$$p_i = (y_i - \hat{\mu}_i) / \hat{\mu}_i^{1/2} . \quad (2.6)$$

Pearson's statistic is the sum of squared Pearson residuals

$$P = \sum_{i=1}^N (y_i - \hat{\mu}_i)^2 / \hat{\mu}_i . \quad (2.7)$$

The deviance is rarely used in econometrics but more widely used in the statistics literature. Let  $l(\mu)$  denote the log-likelihood function for a generalized linear model, defined in section 2.5, such as the Poisson, where  $\mu$  is the  $N \times 1$  vector with  $i$ -th entry  $\mu_i$ . Then the fitted log-likelihood is  $l(\hat{\mu})$ , while the maximum log-likelihood achievable, i.e. that in a full model with  $N$  parameters, is  $l(y)$ , where  $\hat{\mu}$  and  $y$  are  $N \times 1$  vectors with  $i$ -th entries  $\hat{\mu}_i$  and  $y_i$ . The deviance is defined to be

$$D(y, \hat{\mu}) = 2\{l(y) - l(\hat{\mu})\} \quad (2.8)$$

which is twice the difference between the maximum log-likelihood achievable and the log-likelihood of the fitted model. The squared deviance residual is the contribution of the  $i$ -th observation to the deviance.

For the Poisson log-density defined in (2.1) the *deviance residual* is

$$d_i = \text{sign}(y_i - \hat{\mu}_i) \cdot [2\{y_i \log(y_i / \hat{\mu}_i) - (y_i - \hat{\mu}_i)\}]^{1/2}, \quad (2.9)$$

where  $y \log(y) = 0$  for  $y = 0$ . The *deviance* is

$$D(y, \hat{\mu}) = \sum_{i=1}^N 2\{y_i \log(y_i / \hat{\mu}_i) - (y_i - \hat{\mu}_i)\}, \quad (2.10)$$

which usually simplifies due to (2.4).

## 2.2 R-Squared for Poisson Model based on Raw Residuals

We first consider using the usual R-squared for the linear regression model, i.e. measures based on unweighted residual sums of squares. The benchmark is the residual sum of squares in the intercept-only model, with fitted mean  $\bar{y}$ . There are several equivalent ways to express  $R^2$  in the linear regression model, but their analogs for nonlinear models differ.

Using the (unweighted) *residual sum of squares* yields

$$R_{RES}^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{\mu}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}. \quad (2.11)$$

$R_{RES}^2$  is clearly bounded from above by unity, but it may take negative values even if a constant term is included in the regression. Intuitively  $\sum_i (y_i - \hat{\mu}_i)^2 \leq \sum_i (y_i - \bar{y})^2$ , but this is not guaranteed in small samples as the Poisson MLE minimizes  $\sum_i (\hat{\mu}_i - y_i \log \hat{\mu}_i)$  rather than the sum of squared residuals. For similar reasons  $R_{RES}^2$  may decrease as additional regressors are added.

Using instead the (unweighted) *explained sum of squares* yields the measure

$$R_{EXP}^2 = \frac{\sum_{i=1}^N (\hat{\mu}_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2}. \quad (2.12)$$

This may exceed unity in small samples and also need not increase as regressors are added.

$R_{EXP}^2$  differs from  $R_{RES}^2$  since

$$\begin{aligned} \sum_{i=1}^N (y_i - \bar{y})^2 &= \sum_{i=1}^N (y_i - \hat{\mu}_i)^2 + \sum_{i=1}^N (\hat{\mu}_i - \bar{y})^2 \\ &\quad + 2 \sum_{i=1}^N (y_i - \hat{\mu}_i)(\hat{\mu}_i - \bar{y}). \end{aligned}$$

Unlike the case for the linear regression model the third term on the r.h.s. is not zero, and the two measures of  $R^2$  differ. For logit models where a similar difference occurs and has been well studied, Lave (1970) proposed use of the first measure. An additional complication in

defining  $R_{EXP}^2$  arises in Poisson models with  $\mu_i \neq \exp(X_i' \beta)$ . Then an alternative to (2.12) is to replace the sample mean of  $y_i$  in the numerator by the sample mean of the fitted values, as these two differ when (2.4) does not hold. Such a modified  $R_{EXP}^2$  still differs from  $R_{RES}^2$ , and in practice this modification makes relatively little difference to the value of  $R_{EXP}^2$ . It seems preferable to still use (2.12) which is motivated by decomposing  $(y_i - \bar{y})$  into the sum of the residual  $(y_i - \hat{\mu}_i)$  and the remainder  $(\hat{\mu}_i - \bar{y})$ .

A third related measure is the *squared sample correlation coefficient* between  $y_i$  and  $\hat{\mu}_i$

$$R_{COR}^2 = \frac{\left( \sum_{i=1}^N (y_i - \bar{y})(\hat{\mu}_i - \bar{\mu}) \right)^2}{\sum_{i=1}^N (y_i - \bar{y})^2 \cdot \sum_{i=1}^N (\hat{\mu}_i - \bar{\mu})^2}, \quad (2.13)$$

where  $\bar{\mu} = N^{-1} \sum_i \hat{\mu}_i$ . This measure differs from the first two, is clearly bounded between 0 and 1, and may decrease as regressors are added.

In summary, in small samples the three  $R^2$  measures based on raw residuals differ, and the only one of criteria 1-5 satisfied is criterion 1 by  $R_{COR}^2$ .

### 2.3 R-Squared for Poisson Model based on Pearson Residuals

Since the Poisson regression model is a heteroscedastic regression model, a more natural procedure is to use standardized rather than unweighted residuals. An obvious choice for the numerator of  $R^2$  is the Pearson residuals from the fitted model. More problematic is the choice of weight in the denominator. We propose  $\bar{y}$ , which is equivalent to using the Pearson residuals in the most restricted model where only an intercept is included. Then for the Poisson model

$$R_{P,P}^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{\mu}_i)^2 / \hat{\mu}_i}{\sum_{i=1}^N (y_i - \bar{y})^2 / \bar{y}}. \quad (2.14)$$

In small samples  $R_{P,P}^2$  is less than unity, but may be negative and may decrease as regressors are added.

We could use  $\hat{\mu}_i$  instead of  $\bar{y}$  as the weight for the denominator term of  $R_{P,P}^2$ . Superficially this seems similar to the measure of Buse (1973). Buse analyzed models with heteroscedastic errors in the context of GLS estimation with *known* variance and proposed  $R_{BUSE}^2 = 1 - [\sum_i (y_i - \hat{\mu}_i)^2 / \sigma_i^2] / [\sum_i (y_i - \bar{y}^*)^2 / \sigma_i^2]$ , where  $\bar{y}^*$  is the weighted average of  $y$  obtained by GLS in the model with just a constant term. To apply this to the Poisson model requires caution, however, because unlike the case considered by Buse  $\sigma_i^2$  depends on the same parameters as the conditional mean. Essentially it only makes sense to consider how much of the marginal variance of  $\frac{y}{\sigma}$  is explained by the conditional variance of  $\frac{y}{\sigma}$  given  $X$  if  $\sigma$  does not depend on  $X$ .

Another possible variation on  $R_{P,P}^2$  is a weighted version of  $R_{EXP}^2$  in (2.12). In applications the obvious quantity  $[\sum_i (\hat{\mu}_i - \bar{y})^2 / \hat{\mu}_i] / [\sum_i (y_i - \bar{y})^2 / \bar{y}]$  differs markedly from  $R_{P,P}^2$ . This is not surprising as theoretically we need to decompose  $(y_i - \bar{y}) / \bar{y}^{1/2}$  into the sum of the residual  $(y_i - \hat{\mu}_i) / \hat{\mu}_i^{1/2}$  and a remainder term that will be awkward, lack interpretation, and differ from  $(\hat{\mu}_i - \bar{y}) / \hat{\mu}_i^{1/2}$ .

R-squared measures based on Pearson residuals satisfy none of criteria 1 to 5.

