


Goodness of fit measures for logistic regression model: an application for students' evaluations of university teaching

Biagio Simonetti¹  · Pasquale Sarnacchiaro² ·
M. Rosario González Rodríguez³

Published online: 15 September 2016
© Springer Science+Business Media Dordrecht 2016

Abstract This study investigates the assessment of university teachers. The evaluation of university teaching plays an important role in the university system in the light of the fact, that a part of the financial transfer from educational ministry is based on the quality of teaching. From Statistical point of view the use of regression models has been widely applied in this area. The logistic regression model is particularly used as discrete choice model using dichotomous dependent variable. For many regression analyses the lack of a goodness-of-fit measure is more important than coefficient interpretability. The goal of this paper is to present an overview of a few easily employed methods for assessing the model fitness of Logistic Regression Model by Pseudo- R^2 . Moreover the assessment is carried out through a simulation study to analyse the pattern (behaviour) of each measure, with precise focus on change of multiple correlation among the variables. In this paper a survey on student satisfaction (SS) of university teaching system was conducted. The instrument used in this paper is a questionnaire proposed by a different research group in 2010. The collected data was elaborated by a full reflective Structural Equation Model using PLS path model estimation. The initial results showed that the influence of the Organization and Infrastructures on the Student Satisfaction were statistically insignificant. Therefore a more complex model was supposed, the final results showed that the influence of Organization and Infrastructures on the SS was indirect, that is the Organization and the Infrastructures exert an influence upon the SS through the Didactics.

Keywords Logistic Regression · Pseudo- R^2 · Students' Evaluations · University teaching

✉ Biagio Simonetti
simonetti@unisannio.it

¹ University of Sannio, Benevento, Italy

² University Unitelma-La Sapienza di Roma, Rome, Italy

³ University of Seville, Seville, Spain

1 Introduction

In literature, quality of university teaching has been analysed by using structured questionnaire to collect information regarding university teaching and the application of regression models. Many such analyses involve a dichotomous outcome, or a dependent variable (present/absent, yes/no, live/die, etc.). In these studies the logistic regression model (LRM) has been used as statistical model of discrete choice. The reasons of choosing LRM are: (1) The ease of interpretation of the estimated coefficients as “adjusted log odds ratios,” (2) The ability to estimate the probability that a particular subject will develop the outcome, (3) The wide availability of easily used and reliable software to perform the computations.

However, one of the shortcomings of logit analyses is the relative lack of diagnostics that regression analysts expect (Hagle and Mitchell 1992). Computer algorithms for LRM present estimated coefficients and standard errors. Although logit coefficients lack the intuitive interpretation of Ordinary Least Squares (OLS) coefficients, but they can be used for hypothesis testing (Aldrich and Nelson 1984). For many regression analysts, the lack of a goodness-of-fit measure is more important than coefficient interpretability. Unfortunately, no direct equivalent to OLS- R^2 exists for logit models. Some regression analysts find R^2 to be of limited utility (King 1990). Others find R^2 to be of substantial usefulness (Lewis-Beck et al. 1990).

We observed that R^2 is frequently employed as a measure of the percentage variation in the dependent variable *explained* by the regression (Hanushek and Jackson 1977). In addition, R^2 can be used to generate the familiar *F-test* statistic that tests the hypothesis that all of the coefficients except the intercept are zero (Achen 1979).

We can understand the reason why the analysts who find the familiar R^2 statistic to be of considerable utility, desire a similar measure for evaluating LRM performance. A wide range of pseudo- R^2 have been proposed in the past three decades (see, e.g., MacFaden 1973; Maddala 1983; Dhrymes 1986; Nagelkerke 1991). Many LRM computer algorithms display one or more pseudo- R^2 statistics. Although the sample properties of pseudo- R^2 statistics haven been studied in the literature (Hagle and Mitchell 1992; Veall and Zimmermann 1995; Windmeijer 1995; Hu et al. 2006; Hensher and Johnson 1981; Hoeteker 2007; Mittlböck 2002; Mittlböck and Heinzl 2001, 2002; Mittlböck and Schemper 1999; Mittlböck and Waldhör 2000; Veall and Zimmermann 1996; Walker and Smith 2016; Zheng and Agresti 2000). The analysts are facing practical difficulty which is selection among the pseudo- R^2 measures. Advices regarding preferred substitute of R^2 and under what circumstances that preference holds, are absent from the literature.