## 2.4 R-Squared based on Deviance Residuals

We can construct a similar measure to  $R_{P,P}^2$ , using deviance residuals rather than Pearson residuals. The sum of squared deviance residuals for the fitted Poisson model, i.e. the deviance, is defined in (2.10). For Poisson with just an intercept the predicted mean is  $\bar{y}$ , and the deviance is  $D(y, \bar{y}) = \sum_{i=1}^N 2 y_i \log(y_i / \bar{y})$ . This yields the deviance R-squared for the

Poisson

$$R_{DEV,P}^2 = 1 - \frac{\sum_{i=1}^N \{y_i \log(y_i / \hat{\mu}_i) - (y_i - \hat{\mu}_i)\}}{\sum_{i=1}^N y_i \log(y_i / \bar{y})}, \quad (2.15)$$

which simplifies due to (2.4) when  $\mu_i = \exp(X_i' \beta)$ .

This deviance measure has a number of attractive properties. From (2.8) we have

$$\begin{aligned} R_{DEV,P}^2 &= 1 - 2\{l(y) - l(\hat{\mu})\} / 2\{l(y) - l(\bar{y})\} \\ &= 2\{l(\hat{\mu}) - l(\bar{y})\} / 2\{l(y) - l(\bar{y})\} \quad , \end{aligned} \quad (2.16)$$

Since the fitted log-likelihood increases as regressors are added and the maximum value is  $l(y)$  it follows that  $R_{DEV,P}^2$  lies between 0 and 1 and does not decrease as regressors are added.

From (2.16),  $R_{DEV,P}^2$  can be equivalently expressed as

$$R_{DEV,P}^2 = \frac{\sum_{i=1}^N \{y_i \log(\hat{\mu}_i / \bar{y}) - (\hat{\mu}_i - \bar{y})\}}{\sum_{i=1}^N y_i \log(y_i / \bar{y})} .$$

If we define for the Poisson model a generalized deviance function between any two estimates  $a$  and  $b$  of the vector mean  $\mu$  to be  $D(a, b) = \sum_{i=1}^N 2\{y_i \log(a_i / b_i) - (a_i - b_i)\}$ ,

then the numerator term in the expression above is the explained deviance  $D(\hat{\mu}, \bar{y})$ . Unlike R-squared measures based on unweighted residuals or Pearson residuals, therefore, that based on the deviance residuals has the advantage that the measure based on residual variation coincides with the measure based on explained variation.

Also from (2.16),  $R_{DEV,P}^2$  equals the log-likelihood ratio test statistic for overall fit of the model, divided by a scalar  $2\{l(y) - l(\bar{y})\}$  that depends only on the dependent variable  $y$  and not the regressors  $X$ .

Finally, from Hastie (1987) the deviance (2.8) equals twice the estimated Kullback-Leibler divergence between the  $N \times 1$  vectors  $\hat{\mu}$  and  $y$ . If we interpret the deviance  $D(y, \bar{y})$  in the intercept-only model as the information, measured by Kullback-Leibler divergence, in the sample data on  $y$  potentially recoverable by inclusion of regressors, then  $R_{DEV,P}^2$  measures the proportionate reduction in this potentially recoverable information.

Thus  $R_{DEV,P}^2$  satisfies all of criteria 1-5.

## 2.5 Deviance Residuals and Generalized Linear Models

In the simplest generalized linear model (glm) the density is a member of a natural exponential family, in which case the only parameter is the mean or a transformation of the mean. (By glm we mean a model with log-density (2.17). A broader definition allows for a multiplicative dispersion parameter.) The log-density is

$$l_i(\mu_i) = \{y_i b(\mu_i) - c(\mu_i)\} + d(y_i) \quad (2.17)$$

where  $E[y_i] = \mu_i$ ,  $\text{Var}(y_i) = b'(\mu_i)^{-1}$ , and we have parameterized the model in terms of the mean, rather than using the canonical parameterization of McCullagh and Nelder (1989, p.28). Different choices of the functions  $b(\cdot)$  and  $c(\cdot)$  correspond to different members, such as Poisson, normal ( $\sigma^2$  known), gamma and binomial (number of trials known).

It is for this class of models that the deviance (2.8) and associated residual are defined, while the Pearson residual generalizes to  $p_i = (y_i - \hat{\mu}_i) / b'(\hat{\mu}_i)^{-1/2}$ . Both deviance and Pearson residuals have been used and studied extensively for this class of models. The Pearson residuals often exhibit a skewed distribution when  $y$  is non-normally distributed, while deviance residuals are less skewed.

For the Poisson distribution, a residual that approximates normality is the Anscombe residual defined by  $a_i = 1.5 \cdot (y_i^{2/3} - \hat{\mu}_i^{2/3}) / \hat{\mu}_i^{1/6}$ . Although  $d_i$  and  $a_i$  have different functional forms, they take very similar values (see McCullagh and Nelder 1989, p. 39) and the deviance residuals are very nearly normally distributed after taking account of the discreteness by making a continuity correction (Pierce and Schafer 1986). This continuity correction consists of adding or subtracting  $\frac{1}{2}$  to/from  $y$ , toward the center of the distribution. Because of the approximate normality of the deviance residuals, they are very useful tools for identifying individual poorly fitting observations, or for investigating effects of potential new covariates or nonlinear effects of those already in the fitted model. Also, Davison and Gigli (1989) advocate using deviance residuals in normal scores plots to check the distributional assumptions.

For case-influence diagnostics, comparing the fit of a model including all  $N$  cases to the fit of a model based on  $N-1$  cases, after deletion of a single observation, Pregibon (1981) found the studentized Pearson and deviance residuals most useful. The studentized residuals are defined as  $p_i^* = p_i / \sqrt{1-h_i}$  and  $d_i^* = d_i / \sqrt{1-h_i}$ , where  $h_i$  is the  $i$ -th diagonal element of the “hat” matrix  $H = W^{1/2} X (X'WX)^{-1} X'W^{1/2}$ ,  $W = \text{diag}(\hat{w}_i)$ , and  $w_i = \text{Var}(y_i)^{-1} (\partial \mu_i / \partial (X_i' \beta))^2$ . For the detection of outliers, Williams (1987) introduced the likelihood residual. The squared likelihood residual is a weighted average of the squared studentized deviance and Pearson residuals,  $l_i^2 = (1-h_i)d_i^{*2} + h_i p_i^{*2}$ , and is approximately equal to the likelihood ratio test for testing whether an observation is an outlier. Because the average value,  $k/N$ , of  $h_i$  is small, the likelihood residual  $l_i$  is much closer to  $d_i$  than to  $p_i$ , and is therefore also approximately normally distributed. Williams (1987) uses half-normal plots to analyze the results.

Pearson’s statistic and the deviance are used as overall goodness-of-fit tests. A large value for Pearson’s statistic or the deviance is usually interpreted as indicating poor “goodness of fit” of the Poisson model, in the sense that given the assumed conditional variance, the fitted conditional mean is on average far from the actual value of the dependent variable. However, it may instead indicate inappropriateness of the Poisson restriction that the conditional variance equals the conditional mean. Both measures are asymptotically chi-squared distributed if data is grouped. For the more usual case of non-grouped data, which we analyze, the Pearson statistic is often found to be more nearly chi-squared distributed than the deviance (Pierce and Schafer 1986), but these authors still argue that the deviance is a more appropriate measure than the Pearson statistic. Landwehr et al. (1984) also prefer the deviance as a goodness-of fit measure in the logit model, explicitly stating the fact that it will decrease when regressors are added, which is not necessarily the case for Pearson’s statistic. The analysis of deviance, McCullagh and Nelder (1989, p.35), is a generalization of analysis of variance in normal error models that uses the incremental change in deviance as regressors are added. This is used in the log-linear model literature, e.g. Bishop, Feinberg and Holland (1975), which uses the term  $G^2$  statistic for the deviance. McCullagh and Nelder (1989)

merely see the Pearson statistic as a measure of residual variation, and use  $P/(N-k)$  as an estimator of the dispersion parameter  $\rho$ , when the variance is specified as  $\rho\mu_i$ . McCullagh (1986) derives the asymptotic distribution of both Pearson's statistic and the deviance, conditional on the estimate of  $\beta$ . Both statistics are conditionally asymptotically normally distributed. The moments are not easily evaluated, though, and care has to be taken when the  $\mu_i$  are small, as the deviance may become degenerate.