The goal of this paper is to present an overview of a few easily employed methods for assessing the model fitness of Logistic Regression Model by Pseudo- R^2 . Moreover the assessment is carried out through a simulation study to analyse the pattern (behaviour) of each measure, with precise focus on change of multiple correlation among the variables.

The paper is organized as follows: Sect. 2 presents a classification of the Pseudo- R^2 statistics considered in four classes. It also provides a discussion about the properties of each measure and the relationship among them. In Sect. 4 we discuss the use of the different measures and some considerations about further developments at end of the paper.

The practice of student’s evaluations of university teaching via teaching evaluation questionnaires is now widespread (Achen 1979; Aldrich and Nelson 1984). In Italy, all of the Universities carry out surveys to measure Student Satisfaction (SS). Most of these

surveys are conducted through the administration of a evaluation questionnaire. In 2008 the CNVSU (Comitato Nazionale per la Valutazione del Sistema Universitario—National Committee for University System Evaluation) charged a very large Research Group (RG) to build and validate a questionnaire for the assessment of teaching to be administered via web (Amemiya 1981). This questionnaire has been used for a SS assessment at the University Federico II of Naples. Aim of this paper is to propose a full reflective Structural Equation Model (SEM) and to analyze the collected data. In the second section, the main results of the research, conducted by the RG, are shown. In the third section, it is presented both the SS survey conducted at the University Federico II of Naples and the SEM used to analyse the SS and to detect the drivers of their behaviour. In the fourth section, the results of the SEM are pointed out, a brief conclusion ends the paragraph.

2 The Pseudo- R^2

In LRM, there is no true R^2 value as is the case with OLS regression. Deviance is the lack of fit between observed and predicted values, so it can be regarded as a measure of how poorly the model fits. An analogy can be made to sum of squares residual in OLS. In particular, in logistic regression $-2 \log(L_0)$, the null model has the constant estimated by likelihood. Null Model is analogous to the total sum of squares in OLS and $-2 \log(L_1)$. The alternative model estimated by likelihood, is analogous to the residual sum of squares in OLS. The *log-likelihood* is not really a sum of squares, so this measure does not have an *explained variance* interpretation. Rather it indicates the relative improvement in the likelihood of observing the sample data under the hypothesized model as compared to a model with the intercept alone. The proportion of unaccounted variance that is reduced by adding variables to the model is the same as the proportion of variance accounted for, or R^2 .

Starting from this observation, statistical literature has proposed several substitutes for the R^2 statistic, also called Pseudo- R^2 . Many LRM computer algorithms present one or more pseudo- R^2 statistics. In this section the adjusted coefficients of determination for LRM are reviewed, starting from a classification of these measures in four classes. The criterion that characterizes this classification is the way in which these measures have been constructed as seen in Table 1. The first class measure is based on the variance decomposition of the estimated logit. Second and third class include those measures constructed using the likelihood method and the log-likelihood method respectively. The last class measures are based on the estimated probabilities.

2.1 Basing on the variance decomposition of the estimated logits class

McKelvey and Zavoina (1975) proposed the most commonly employed pseudo- R^2 . The estimated logit coefficients can be utilized to calculate explained variance by computing the variance of the forecasted values for the latent dependent variable.

In R^2_{MZ} the $\hat{V}(\alpha + \sum \beta_k X_k)$ is the sample variance of the linear predictor, or the variance accounted for by the predictor set, and the denominator is, once again, an estimate of the variance of Y^* . The R^2_{MZ} is the best indicator for the true R^2 of the OLS regression.