## 2.6 Discussion

For the Poisson model, the deviance  $R^2$  satisfies all five criteria presented in the introduction, while the other proposed  $R^2$  fail all these criteria, except  $R_{COR}^2$  which always lies between 0 and 1. The deviance  $R^2$  is based on a residual that while unfamiliar to econometricians has desirable properties detailed in section 2.5. It can be computed from the output of any Poisson regression package that reports the fitted log-likelihood or a likelihood ratio test, see (2.16), given separate once-only computation of  $l(y)$ . It is even more easily computed as  $1 - G^2(\hat{\mu})/G^2(\bar{y})$ , where  $G^2$  is the  $G^2$  statistic reported by the econometrics package LIMDEP.

The only study we have found that considers or reports an  $R^2$  for Poisson models is Merkle and Zimmermann (1992, p.291). They present a measure equal to  $R_{DEV,P}^2$  which is motivated as the explained increase in the log-likelihood relative to the maximum increase possible, and  $R_{P,P}^2$ . They give little discussion of properties of these, and do not attempt to discriminate between them.

The only statistical package we have found that reports an  $R^2$  is STATA, which uses the likelihood ratio index,

$$R_{LRI}^2 = 1 - \frac{l(\hat{\mu})}{l(\bar{y})}$$

which compares the log-likelihoods of the fitted and intercept-only models.  $R_{LRI}^2$  is a scalar multiple of  $R_{DEV}^2$ , since

$$R_{LRI}^2 = \{1 - l(y)/l(\bar{y})\} \cdot R_{DEV}^2$$

using (2.16). For the Poisson these differ and  $R_{LRI}^2$  has the disadvantage of having a theoretical upper limit less than unity.

All of the measures defined above generalize in a straightforward way to other generalized linear models. In particular, the general formula for  $R_{DEV}^2$  given in (2.16) is evaluated at densities aside from Poisson that are also special cases of (2.17). For the normal ( $\sigma^2$  known) the deviance residual is the usual residual and the deviance  $R^2$  coincides with the usual  $R^2$ , and for the Bernoulli, e.g. logit and probit, the deviance  $R^2$  coincides with the  $R^2$  of McFadden (1974) which equals  $R_{LRI}^2$ . Generalization of the deviance  $R^2$  to generalized linear models is studied in detail by Cameron and Windmeijer (1994).

Our concern here is with count data models. The Poisson model imposes the restriction that the conditional variance equals the conditional mean, but in practice this restriction is often unreasonable. More general count data models relax this restriction by introducing a dispersion parameter which complicates computation of both deviance  $R^2$  and Pearson  $R^2$ . Furthermore, the deviance is a construct for generalized linear models, yet not all count data models fall in this class. The next section nonetheless proposes deviance  $R^2$  and Pearson  $R^2$  for the commonly-used non-Poisson count data models.

### 3. R-SQUARED FOR OVERDISPERSED COUNT DATA MODELS

#### 3.1 General Results

The Poisson model imposes the restriction that the conditional variance equals the conditional mean, but in practice this restriction is often unreasonable. More general count data models relax this restriction. Basic models for count data are presented in Cameron and Trivedi (1986), while Gurmu and Trivedi (1992) provide a more recent comprehensive survey.

The variance-mean relationship is usually specified to be either that of the Negbin 1 variance function

$$\text{Var}(y_i|X_i) = \mu_i + \alpha \cdot \mu_i \quad (3.1)$$

or the Negbin 2 variance function

$$\text{Var}(y_i|X_i) = \mu_i + \alpha \cdot \mu_i^2 \quad (3.2)$$

where we use the terminology of Cameron and Trivedi (1986) with 1 and 2 denoting the highest power to which the mean is raised in the variance function. The scalar  $\alpha \geq 0$  is called the dispersion parameter.

Two possible modeling approaches may be taken. The quasi-likelihood (QL) approach assumes mean function (2.2) and variance function  $\text{Var}(y_i|X_i) = v(\mu_i, \alpha)$  for specified function  $v$  as in (3.1) or (3.2), and estimates the parameters  $\beta$  and  $\alpha$  on the basis of the assumed first two moments. The QL score equations for  $\beta$  have the form

$$\sum_{i=1}^N \frac{(y_i - \mu_i)}{v(\mu_i, \alpha)} \frac{\partial \mu_i}{\partial \beta} = 0 \quad (3.3)$$

for a given value of  $\alpha$ .

The *maximum likelihood (ML)* approach additionally specifies a density, usually the negative binomial, and estimates the parameters  $\beta$  and  $\alpha$  by maximum likelihood methods. In econometrics, the MLE for negative binomial with Negbin 2 variance function is commonly used, and is the estimator provided, for example, by the LIMDEP and STATA packages. We consider both QL and negative binomial ML estimators for the Negbin 1 and Negbin 2 variance functions.

For all these models the definition of the *unweighted residual* is that in (2.5). The *unweighted residual R-squared* measures  $R_{RES}^2$ ,  $R_{EXP}^2$  and  $R_{COR}^2$  are as defined in the Poisson case, except that they are evaluated at the mean fitted by the relevant estimation procedure. These measures have the same properties, and deficiencies, as in the Poisson case. In the special case where  $\mu_i = \exp(X_i' \beta)$  and  $X_i$  includes a constant, the raw residuals can be shown to sum to zero, i.e. (2.4) holds, for the Negbin 1 model (both QL and ML

estimates), but not for the Negbin 2 model. The latter result can lead to quite different performance of  $R_{RES}^2$  and  $R_{EXP}^2$  in Negbin 2 applications.

For all these models, the *Pearson residual* generalizes to

$$p_i = (y_i - \hat{\mu}_i) / v(\hat{\alpha}, \hat{\mu}_i)^{1/2} \quad (3.4)$$

where the variance function in the denominator is defined in (3.1) or (3.2). To construct an  $R_P^2$  measure for count data models more general than the Poisson we need to obtain a dispersion parameter estimate  $\hat{\alpha}$  for both the fitted and intercept-only models. We use the same estimate of the variance parameter for both the fitted and intercept-only models, as we wish to measure the fit due to the regressors, not the dispersion parameter. On this point, it is instructive to consider the homoscedastic error classical regression model. If the variance parameter  $\sigma^2$  is known, the Pearson residual equals  $(y_i - \hat{\mu}_i) / \sigma$  in the fitted model and  $(y_i - \bar{y}) / \sigma$  in the intercept-only model. When  $\sigma$  is unknown we use the same estimate of  $\sigma$  for both the fitted model and intercept-only model, cancellation occurs, and  $R_P^2$  reduces to the usual  $R^2$  for the classical regression model. If instead we used  $\hat{\sigma}^2 = (N - k)^{-1} \sum_i (y_i - \hat{\mu}_i)^2$  for the fitted model and  $\hat{\sigma}_0^2 = (N - 1)^{-1} \sum_i (y_i - \bar{y})^2$  for the intercept-only model, then  $R_P^2$  would always equal  $(k - 1) / (N - 1)$ . The same would occur for the QL estimator with Negbin 1 variance function, defined in section 3.3.

We therefore define for general count data models the *Pearson R-squared*

$$R_P^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{\mu}_i)^2 / v(\hat{\mu}_i, \hat{\alpha})}{\sum_{i=1}^N (y_i - \hat{\mu}_0)^2 / v(\hat{\mu}_0, \hat{\alpha})} . \quad (3.5)$$

where  $\hat{\mu}_i$  and  $\hat{\alpha}$  are evaluated at the estimates  $(\hat{\alpha}, \hat{\beta})$  of the fitted model, and  $\hat{\mu}_0 = \hat{\mu}_0(\hat{\alpha})$  denotes the predicted mean in the intercept-only model estimated under the constraint that  $\alpha = \hat{\alpha}$ . For the Poisson model  $R_P^2$  simplifies to  $R_{P,P}^2$  as  $v(\hat{\mu}_i, \hat{\alpha}) = \hat{\mu}_i$  and  $\hat{\mu}_0 = \bar{y}$ . Computation of  $R_P^2$  for other models is deferred to the relevant sub-sections below. In all cases this measure satisfies none of criteria 1-5.