Table 1 Classification of Pseudo R^2

Class	Author	Formula	Range
Basing on the variance decomposition of the estimated Logits	McKelvey and Zavoina (1975)	$R_{MZ}^2 = \frac{\hat{V}(x + \sum \beta_k X_k)}{\hat{V}(x + \sum \beta_k X_k) + \pi^2/3}$	
Likelihood-based measures:	Maddala (1983)	$R_M^2 = 1 - \left(\frac{L_A}{L_0}\right)^{\frac{2}{n}}$	$\left[0, \left(1 - L_A^{\frac{2}{n}}\right)\right]$
	Cragg–Uhler (1970)	$R_{CU}^2 = \frac{L_A^{\frac{2}{n}} - L_0^{\frac{2}{n}}}{1 - L_0^{\frac{2}{n}}}$	$[0, 1]$
Log-Likelihood-based measures:	Cox and Snell (1989)	$R_{CS}^2 = 1 - \left[\frac{-2 \ln(L_0)}{-2 \ln(L_1)}\right]^{2/n}$	$0 \leq R_{CS}^2 < 1$
	Nagelkerke (1991)	$R_N^2 = \frac{n \times R_{CS}^2}{1 - e^{[2 \ln(L_0)]}}$	$[0, 1]$
	McFadden	$R_{mF}^2 = 1 - \frac{\ln(L_A)}{\ln(L_0)}$	$[0, 1]$
	Aldrich–Nelson (1984)	$R_{AN}^2 = \frac{2[\ln(L_A) - \ln(L_0)]}{2[\ln(L_A) - \ln(L_0)] + (n_{M_A} \times \frac{\pi^2}{3})}$	$\left[0; \frac{-2 \ln(L_0)}{\frac{\pi^2}{3} n - 2 \ln(L_0)}\right]$
Basing on the estimated probabilities	Efron (1978)	$R_{LE}^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{p}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$	$[0, 1]$
	Squared Pearson correlation (Mittlböck and Schemper 1996)	$R_{MS}^2 = \frac{\left[\sum_{i=1}^n (y_i - \bar{y})(\hat{p}(x_i) - \bar{p}(x))\right]^2}{\left[\sum_{i=1}^n (y_i - \bar{y})^2\right] \left[\sum_{i=1}^n (\hat{p}(x_i) - \bar{p}(x))^2\right]}$	$[0, 1]$

2.2 Likelihood class

The pseudo- R^2 statistics belonging to the Likelihood class are constructed using L_A and L_0 . They represent the null and alternative models which are estimated by likelihood, where n is the sample size. The measure R_M^2 is identical to standard multiple R^2 when applied to generalized liner model (Maddala 1983). R_M^2 can never reach a value of one even if the model predicts perfectly, so it has been suggested in the statistics R_{CU}^2 which are obtained by dividing R_M^2 by his maximum value (Cragg–Uhler 1970). Hypothetically, R_{CU}^2 can reach a value of one but the correction appear to be cosmetic as R_{CU}^2 has to be 100 % for complete agreement. There is no indication why scaling of the intermediate values of R_{CU}^2 should be adequate. Moreover R_M^2 and R_{CU}^2 do not have good interpretability in terms of probability (p_i). R_M^2 has the following properties (Nagelkerke 1991)

- R_M^2 is consistent with classical R^2 , e.g. linear regression yields the classical R^2 .
- R_M^2 is consistent with maximum likelihood i.e. the maximum likelihood estimates of the model parameters maximize value of R^2 .
- R_M^2 is asymptotically independent of the sample size (n)
- R_M^2 is interpreted as the proportion of *explained variation*, or rather, $(1 - R^2)$ is interpreted as the proportion of *unexplained variation*. Variation should be construed generally as a measure of the extent to which a distribution is not degenerate.
- R_M^2 is dimensionless (i.e. it does not depend on the units used).

- By using Taylor Expansion & considering y has a probability density $p(y|\beta x + \alpha)$, it can be shown that to a first order approximation, R_M^2 is the square of the Pearson correlation between x and p (the efficient score of the model). It is the derivative with respect to β of $\log\{p(y|\beta x + \alpha)\}$ at $\beta = 0$.

2.3 Log-likelihood class

In this class we find the Pseudo- R^2 based on the criteria of the log-Likelihood. The term $\ln(L_A)$ is the maximized *log likelihood* for the fitted model and $\ln(L_0)$ is the maximized *log likelihood* for the null model containing only an intercept term. n is the simple size. To interpret R_{CS}^2 (Cragg and Uhler 1970) the statistics can be rewritten as follow: $-\ln(1 - R_{CS}^2) = 2[(\ln L_A - \ln L_0)/n]$. The right side of this equation can be interpreted as the amount of information gained when including the predictors into model A , in comparison with the *null* model. The R_{CS}^2 cannot attain a value of 1. To address this disadvantage Nagelkerke (1991) proposed the R_N^2 , which can attain a maximum value of 1.