The deviance is traditionally defined for generalized linear models, a class that excludes the Negbin 2 model, unless  $\alpha$  is known, and the Negbin 1 model even if  $\alpha$  is known. We propose the following generalization of the *deviance* to cover these models when estimated by ML

$$D(\hat{\mu}_s, \hat{\mu}, \hat{\alpha}) = \sum_{i=1}^N 2 \{l_i(\hat{\mu}_{si}, \hat{\alpha}) - l_i(\hat{\mu}_i, \hat{\alpha})\}, \quad (3.6)$$

where  $l_i(\hat{\mu}_i, \hat{\alpha})$  is the log-density evaluated at the ML estimator of the fitted model, and  $\hat{\mu}_{si} = \hat{\mu}_{si}(\hat{\alpha})$  is the ML estimate of the fitted mean in the saturated model, i.e. the model with separate intercept for each observation, estimated with the constraint  $\alpha = \hat{\alpha}$ . The deviance is twice the difference between the maximum log-likelihood (given  $\alpha = \hat{\alpha}$ ) and the log-likelihood for the fitted model, and is clearly nonnegative. (We call this generalization the deviance. The discussion in McCullagh and Nelder (1989, p. 33) suggests it could also be called the discrepancy, or in some applications the scaled deviance. The working paper version of this paper had a slightly different definition which, as pointed out by a referee, could potentially be negative.) To generate an R-squared measure from (3.6) we evaluate the deviance for the benchmark intercept-only model at the same estimated value of  $\alpha$  as for the fitted model, taking the same approach as for the Pearson R-squared.

The *Deviance R-squared* is

$$\begin{aligned} R_{DEV}^2 &= 1 - D(\hat{\mu}_s, \hat{\mu}, \hat{\alpha}) / D(\hat{\mu}_s, \hat{\mu}_0, \hat{\alpha}) \\ &= 1 - \frac{\sum_{i=1}^N \{l_i(\hat{\mu}_{si}, \hat{\alpha}) - l_i(\hat{\mu}_i, \hat{\alpha})\}}{\sum_{i=1}^N \{l_i(\hat{\mu}_{si}, \hat{\alpha}) - l_i(\hat{\mu}_0, \hat{\alpha})\}} \end{aligned} \quad (3.7)$$

where  $\hat{\mu}_i$ ,  $\hat{\alpha}$  and  $\hat{\mu}_0$  are defined after (3.5). This nests  $R_{DEV,P}^2$  for the Poisson model, as then  $\hat{\alpha} = 0$ ,  $\hat{\mu}_{si} = y_i$ , and  $\hat{\mu}_0 = \bar{y}$ . For the negative binomial density with Negbin 1 or 2 variance function,  $R_{DEV}^2$  approaches  $R_{DEV,P}^2$  as  $\hat{\alpha} \rightarrow 0$ , since the negative binomial density then simplifies to the Poisson density.  $R_{DEV}^2$  always satisfies criteria 1 and 4, directly from (3.7), and criterion 3 by re-expressing (3.7) as  $R_{DEV}^2 = D(\hat{\mu}, \hat{\mu}_0, \hat{\alpha}) / D(\hat{\mu}_s, \hat{\mu}_0, \hat{\alpha})$ . Criterion 2 is generally not satisfied, as changes in  $\hat{\alpha}$  as regressors are added leads to changes in all terms

in (3.7). Criterion 5 is generally not satisfied, but will be if the model is a generalized linear model.

QL estimation does not require specification of a density. It can be shown, however, that for the Negbin 1 or 2 variance functions there is a specific likelihood function for which the ML estimator of  $\beta$  (for given  $\alpha$ ) equals the QL estimator. The deviance concept can therefore be applied to QL estimation. Details are given in the following sub-sections where we consider computation of  $R_P^2$  and  $R_{DEV}^2$  for models with Negbin 2 and Negbin 1 variance functions estimated by both ML and QL methods.

### 3.2 R-squared for Negbin 2

The estimator for overdispersed count data most commonly used by econometricians is the MLE for the negative binomial distribution with mean (2.2) and Negbin 2 variance function (3.2) quadratic in the mean. This estimator is provided, for example, by LIMDEP and STATA. The log density is

$$l(\mu_i, \alpha) = \log(\Gamma(y_i + \alpha^{-1}) / \Gamma(\alpha^{-1})) - (y_i + \alpha^{-1}) \log(1 + \alpha\mu_i) + y_i \log(\alpha\mu_i) - \log y_i!. \quad (3.8)$$

The ML estimator jointly maximizes the log-likelihood function w.r.t.  $\beta$  and  $\alpha$ . The QL estimator of  $\beta$  solves (3.3) with  $\alpha$  replaced by a consistent estimator, e.g.  $\hat{\alpha} = \sum_i \hat{\mu}_i^2 \{(y_i - \hat{\mu}_i)^2 - \hat{\mu}_i\} / \sum_i \hat{\mu}_i^4$ , where  $\hat{\mu}_i$  is the Poisson estimated mean. This can be shown to be equivalent to maximizing (3.8) w.r.t.  $\beta$  with  $\alpha = \hat{\alpha}$  (see Cameron and Trivedi 1986, pp.45-46).

The Pearson R-squared is

$$R_{P,NB2}^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{\mu}_i)^2 / (\hat{\mu}_i + \hat{\alpha} \hat{\mu}_i^2)}{\sum_{i=1}^N (y_i - \bar{y})^2 / (\bar{y} + \hat{\alpha} \bar{y}^2)}, \quad (3.9)$$

as it can be shown that  $\hat{\mu}_0 = \bar{y}$  for both the ML and QL estimators.

The deviance for ML estimation is

$$\begin{aligned}
D_{NB2}(\hat{\mu}_y, \hat{\mu}, \hat{\alpha}) \\
= \sum_{i=1}^N 2 \left\{ y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i + \hat{\alpha}^{-1}) \log\left(\frac{(y_i + \hat{\alpha}^{-1})}{(\hat{\mu}_i + \hat{\alpha}^{-1})}\right) \right\}
\end{aligned} \tag{3.10}$$

using  $\hat{\mu}_{si} = y_i$  for the log-density (3.8). Thus

$$\begin{aligned}
R_{DEV, NB2(ML)}^2 \\
= 1 - \frac{\sum_{i=1}^N \left\{ y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i + \hat{\alpha}^{-1}) \log\left(\frac{(y_i + \hat{\alpha}^{-1})}{(\hat{\mu}_i + \hat{\alpha}^{-1})}\right) \right\}}{\sum_{i=1}^N \left\{ y_i \log\left(\frac{y_i}{\bar{y}}\right) - (y_i + \hat{\alpha}^{-1}) \log\left(\frac{(y_i + \hat{\alpha}^{-1})}{(\bar{y} + \hat{\alpha}^{-1})}\right) \right\}}
\end{aligned} \tag{3.11}$$

Criteria 1, 3 and 4 are always satisfied as noted in section 3.1. Criterion 2 is not, but we find in examples that  $R_{DEV, NB2(ML)}^2$  increases as regressors are added. Criterion 5 is satisfied if  $\alpha$  is known, in which case  $\hat{\alpha}$  above becomes  $\alpha$ , as then the model is a generalized linear model, i.e. (3.8) is a special case of (2.17), so the deviance equals twice the estimated Kullback-Leibler divergence. A leading example of known  $\alpha$  is the geometric count model, which sets  $\alpha = 1$ .

If estimation is by QL, the same formula (3.11) is used for  $R_{DEV, NB2(QL)}^2$ , but the MLE for  $\alpha$  is replaced by a consistent estimate  $\hat{\alpha}$ , and  $\hat{\mu}_i$  is replaced by the QL estimate of  $\mu_i$ .

### 3.3 R-squared for Negbin 1

The log density for the negative binomial with mean (2.2) and Negbin 1 variance function (3.1) linear in the mean is

$$\begin{aligned}
l(\mu_i, \alpha) &= \log(\Gamma(y_i + \alpha^{-1}\mu_i) / \Gamma(\alpha^{-1}\mu_i)) \\
&\quad - (y_i + \alpha^{-1}\mu_i) \log(1 + \alpha) + y_i \log \alpha - \log y_i!
\end{aligned} \tag{3.12}$$

The ML estimator jointly maximizes the log-likelihood function w.r.t.  $\beta$  and  $\alpha$ . The QL estimator which solves (3.3) for given  $\alpha$ , can be shown to maximize w.r.t.  $\beta$  the log-likelihood based on the scaled Poisson log density

$$l(\mu_i, \alpha) = (y_i \log \mu_i - \mu_i) / (1 + \alpha) + c(y_i, \alpha), \quad (3.13)$$

where  $c(y_i, \alpha)$  is a normalizing constant. Clearly, no matter what value  $\alpha$  takes, the QL estimator for  $\beta$  is equal to the Poisson ML estimator.

The Pearson R-squared for the ML estimator is given by

$$R_{P, NB1(ML)}^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{\mu}_i)^2 / \hat{\mu}_i}{\sum_{i=1}^N (y_i - \hat{\mu}_0)^2 / \hat{\mu}_0}. \quad (3.14)$$

Further simplification occurs for the QL estimator as then  $\hat{\mu}_0 = \bar{y}$ , so  $R_{P, NB1(QL)}^2 = R_{P, P}^2$ .

The deviance for the ML estimator which maximizes the likelihood based on (3.12) is

$$D_{NB1}(\hat{\mu}_s, \hat{\mu}, \hat{\alpha}) = \sum_{i=1}^N 2 \left\{ \log \left( \frac{\Gamma(y_i + \hat{\alpha}^{-1} \hat{\mu}_{si}) \cdot \Gamma(\hat{\alpha}^{-1} \hat{\mu}_i)}{\Gamma(y_i + \hat{\alpha}^{-1} \hat{\mu}_i) \cdot \Gamma(\hat{\alpha}^{-1} \hat{\mu}_{si})} \right) - (\hat{\alpha}^{-1} \hat{\mu}_{si} - \hat{\alpha}^{-1} \hat{\mu}_i) \cdot \log(1 + \hat{\alpha}) \right\} \quad (3.15)$$

where the saturated model estimator  $\hat{\mu}_{si}$  equals 0 for  $y_i = 0$ , and for  $y_i > 0$  is the solution to the nonlinear system of equations  $\log(1 + \hat{\alpha}) = \sum_{j=1}^{y_i} (y_i - j + \hat{\alpha}^{-1} \hat{\mu}_{si})^{-1}$ . The solution is very well approximated by  $\hat{\mu}_{si} = y_i + \hat{\alpha} / 2$ . (A Gauss program to compute all the R-squareds in this paper is available from the JBES ftp-site, and from the authors upon request.) The deviance R-squared is

$$R_{DEV, NB1(ML)}^2 = 1 - \sum_{i=1}^N \left( \log \left( \frac{\Gamma(y_i + \hat{\alpha}^{-1} \hat{\mu}_{si}) \cdot \Gamma(\hat{\alpha}^{-1} \hat{\mu}_i)}{\Gamma(y_i + \hat{\alpha}^{-1} \hat{\mu}_i) \cdot \Gamma(\hat{\alpha}^{-1} \hat{\mu}_{si})} \right) - (\hat{\alpha}^{-1} \hat{\mu}_{si} - \hat{\alpha}^{-1} \hat{\mu}_i) \cdot \log(1 + \hat{\alpha}) \right)$$

$$/ \sum_{i=1}^N \left\{ \log \left( \frac{\Gamma(y_i + \hat{\alpha}^{-1} \hat{\mu}_{si}) \cdot \Gamma(\hat{\alpha}^{-1} \hat{\mu}_0)}{\Gamma(y_i + \hat{\alpha}^{-1} \hat{\mu}_0) \cdot \Gamma(\hat{\alpha}^{-1} \hat{\mu}_{si})} \right) - (\hat{\alpha}^{-1} \hat{\mu}_{si} - \hat{\alpha}^{-1} \hat{\mu}_0) \cdot \log(1 + \hat{\alpha}) \right\} \quad (3.16)$$

Criteria 1, 3 and 4 are always satisfied. Criterion 2 is not, but we find in examples that  $R_{DEV, NB1(ML)}^2$  increases as regressors are added. Criterion 5 is not satisfied, as even if  $\alpha$  is known the Negbin 1 model is not a generalized linear model.

For the Negbin 1 variance function the ML estimator is rarely used. The QL estimator, however, is the generalization of the Poisson most commonly used by applied statisticians. As indicated above, this estimator of  $\beta$  is exactly the same as the Poisson MLE, while the standard errors of the QL estimator of  $\beta$  are the usual reported Poisson MLE standard errors multiplied by the scalar  $(1 + \hat{\alpha})^{1/2}$ , where  $1 + \hat{\alpha} = (N - k)^{-1} \sum_i (y_i - \hat{\mu}_i)^2 / \hat{\mu}_i$  and  $\hat{\mu}_i$  is the Poisson fitted mean. The QL estimator could be viewed as providing a consistent but inefficient estimator in the model with log-density (3.12). It is better to view it as maximizing w.r.t.  $\beta$  the log-likelihood based on (3.13), for which  $R_{DEV}^2$  reduces to the Poisson  $R_{DEV}^2$ , due to cancellation of the common factor  $1 + \hat{\alpha}$ . We use this version of  $R_{DEV}^2$  for the QL estimator, i.e.  $R_{DEV, NB1(QL)}^2 = R_{DEV, P}^2$ . It satisfies criteria 1-4.

### 3.4 Discussion

If a constant term is excluded from the full model, it no longer nests the reference intercept-only model and all  $R^2$ , with the exception of  $R_{EXP}^2$  and  $R_{COR}^2$ , may be negative. This is similar to the linear regression model without intercept.

The goal has been to present  $R^2$  measures that do not decrease as regressors are added. For the linear regression model, a degrees-of-freedom adjusted  $R^2$  has been developed to provide a penalty as the number of regressors increase. Similar measures can be developed here. The obvious correction to measures such as  $R_{RES}^2$ ,  $R_P^2$ , or  $R_{DEV}^2$  is to divide the numerator by  $(N - k)$  and the denominator by  $(N - 1)$ . Another correction is to instead use

small-sample adjusted residuals defined in Pierce and Schafer (1986), for example. Our concern, however, is with the more fundamental issue of which residual should be used in the first place.

The estimators discussed specify at least the first two moments of  $y_i$ . A consistent but inefficient estimator using only the first moment is the nonlinear least squares (NLLS) estimator with nonlinear mean function (2.2). Heteroscedasticity consistent standard errors can be computed, but  $R_{RES}^2$  is inadequate due to heteroscedasticity, while the  $R_P^2$  or  $R_{DEV}^2$  measures defined above will satisfy none of criteria 1 to 5 when evaluated at the NLLS estimate.

The proposed  $R^2$  measures are not intended to discriminate between Poisson and negative binomial (types 1 and 2) models. A variation of (3.7) that does this for ML estimation is

$$R_{DP}^2 = 1 - \frac{\sum_{i=1}^N \{l_{pi}(y_i) - l_i(\hat{\mu}_i, \hat{\alpha})\}}{\sum_{i=1}^N \{l_{pi}(y_i) - l_{pi}(\bar{y})\}} \quad (3.17)$$

where  $l_{pi}$  denotes the Poisson density and  $l_i(\hat{\mu}_i, \hat{\alpha})$  denotes the log-density for the fitted Poisson or negative binomial model. The motivation for  $R_{DP}^2$  is that  $l_{pi}(y_i)$  is the log-density in the saturated model not only for Poisson, but also negative binomial if maximization is over both  $\mu_i$  and  $\alpha$ , and  $l_{pi}(\bar{y})$  is the obvious benchmark intercept-only model for all densities. For Poisson  $R_{DP}^2 = R_{DEV,P}^2$  and hence satisfies criteria 1-5. For negative binomial models  $R_{DP}^2$  satisfies criteria 1-4. A high value  $R_{DP}^2$  will arise if the regressors have high explanatory power, or if the data are considerably overdispersed so that even without regressors the negative binomial models fits the data much better than the Poisson. Comparisons of the ML Negbin  $R_{DP}^2$ 's with  $R_{DEV,P}^2$  are implicitly tests of  $\alpha = 0$ .

#### 4. EMPIRICAL PERFORMANCE OF MEASURES

Three estimators are considered: (1) Poisson MLE or equivalently Negbin 1 QLE, (2) Negbin 1 MLE, and (3) Negbin 2 MLE. These estimators are abbreviated to POISSON, NB1 and NB2. We could alternatively call the first estimator Negbin 1 QLE since from section 3.3 the estimators of  $\beta$  and the associated R-squared measures are the same for Poisson MLE and Negbin 1 QLE. The only difference is in the reported standard errors. Thus we consider all the common count data estimators plus the less used Negbin 1 MLE. In all models the conditional mean is specified as  $\mu = \exp(X' \beta)$ .