If we work with *single-trial* syntax, then the saturated model has a dummy variable for each observation (log saturated model = 0). Resultantly, the deviance R^2 simplifies to R_{mF}^2 (McFadden 1973).

The R_{mF}^2 has been criticized on the following grounds: the denominator is constant and the numerator is one-half the likelihood ratio test for significance; thus the quantity R_{mF}^2 is “*nothing more than an expression of the likelihood ratio test and, as such, is not a measure of goodness of fit*”. We disagree with this opinion. From our point of view, the basic idea behind R_{mF}^2 is to compare the log-likelihood gain achieved by the fitted model (numerator) with the maximum potential log-likelihood gain (denominator). As R_{mF}^2 is a measure of comparison of two log-likelihood gains, it can be treated as an indicator of goodness of fit. Also R_{mF}^2 has the additional advantage of interpretation in terms of reduction in recoverable information. Also R_{mF}^2 can be large even if the strength of association is weak. As a matter of fact, R_{CS}^2 and R_{mF}^2 sometimes work in opposite directions: R_{CS}^2 can be too low and R_{mF}^2 can be too high in some cases. it creates an additional incentive to use them in combination with each other. Thus, the use of R_{mF}^2 is not questionable, what actually matters is the use of R_{mF}^2 in combination with R_{CS}^2 .

The R_{AN}^2 statistics (Aldrich and Nelson 1984) employs log-likelihood ratios based on two passes through the data. The first pass generates the likelihood value for the null model (i.e., with an intercept only). The likelihood value of the null hypothesis is commonly denoted as L_0 . The second pass generates the likelihood value for the full model, commonly denoted as L_A . Once the logarithms of the likelihood values are generated, the calculation of R_{AN}^2 is straight forward. The Aldrich-Nelson measure employs the familiar χ^2 statistic. $-2LLR$ is defined as $-2 \ln(L_0/L_1)$ with k , the number of independent variables estimated (excluding the constant) and degrees of freedom. The χ^2 statistic is often included as part of the output for logit packages, even if no *pseudo* R^2 is reported. The Dhrymes measure is a straight forward application of the R_{AN}^2 . Achen contends that *pseudo* R^2 statistics presented above suffer from one defect i.e. they are not acceptable test statistics (Achen 1979). Achen argues that they cannot be used successfully to determine whether all of the coefficients or a subset of the coefficients in a model are statistically different from zero (except for the intercept). This is not as serious a failing as Achen would have us believe, since the χ^2 statistic is the analog to the F-test in regression analysis

(Aldrich and Nelson 1984). Achen provides a definition for a *pseudo* R^2 that gives asymptotically the same test as the χ^2 statistic (Achen 1979).

Starting from the R^2_{AN} Hagle and Mitchell (1992) proposed a modified version, where $\max(R^2_{AN}|\hat{\pi})$ is the maximum value attainable by R^2_{AN} , provided $\hat{\pi}$ which is the sample proportion with Y is equal to 1. The formula for $\max(R^2_{AN}|\hat{\pi})$ is:

$$-2[\hat{\pi} \log \hat{\pi} + (1 - \hat{\pi}) \log(1 - \hat{\pi})] / \{1 - 2[\hat{\pi} \log \hat{\pi} + (1 - \hat{\pi}) \log(1 - \hat{\pi})]\}.$$

This modification R^2_{HM} , ensures that the measure will be bounded by 0 and 1. Despite the fact that it is so bounded, here 1 indicates perfect predictive efficacy, so the value is not really interpretable.

2.4 Basing on the estimated probabilities class

The first measure proposed in this class starts from the general form of the Proportion of the explained variation:

$$PEV = \frac{\left[\sum_{i=1}^n D(y_i) - \sum_{i=1}^n D(y_i|x_i) \right]}{\sum_{i=1}^n D(y_i)}$$

Where $D(y_i)$ denotes a measure of the distance of y_i from an unconditional central location parameter and $D(y_i|x_i)$ denotes a measure of the distance of y_i from conditional central location parameter. The measures belonging to this class differ in their specification of $D(y_i)$ and $D(y_i|x_i)$. In the R^2_{LE} measure (Efron 1978), $D(y_i)$ and $D(y_i|x_i)$ represent the squared distance between observed (y_i) and predicted (\bar{p} and \bar{p}_i) outcomes under the *null* model (only with intercept) and under the *full* model (with covariates).

When both models are applicable, the R^2_{LE} is found to be consistent with the results of R^2 in the general linear model. Furthermore, the interpretation of R^2_{LE} is more intuitive to many people, as the use of squared residuals is a very basic in statistics whereas the use of likelihood is not so clear in its interpretation. In the case of small sample size the R^2_{LE} may give artificially inflated values and use of its corrected version R^2_{MS} is recommended (Mittlböck and Schemper 1996). This last measure is computed using the squared correlation of y and \hat{p} . The \hat{p}_i denotes the estimates from a logistic regression by $\Pr(y_i = 1|x_i) = \hat{p}_i = \exp(\hat{\beta}x_i) / 1 + \exp(\hat{\beta}x_i)$, with $\hat{\beta}$ denoting the estimated parameter vector. Moreover, $\Pr(y_i = 1) = \bar{p} = \sum_i (y_i/n)$.

3 Case study

3.1 The construction of the questionnaire

This paragraph summarizes the main results of the project carried out in 2010. It is supported by CNVSU and developed by RG. The aim is to design, build and evaluate a questionnaire for the student evaluation of university teaching in Italy (Begg and Gray 1984). After conducting a survey regarding the questionnaires used in Italy and abroad for

the SS assessment, RG has defined four different questionnaires to compare with a survey, which are as under:

- The Standard Questionnaire (SQ) is proposed by CNVSU. It is composed of 15 questions with items rated on a 4-point scale.
- A revisited version of the SQ is composed of 15 questions with items rated on a 10-point scale.
- The Experimental Questionnaire (EQ) is proposed by RG and composed of 9 questions with items rated on two joint 4 and 10-point scales.
- The EQ revisited Questionnaire is composed of 9 questions with items rated on two disjoint 4 and 10-point scales.

(For information about Questionnaires, Visit (www.cnvsu.it/library/downloadfile.asp?id=1177).

The four questionnaires were administered to a sample of about 1500 students of the University of Brescia and of the University of Sannio. In order to investigate the importance of the questions and to evaluate the SS, the RG adopted the following strategies:

For the first point a factorial analysis which was rotated and not rotated with Promax rotation, was performed. Then a logistic regression analysis was performed. Purpose was to detect the model that best describes the 4 and 10-point data obtained from the questionnaires. It was determined by the minimum residual method (Colin Cameron and Windmeijer 1996) on the polychoric correlation matrix. The software used was Prelis (Version 2.54). The selection of questions to be retained in the questionnaire was made considering the results of factor analysis and using the criterion proposed by Tabachnick and Fidell (Cragg–Uhler 1970). For verifying the stability of parameters or the replicability of pattern/structure coefficients (loadings), a nonparametric bootstrap factor analysis have been performed (Begg and Gray 1984). In order to evaluate the importance of questions in the questionnaires with respect to the overall SS, a logistic regression analysis (using the forward selection approach) was used.

The Rasch Analysis with the Rating Scale Model (Dhrymes 1986) was used to assess the properties of all the ordinal 4 and 10point scales of response used in the questionnaires. The performed scaling analysis has pointed out the greater flexibility offered by the 10-point scale. Its votes can be merged in a 4-point scale preserving the comparisons with previous surveys. Moreover, more informative statistical analysis could be conducted for the future surveys. For these reasons, the RG proposed a new scale (Amemiya 1981).

As a final product of this research, the RG proposed a new version of the questionnaire for evaluating university teaching composed of twelve questions (Table 1) with the scale proposed above. This questionnaire has been used for a survey conducted on a sample of students of the University of Naples Federico II.