We compare six R-squared measures. Three are based on unweighted residuals ( $R_{RES}^2$ ,  $R_{EXP}^2$ , and  $R_{COR}^2$ ), one is based on Pearson residuals ( $R_P^2$ ), and two are based on deviance residuals ( $R_{DEV}^2$  and  $R_{DP}^2$ ) with  $R_{DEV}^2 = R_{DP}^2$  for Poisson. Formulae for these measures are collected in Table 1.

**[TABLE 1 ABOUT HERE]**

We use the same models and sample of 5,190 individuals drawn from the 1977-78 Australian Health Survey, Australian Bureau of Statistics (1978) and Social Science Data Archives (1983), as used by Cameron and Trivedi (1986). No goodness of fit measure, aside from the fitted log-likelihood, is reported in their study. We consider the extent to which utilization of health care services varies with socioeconomic characteristics (SEX, AGE, AGESQ, INCOME), health insurance type (the three included mutually exclusive dummies for type of health insurance are LEVYPLUS, FREEPOOR and FREEREPA with LEVY the omitted category), recent health (ILLNESS, ACTDAYS), and long-term health status (HSCORE, CHCOND1 and CHCOND2). A summary description of the independent variables is given in Table 2.

**[TABLE 2 ABOUT HERE]**

Three dependent variables are modeled, *viz.* DOCTORCO, the number of doctor consultations in the past two weeks; HOSPADMINI, the number of hospital admissions in the past 12 months; and MEDICINES, the number of medicines taken in the past two days. The values of the R-squared measures for the models with all 12 regressors are given in Table 3,

together with bootstrap standard error estimates. The bootstrap resamples  $(y_i, X_i)$  jointly, with 200 bootstrap replications performed following the suggestion for standard error estimation of Efron and Tibshirani (1993, p. 52).

**[TABLE 3 ABOUT HERE]**

The first dependent variable, the number of doctor consultations ( DOCTORCO) in the past two weeks, is a discrete variable taking the values 0, 1, 2, ..., 9 with probabilities of respectively 0.798, 0.151, 0.034, 0.006, 0.005, 0.002, 0.002, 0.002, 0.001, 0.000. Estimates of the various count data models are given in Cameron and Trivedi (1986, p.47) and are not reproduced here. The data are overdispersed, with the ML estimate of the variance parameter  $\alpha$  being 0.456 in the NB1 model and 1.077 in the NB2 model. The log-likelihood statistics are respectively  $-3355.5$ ,  $-3226.6$  and  $-3198.7$  in the Poisson, NB1 and NB2 models, so that the parameter  $\alpha$  is estimated to be highly significantly different from zero in both the NB1 and NB2, with the NB2 model providing the best fit for the data.

Looking across different count data models for DOCTORCO, the preferred measure  $R_{DEV}^2$  for the full model takes respective values of 0.223 for Poisson, 0.171 for NB1 and 0.229 for NB2.  $R_p^2$  takes respective values of 0.373, 0.413 and 0.373.  $R_{COR}^2$  changes little across the models, due to the fitted values being highly correlated across models, taking values 0.164, 0.162 and 0.150. The other two measures,  $R_{EXP}^2$  and especially  $R_{RES}^2$ , change appreciably across the different models, and for this reason may be inferior. In particular for the NB2 model estimated by MLE  $R_{EXP}^2 = 0.502$ ,  $R_{COR}^2 = 0.150$  and  $R_{RES}^2 = 0.051$ . (This large difference for NB2 arises because, in contrast to the Poisson and NB1 model, the NB2 model raw residuals do not sum to zero for  $\mu = \exp(X'\beta)$ ). This leads to residuals that are considerably larger than in the Poisson and NB1 case, especially for some outlying observations, in this case people who did not visit the doctor but had high values for ILLNESS and ACTDAYS. At the same time, the Poisson and NB2 fitted values are still very highly correlated, which leads to little difference in  $R_{COR}^2$  across the two models. For the full model a regression of the NB2 fitted values on the Poisson fitted values has a slope coefficient of 1.40, considerably different from unity, while the squared correlation

coefficient between NB2 and Poisson fitted values is 0.97. Corresponding figures in comparing NB1 and Poisson are 0.92 for the slope and 0.98 for the squared correlation coefficient.) The final measure  $R_{DP}^2$ , which can be meaningfully compared across models, increases from 0.223 for POISSON to 0.268 for NB1 and 0.278 for NB2.

The second dependent variable, the number of hospital admissions (HOSPADMI) in the past 12 months, takes the values 0, 1, 2, ..., 5 with probabilities of respectively 0.865, 0.108, 0.018, 0.005, 0.001, and 0.002. The data are moderately overdispersed, with the (unconditional) sample variance 1.48 times the sample mean. HOSPADMI is a priori difficult to predict. This difficulty in prediction is illustrated in Table 3, where the largest value of  $R^2$  is 0.181 for  $R_p^2$  in NB1.

The third dependent variable is the number of medicines (MEDICINES) taken in the past two days, which takes the values 0, 1, 2, ..., 8 with probabilities of respectively 0.429, 0.268, 0.139, 0.076, 0.042, 0.020, 0.012, 0.006 and 0.008. This is more easily predicted, with higher  $R^2$  than for the other two utilization measures.

The deviance  $R^2$  takes similar values for the Poisson and NB2 models for all three dependent variables, but for the NB1 model the value is always lower than the other two. For the Pearson  $R^2$  also the values of the Poisson and NB2 models are quite similar, and for this measure the NB1 results are higher. The raw residuals based  $R^2$  measures are similar for all three models for HOSPADMI and MEDICINES, but not for DOCTORCO. For the last variable, the differences between the various  $R^2$  measures are largest, with for example  $R_p^2$  being more than twice the value of  $R_{RES}^2$  in both the Poisson and NB1 model. Also, the difference between  $R_{RES}^2$  and  $R_{EXP}^2$  is largest for DOCTORCO. For all applications, except NB2 model for DOCTORCO,  $R_{RES}^2$  is very close to  $R_{COR}^2$ .

In all cases the variability of the deviance  $R^2$  is less than that of the other measures, even if we standardize for different scale of different measures by comparing the ratio of the bootstrap estimated standard error to the calculated  $R^2$ . The variability of the Pearson  $R^2$  is largest.

**[FIGURE 1 ABOUT HERE]**

Figure 1 gives nonparametric estimates of the bootstrap densities of the various  $R^2$  for the model with all regressors included. The bandwidth for the various densities is 0.008, except for the more variable  $R_p^2$  for which it is 0.012, and there are 1,000 replications.  $R_{DEV}^2$  is symmetrically distributed, and more concentrated around its mean than the other measures.  $R_p^2$  is asymmetrically (skewed to the left) distributed, whereas  $R_{RES}^2$  and  $R_{COR}^2$  are skewed to the right. The density of  $R_{EXP}^2$  is symmetric. Plots corresponding to Figure 1 for the other estimators and models show symmetry for  $R_{DEV}^2$  throughout, asymmetries for other  $R^2$  measures for NB1 and NB2 estimates of DOCTORCO similar to that for POISSON, and relatively little asymmetries for all measures and all models for HOSPADMI and MEDICINES.

**[FIGURE 2 ABOUT HERE]**

Figure 2 plots the  $R^2$  measures for DOCTORCO as regressors are progressively added. Recall that while Figure 2 is labeled as the Poisson model it could just as well be labeled as the NB1 variance function model estimated by the QL method. The plot illustrates that  $R_p^2$  can indeed be negative and decrease as regressors are added. All R-squared measures tell a similar story as regressors are added. Health status variables, especially recent health, appear much more important than socioeconomic characteristics and health insurance status in determining the number of doctor consultations in the past 2 weeks. This interpretation is dependent on the order with which regressors are added. If the order of the regressors is reversed, the  $R^2$  measures range from 0.05 to 0.09 when just long-term health status variables are included, and range from 0.15 to 0.36 if the short-term health status variables are additionally included. These latter variables, either ILLNESS or ACTDAYS alone, have the greatest contribution to  $R^2$ .

Some studies calculate a fitted frequency distribution as the average over the predicted probabilities for each count, denoted  $\tilde{p}_j$ ,  $j = 0, 1, 2, \dots$ , tabulate this along with the observed frequencies  $\bar{p}_j$ , and use correspondence between the two as a measure of model adequacy.