3.2 Analysis of the student satisfaction by logistic regression

The proposed questionnaire was administrated to a total sample of 511 students in the University of Naples Federico II. Sample was selected through simple random sampling. Firstly, a Factorial Analysis has been elaborated to identify the main aspects that influence the SS. According to the Kaiser and Guttman rule, four-factor was retained, which were labelled as follows: *Organization*, *Infrastructures*, *Didactics* and *SS*. In order to formalize a scheme for the interpretation of the *SS* and detecting the drivers of their behaviour, a SEM was elaborated. *Organization* (ξ_1), *Infrastructures* (ξ_2) and *Didactics* (ξ_3) were considered exogenous latent variables (LVs), while *SS* (η_1) was considered as an endogenous latent variable (LV). The use of the SEM in SS study is quite widespread (Efron 1978). In SEM techniques we distinguish

Table 2 The questionnaire—sections and items

Sections			
Organization	Infrastructure	Didactics	Student satisfaction
Questions			
Clarity of exam modality (X_1)	Appropriate classroom for exercitations (X_4)	Supplementary activities are useful (X_6)	Overall satisfaction (Y_1)
Teacher available for explanation (X_2)	Appropriate classroom for lessons (X_5)	Teacher stimulate the interest (X_7)	Interest for the topics (Y_2)
Teaching timetable respected (X_3)		Teacher exposes clearly (X_8)	
		Charged study proportional to the CFU (X_9)	
		Appropriate teaching materials (X_{10})	

Table 3 Pseudo R^2 values

Index author's name	Index value
McKelvey and Zavoina	0.56
Maddala	0.57
Cragg–Uhler	0.60
Cox and Snell	0.72
Nagelkerke	0.44
McFadden	0.39
Aldrich–Nelson	0.61
Efron	0.45
Mittlebock–Schemper	0.64

between two families: covariance-based techniques and variance-based techniques. Covariance-based techniques are represented by linear structural relations (LISREL) (Hagle and Mitchell 1992). Partial least squares (PLS) path modelling is the most prominent representative of variance-based techniques (Hanushek and Jackson 1977) (Table 2).

In PLS approach, there are less probabilistic hypotheses. Data are modelled by a succession of simple or multiple regressions and there is no identification problem. In LISREL, the estimation is done by maximum likelihood, basing on the hypothesis of multi-normality and allows the modelisation of the variance–covariance matrix. However, identification problems and non-convergence of the algorithm are sometimes encountered. PLS approach was chosen because it has less stringent assumptions about the distribution of variables and error terms. Although PLS algorithm adopts a formative scheme, it is currently being used with both kinds of models. However, whether it should consistently used in reflective schemes or not, is now debated (Heinzl and Mittlböck 2002; Heinzl et al. 2002).

In order to measure the model goodness of fit, several indices have been calculated. The results are shown in Table 3.

4 Conclusion

We can conclude that in order to make a judgement on these measures, we require a criterion that a suitable measure should obey. We think that a good measure of explained variation with logistic regression should possess the following properties (Mittlböck and Schemper 1996):

Intuitively clear interpretation;

The potential range of values of a measure should be $[0, 1]$ with the end points corresponding to complete lack of predictability and perfect predictability, respectively.

Moreover it could be a good idea to use two or three Pseudo- R^2 belonging to different classes, in order to assess how the choice model fits. The above approaches to calculating R^2 with logistic regression are only two of several different approaches. At this point, there does not seem to be much agreement on which approach is best. Moreover, researchers do not seem to report either very often when logistic analyses are performed. My recommendation would be to use these to make some reference to their “approximate” accuracy without considering them to be definitive values for the percentage of variance.

References

- Achen, C.H.: Alternative Measure of association in the probit model. Unpublished manuscript, University of California, Berkeley, 1979
- Aldrich, J.H., Nelson, F.D.: Linear Probability, Logit, and Probit Models. Sage University Press, Beverly Hills (1984)
- Amemiya, T.: Qualitative response models: a survey. *J. Econ. Lit.* **19**, 1483–1536 (1981)
- Begg, C.B., Gray, R.: Calculation of polychotomous logistic regression parameters using individualized regression. *Biometrika* **71**, 11–18 (1984)
- Colin Cameron, A., Windmeijer, F.A.G.: An R-squared measure of goodness of fit for some common non linear regression models. *J. Econom.* **77**, 329–342 (1996)
- Cox, D.R., Snell, E.J.: The analysis of binary data, 2nd edn. Chapman and Hall, London (1989)
- Cragg, S.G., Uhler, R.: The demand for automobiles. *Can. J. Econ.* **3**, 386–406 (1970)
- Dhrymes, P.J.: Limited dependent variables. In: Griliches, Z., Intriligator, M.D. (eds.) *Handbook of Econometrics*, vol. III, pp. 1567–1631. North Holland, Amsterdam (1986)
- Efron, B.: Regression and Anova with zero-one data. Measures of residual variation. *J. Am. Stat. Assoc.* **73**, 113–121 (1978)
- Hagle, T.M., Glenn, E.M.: Goodness of fit measures for probit and Logit. *Am. J. Polit. Sci.* **36**(3), 762–784 (1992)
- Hanushek, E.A., Jackson, J.E.: *Statistical Methods for social scientists*. Academy Press, New York (1977)
- Heinzl, H., Mittlböck, M.: R-squared measures for the inverse Gaussian regression model. *Computational Statistics* **17**(4), 2002 (to appear)
- Heinzl, H., Stare, J., Mittlböck, M.: A measure of dependence for the stratified Cox proportional hazards regression model. *Biom. J.* **44**, 671–682 (2002)
- Hensher, D.A., Johnson, L.W.: *Applied Discrete Choice Modelling*. Croom Helm, London (1981)
- Hoeteker, G.: The use of logit and probit models in strategic management research: critica issues. *Strateg. Manag. J.* **28**, 331–343 (2007)
- Hu, H., Saho, J., Palta, M.: Pseudo- R^2 in logistic regression model. *Stat. Sin.* **16**, 847–860 (2006)
- King, G.: When not to use R-squared. *Polit. Methodol.* **3**(2), 9–11 (1990)
- Lewis-Beck, Michael S., Skalaban, Andrew: When not to use R-squared. *Polit. Methodol.* **3**(2), 11–13 (1990)
- Maddala, G.S.: *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, Cambridge (1983)
- McFadden, D.: Conditional logit analysis of qualitative choice behaviour. In: Zarembka, P. (ed.) *Frontiers in Econometrics*, pp. 105–142. Academic Press, New York (1973)

- McKelvey, Richard D., Zavoina, W.: A statistical model for the analysis of ordinal level dependent variables. *J. Math. Sociol.* **4**, 103–120 (1975)
- Mittlböck, M.: Calculating adjusted R^2 measures for poisson regression models. *Comput. Methods Programs Biomed.* **68**, 205–214 (2002)
- Mittlböck, M., Heinzl, H.: A note on R^2 measures for Poisson and logistic regression models when both models are applicable. *J. Clin. Epidemiol.* **54**, 99–103 (2001)
- Mittlböck, M., Heinzl, H.: Measures of explained variation in Gamma regression model. *Commun. Stat. Simul. Comput.* **31**, 61–73 (2002)
- Mittlböck, M., Schemper, M.: Explained variation for logistic regression. *Stat. Med.* **15**, 1987–1997 (1996)
- Mittlböck, M., Schemper, M.: Computing measures of explained variation for logistic regression models. *Comput. Methods Programs Biomed.* **58**, 17–24 (1999)
- Mittlböck, M., Waldhör, T.: Adjustments for R^2 -Measures for Poisson regression models. *Comput. Stat. Data Anal.* **34**, 461–472 (2000)
- Nagelkerke, N.J.D.: A note on a general definition of the coefficient of determination. *Biometrika* **78**(3), 691–693 (1991)
- Veall, M.R., Zimmermann, K.F.: Comments on goodness of fit-measures in binary choice models. *Econom. Rev.* **14**, 117–1120 (1995)
- Veall, M.R., and Zimmermann, K.F. (1996). Pseudo- R^2 Measures for some Common Limited Dependent Variable Models. Sonderforschungsbereich 386, Paper 18
- Walker, D. A. and Smith, T. J.: Nine Pseudo R^2 indices for binary logistic regression models. *J. Mod. Appl. Stat. Methods*, 15(1), 2016, Article 43
- Windmeijer, F.A.G.: Goodness-of-fit measures in binary choice models. *Econ. Rev.* **14**, 101–116 (1995)
- Zheng, B.Y., Agresti, A.: Summarizing the predictive power of a generalized linear model. *Stat. Med.* **19**, 1771–1781 (2000)