This suggests an  $R^2$  equal to  $1 - \frac{\sum_{k=0}^{\max(y)} (\bar{p}_k - \tilde{p}_k)^2}{\sum_{k=0}^{\max(y)} (\bar{p}_k - \tilde{p}_{k0})^2}$  where  $\tilde{p}_{k0}$  is the predicted probability from the intercept-only model. For counts 0, 1, 2, ..., 9 of DOCTORCO the observed probabilities of respectively 0.798, 0.151, 0.034, 0.006, 0.005, 0.002, 0.002, 0.002, 0.001, 0.000 are seemingly closely fitted by the Poisson MLE average predicted probabilities of 0.773, 0.178, 0.032, 0.009, 0.004, 0.002, 0.001, 0.000, 0.000, 0.000. But the intercept-only Poisson predicted probabilities of 0.740, 0.223, 0.034, 0.003, 0.000, ... , 0.000 are also reasonable, leading to an  $R^2$  of 0.84. Aggregation leads to an  $R^2$  much higher than the other measures.

## 5. CONCLUSIONS

The preferred R-squared measure for standard count data models is that based on the deviance residual,  $R_{DEV}^2$ . For the Poisson ML estimator and the QL estimator with Negbin 1 variance function this satisfies all five criteria presented in the introduction. For other count data models, we propose an extension to the deviance leading to  $R_{DEV}^2$  which satisfies at least criteria 1, 3 and 4.

The measure based on Pearson residuals,  $R_P^2$ , has the limitation that it satisfies none of the five criteria presented in the introduction. If measures based on raw residuals are used, the squared correlation coefficient between fitted and actual values,  $R_{COR}^2$ , has the advantage over measures based on residual sums of squares,  $R_{RES}^2$ , or explained sums of squares,  $R_{EXP}^2$ , of being bounded by 0 and 1, while the other two measures fail all of criteria 1-5. For the linear regression model these three measures would be equal.

While all these measures are designed to measure the explanatory power of the regressors, one measure,  $R_{DP}^2$ , has been introduced, that allows for cross model comparison, and implicitly tests whether the dispersion parameter in either Negbin 1 or Negbin 2 model is equal to zero, when compared to  $R_{DEV}^2$  for the Poisson model.

The different R-squared measures are compared in an extensive application to data on individual utilization of health care services from the 1977-78 Australian Health Survey. Utilization may vary with socioeconomic characteristics, health insurance status, recent health status and long-term health status. The three unweighted residual  $R^2$  measures can differ substantially.  $R_P^2$  and  $R_{DEV}^2$  can markedly differ from each other, even though in the fitted models the Pearson and deviance residuals are quite similar to each other. For all estimators and models  $R_P^2$  at times decreased with additional regressors, and an explained sums of squares version of  $R_P^2$ , not presented in the figures given here, differed quite substantially from  $R_P^2$ . In application to different health utilization measures, i.e. different dependent variables, the measures considered take higher values for count data that are a priori more easily predictable, and attribute high explanatory power to variables that a priori might be expected to be important.

The various  $R^2$  measures proposed give qualitatively similar results regarding detection of important regressors, with the notable exception of  $R_{EXP}^2$  and  $R_{RES}^2$  in the Negbin 2 model. Use of any of these measures, aside from this exception, is more informative than the current practice of not computing an R-squared.

The preferred deviance  $R^2$  requires specification of the likelihood or quasi-likelihood. If this is misspecified, just criteria 1 and 2 are satisfied. As an indication of the information content of the regressors  $R_{DEV}^2$  may still do well, however, as illustrated by the examples which were not necessarily Poisson or negative binomial. Furthermore, a (nonparametric) bootstrap revealed that the distribution of  $R_{DEV}^2$  was less variable and more symmetric than that of other measures.

## **ACKNOWLEDGEMENT**

This research began at The Australian National University. The authors thank the Economics Program in the Research School of Social Sciences and the Department of Statistics for their support and hospitality, and the Social Sciences Data Archive for providing the data. This

work has benefited from the thoughtful comments of three anonymous referees, an associate editor, and numerous seminar participants. Yong-eok Kim provided excellent research assistance.

## REFERENCES

- Australian Bureau of Statistics (1978), Australian Health Survey 1977-78 [machine-readable file], Canberra: Author.
- Bishop, Y.M.M., Feinberg, S.E. and Holland, P.W. (1975), *Discrete Multivariate Analysis: Theory and Practice*, Cambridge: MIT Press.
- Buse, A. (1973), "Goodness of Fit in Generalized Least Squares Estimation," *The American Statistician*, 27, 106-108.
- Cameron, A.C. and Trivedi, P.K. (1986), "Econometric Models Based on Count Data: Comparisons and Application of Some Estimators and Tests," *Journal of Applied Econometrics*, 1, 29-53.
- Cameron, A.C. and Windmeijer, F.A.G. (1994), "An R-squared Measure of Goodness of Fit for some Common Nonlinear Regression Models," Discussion Papers in Economics 94-12, University College London.
- Computing Resource Center (1992), *Stata Reference Manual: Release 3*, 5th Edition, Santa Monica: Author.
- Davison A.C. and Gigli, A. (1989), "Deviance Residuals and Normal Scores Plots", *Biometrika*, 76, 211-221.
- Dhrymes, P.J. (1986), "Limited Dependent Variables," in Z. Griliches and M.D. Intrilligator eds., *Handbook of Econometrics: Volume 3*, Amsterdam: North-Holland.
- Econometric Software, Inc. (1991), *Limdep User's Manual and Reference Guide: Version 6.0*, New York.
- Efron, B. (1978), "Regression and ANOVA with zero-one Data: Measures of Residual Variation," *Journal of the American Statistical Association*, 73, 113-121.
- Efron, B. and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, London: Chapman and Hall.

- Gurmu S. and Trivedi, P.K. (1992), "Recent Developments in Models of Event Counts: A Survey," Indiana University Working Paper in Economics 92-027.
- Hastie, T. (1987), "A Closer Look at the Deviance," *The American Statistician*, 41, 16-20.
- Landwehr, J.M., Pregibon, D. and Shoemaker, A.C. (1984), "Graphical Methods for Assessing Logistic Regression Models (with discussion)," *Journal of the American Statistical Association*, 79, 61-83.
- Lave, C. A. (1970), "The Demand for Urban Mass Transportation," *The Review of Economics and Statistics*, 37, 320-323.
- Merkle, L. and Zimmermann, K. F. (1992), "The Demographics of Labor Turnover: A Comparison of Ordinal Probit and Censored Count Data Models," *Reserches Economiques de Louvain*, 58, 283-307.
- McCullagh, P. (1986), "The Conditional Distribution of Goodness-of-fit Statistics for Discrete Data," *Journal of the American Statistical Association*, 81, 104-107.
- McCullagh, P. and Nelder, J.A. (1989), *Generalized Linear Models*, 2nd. ed., Chapman and Hall, London.
- McFadden, D. (1974), "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka (ed.), *Frontiers in Econometrics*, New York: Academic Press.
- Pierce, D.A. and Schafer, D.W. (1986), "Residuals in Generalized Linear Models," *Journal of the American Statistical Association*, 81, 977-986.
- Pregibon, D. (1981), "Logistic regression Diagnostics," *The Annals of Statistics*, 9, 705-724.
- Social Science Data Archives (1983), "Australian Health Survey 1977-78: User's guide for the Machine-Readable Data File," SSDA Study No. 196, The Australian National University, Canberra: Author.
- Veall, M. R. and Zimmermann, K. F. (1994), "Pseudo- $R^2$  Measures for Some Common Limited Dependent Variable Models," Department of Economics, McMaster University.
- Williams, D.A. (1987), "Generalized Linear Model Diagnostics Using the Deviance and Single Case Deletions," *Applied Statistics*, 36, 181-191.
- Windmeijer, F.A.G. (1995), "Goodness-of-Fit Measures in Binary Choice Models," *Econometric Reviews*, 14, 101-116.

**Table 1: Summary of R-Squared Measures for Count Data Models**

*Unweighted Residuals (for all models)*

$$R_{RES}^2 = 1 - \frac{\sum_i (y_i - \hat{\mu}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

$$R_{EXP}^2 = \frac{\sum_i (\hat{\mu}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2}$$

$$R_{COR}^2 = \frac{\left( \sum_i (y_i - \bar{y})(\hat{\mu}_i - \bar{\mu}) \right)^2}{\sum_i (y_i - \bar{y})^2 \cdot \sum_i (\hat{\mu}_i - \bar{\mu})^2}$$

*Pearson Residuals*

$$R_{P,P}^2 = R_{P,NB1(QL)}^2 = 1 - \frac{\sum_i \{(y_i - \hat{\mu}_i)^2 / \hat{\mu}_i\}}{\sum_i \{(y_i - \bar{y})^2 / \bar{y}\}}$$

$$R_{P,NB1(ML)}^2 = 1 - \frac{\sum_i \{(y_i - \hat{\mu}_i)^2 / \hat{\mu}_i\}}{\sum_i \{(y_i - \hat{\mu}_0)^2 / \hat{\mu}_0\}}$$

$$R_{P,NB2}^2 = 1 - \frac{\sum_i \{(y_i - \hat{\mu}_i)^2 / (\hat{\mu}_i + \hat{\alpha} \hat{\mu}_i^2)\}}{\sum_i \{(y_i - \bar{y})^2 / (\bar{y} + \hat{\alpha} \bar{y}^2)\}}$$

*Deviance Residuals*

$$R_{DEV,P}^2 = R_{DP,P}^2 = R_{DEV,NB1(QL)}^2 = \frac{\sum_i \{(y_i \log(\hat{\mu}_i / \bar{y}) - (\hat{\mu}_i - \bar{y}))\}}{\sum_i y_i \log(y_i / \bar{y})}$$

$$R_{DEV,NB1(ML)}^2 = 1 - \frac{\sum_i \left\{ \log \left( \frac{\Gamma(y_i + \hat{\alpha}^{-1} \hat{\mu}_{si}) \cdot \Gamma(\hat{\alpha}^{-1} \hat{\mu}_i)}{\Gamma(y_i + \hat{\alpha}^{-1} \hat{\mu}_i) \cdot \Gamma(\hat{\alpha}^{-1} \hat{\mu}_{si})} \right) - (\hat{\alpha}^{-1} \hat{\mu}_{si} - \hat{\alpha}^{-1} \hat{\mu}_i) \cdot \log(1 + \hat{\alpha}) \right\}}{\sum_i \left\{ \log \left( \frac{\Gamma(y_i + \hat{\alpha}^{-1} \hat{\mu}_{si}) \cdot \Gamma(\hat{\alpha}^{-1} \hat{\mu}_0)}{\Gamma(y_i + \hat{\alpha}^{-1} \hat{\mu}_0) \cdot \Gamma(\hat{\alpha}^{-1} \hat{\mu}_{si})} \right) - (\hat{\alpha}^{-1} \hat{\mu}_{si} - \hat{\alpha}^{-1} \hat{\mu}_0) \cdot \log(1 + \hat{\alpha}) \right\}}$$

$$R_{DP,NB1(ML)}^2 = 1 - \frac{\sum_i \left\{ y_i \log(y_i / \hat{\alpha}) - y_i + \log \left( \frac{\Gamma(y_i + \hat{\alpha}^{-1} \hat{\mu}_i)}{\Gamma(\hat{\alpha}^{-1} \hat{\mu}_i)} \right) + (y_i + \hat{\alpha}^{-1}) \cdot \log(1 + \hat{\alpha}) \right\}}{\sum_i y_i \log(y_i / \bar{y})}$$

$$R_{DEV,NB2}^2 = 1 - \frac{\sum_i \{y_i \log(y_i / \hat{\mu}_i) - (y_i + \hat{\alpha}^{-1}) \log((y_i + \hat{\alpha}^{-1}) / (\hat{\mu}_i + \hat{\alpha}^{-1}))\}}{\sum_i \{y_i \log(y_i / \bar{y}) - (y_i + \hat{\alpha}^{-1}) \log((y_i + \hat{\alpha}^{-1}) / (\bar{y} + \hat{\alpha}^{-1}))\}}$$

$$R_{DP,NB2}^2 = 1 - \frac{\sum_i \{y_i \log(y_i / \hat{\alpha} \hat{\mu}_i) - y_i - \log(\Gamma(y_i + \hat{\alpha}^{-1}) / \Gamma(\hat{\alpha}^{-1})) + (y_i + \hat{\alpha}^{-1}) \cdot \log(1 + \hat{\alpha} \hat{\mu}_i)\}}{\sum_i y_i \log(y_i / \bar{y})}$$

---

NOTE:  $\bar{\mu} = N^{-1} \sum_i \hat{\mu}_i$ ,  $\hat{\mu}_0$  is the estimate in the intercept only model, conditional on  $\hat{\alpha}$ , and  $\hat{\mu}_{si}$  is the estimate in the saturated model, conditional on  $\hat{\alpha}$ , and  $y \log y = 0$  when  $y = 0$ .

**Table 2: Description of Regressors for Health Application**

---

---

*Socioeconomic*

SEX	1 if female; 0 otherwise.
AGE	In years.
AGESQ	Age squared.
INCOME	Annual income in dollars.

*Health Insurance*

LEVYPLUS	1 if covered by private health insurance fund for private patient in public hospital, 0 otherwise.
FREEPOOR	1 if covered by government because low income, recent immigrant, unemployed, 0 otherwise.
FREEREPA	1 if covered by government because of old age or disability pension, or because invalid veteran or family of deceased veteran, 0 otherwise.
LEVY	1 if covered by Medibank, 0 otherwise is the omitted dummy.

*Recent Health Status*

ILLNESS	Number of illnesses in the past two weeks
ACTDAYS	Number of days of reduced activity in past two weeks due to illness or injury.

*Long-term Health Status*

HSCORE      General health score using Goldberg's method. High Score indicates bad health.

CHCOND1    1 if chronic condition that does not limit activity, 0 otherwise.

CHCOND2    1 if chronic condition that limits activity, 0 otherwise.

---

**Table 3: R-Squared Measures for Health Application**

Dep. Variable	$R^2$ Measure	Model		
		Poisson	NB1	NB2
DOCTORCO	$R^2_{RES}$	0.157 (0.026)	0.159 (0.025)	0.051 (0.062)
	$R^2_{EXP}$	0.243 (0.024)	0.208 (0.022)	0.502 (0.083)
	$R^2_{COR}$	0.164 (0.024)	0.162 (0.024)	0.150 (0.024)
	$R^2_P$	0.373 (0.033)	0.413 (0.040)	0.373 (0.035)
	$R^2_{DEV}$	0.223 (0.018)	0.171 (0.013)	0.229 (0.016)
	$R^2_{DP}$		0.268 (0.018)	0.278 (0.019)
HOSPADMI	$R^2_{RES}$	0.108 (0.020)	0.108 (0.020)	0.105 (0.020)
	$R^2_{EXP}$	0.116 (0.019)	0.111 (0.018)	0.125 (0.021)
	$R^2_{COR}$	0.108 (0.020)	0.108 (0.020)	0.106 (0.019)
	$R^2_P$	0.144 (0.034)	0.181 (0.042)	0.132 (0.037)
	$R^2_{DEV}$	0.131 (0.016)	0.108 (0.013)	0.131 (0.015)
	$R^2_{DP}$		0.159 (0.016)	0.156 (0.016)
MEDICINES	$R^2_{RES}$	0.370 (0.013)	0.370 (0.013)	0.367 (0.013)
	$R^2_{EXP}$	0.404 (0.015)	0.403 (0.015)	0.428 (0.016)
	$R^2_{COR}$	0.371 (0.013)	0.371 (0.013)	0.369 (0.013)
	$R^2_P$	0.380 (0.020)	0.411 (0.020)	0.372 (0.022)
	$R^2_{DEV}$	0.347 (0.011)	0.325 (0.011)	0.340 (0.011)
	$R^2_{DP}$		0.359 (0.010)	0.357 (0.010)

NOTE: Model with all regressors. Standard errors from 200 bootstrap replications in parentheses.

**Figure 1: Bootstrap Densities of  $R^2$  for DOCTORCO, Poisson Models**

**Figure 2: R-Squared for DOCTORCO, POISSON Model**

See rsq8figs.wp5 for the Wordperfect 5.1 originals