Tilburg University

Evaluating Pseudo-R²'s for Binary Probit Models

Veall, M.; Zimmermann, K.

Publication date:
1990

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):

Veall, M., & Zimmermann, K. (1990). *Evaluating Pseudo-R²'s for Binary Probit Models*. (CentER Discussion Paper; Vol. 1990-57). CentER.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

CBM
CBM
R
8414
1990
57

entER
for
mic Research

Discussion paper

No. 9057

**EVALUATING PSEUDO- R^2 'S FOR
BINARY PROBIT MODELS**

by Michael R. Veall
and Klaus F. Zimmermann

October 1990

ISSN 0924-7815

R20

330.115

Evaluating Pseudo- R^2 's for Binary Probit Models^{*}

Michael R. Veall
McMaster University

and

Klaus F. Zimmermann
University of Munich

May 1990

- * Financial support from the von Humboldt Foundation and the Natural Sciences and Engineering Research Council of Canada to the first author and from the German Science Foundation (DFG) and from the Fulbright Commission to the second author is gratefully acknowledged. We wish to thank John De New for excellent research assistance. Preliminary versions of the paper were presented at the Munich Applied Econometrics Workshop in July 1989 and at the Econometrics Seminar of CentER, Tilburg University, in May 1990. We wish to thank Gerhard Arminger, Gebhard Flaig, François Laisney and other seminar participants for many useful comments.

Correspondence to:

Prof. Dr. Klaus F. Zimmermann
Seminar of Labor and Population Economics
University of Munich
Ludwigstr. 28 RG
8000 München 22
Federal Republic of Germany

Abstract

Many applied researchers using limited dependent variable models have found it disadvantageous that a widely accepted Pseudo- R^2 does not exist for this type of estimation. In addition to the problem of choosing a Pseudo- R^2 for new research, there is a difficulty in comparisons among existing studies which have used different measures. To offer at least some degree of comparability in the binary probit case, this paper contains a Monte Carlo experiment in which a number of alternative Pseudo- R^2 measures are all evaluated on the same data sets. In terms of choosing a single Pseudo- R^2 , it would be useful if comparability also extended to the ordinary least squares (OLS) case. A good Pseudo- R^2 should therefore be able to mimic the OLS- R^2 on the underlying continuous data. The results are surprisingly clear: a measure suggested by McKelvey/Zavoina performs the best under this criterion. However, in the more likely case of low Pseudo- R^2 's, a normalization of a measure proposed by Aldrich/Nelson which we suggest is almost as good as the McKelvey/Zavoina, and is in general easier to calculate. We also provide tables that may give an applied researcher some idea as to what a given Pseudo- R^2 value may imply about the underlying R^2 .

Keywords: Pseudo- R^2 , Binary Probit, Goodness-of Fit

1. Introduction

In recent years, limited dependent variable models have become increasingly popular. However, many researchers have found it disadvantageous that a widely accepted Pseudo- R^2 does not exist for this type of estimation, although many approaches have been suggested. Some measures are based on the likelihood-ratio test statistic, while others evaluate the predicted probabilities or residuals of the binary regressions, or use prediction-realization tables for analysis. (See Amemiya, 1981; Maddala, 1983; Zimmermann, 1986; Greene, 1990.)

The goal of this paper is to provide guidance for researchers in choosing and interpreting Pseudo- R^2 's. We concentrate on the binary probit case. Because the dependent variable is either 0 or 1 there is no clear cut measure based on explained variation and it is well known that measures based on the success in predicting the binary variable are heavily influenced by the proportion of ones in the sample. Hence the first issue is how alternative Pseudo- R^2 's may be compared.

As pointed out for example by Dhrymes (1986, pp. 1578-79), no Pseudo- R^2 in the probit case has all the properties of ordinary least squares (OLS) R^2 with its simple relationship to the F statistic of the hypothesis that all regression slope coefficients are zero and its interpretations both as the squared correlation coefficient of the predicted and actual values of the dependent variable and as the ratio of explained variation to total variation. Our view is that reporting R^2 in the OLS case has become standard because it is a readily available informal index of "degree of explanation" that can be compared across regression models. But comparisons from different equations are tempered with judgement involving the type of data. For example, R^2 's in studies with time series aggregate economic data are often expected to be .7, .8 or higher but with cross section data on individuals, an R^2 of .4 may be viewed as very high.

Given this view, the most important attribute of a Pseudo- R^2 is that it be used consistently so that comparisons may be made for different studies. (Aldrich and Nelson, 1984, pp. 55-59 emphasize this point as well.) Unfortunately, different Pseudo- R^2 's have been used by different researchers. This paper therefore provides some basis for comparability, showing for example that some Pseudo- R^2 's will typically be greater than others from the same data. Moreover it seems most useful to put any Pseudo- R^2 into the context of the vast experience with OLS- R^2 . We therefore rank our candidate Pseudo- R^2 's in their ability to mimic the continuous data OLS- R^2 , which is unobserved in application. The comparisons are based on a Monte Carlo experiment. Continuous data on the endogenous variable are simulated based on a sequence of true models generating OLS- R^2 's in the [0,1]-range. We then explore the relationships among the various Pseudo- R^2 's calculated from the probit of the corresponding classified binary data and consider which is (i) closest to OLS- R^2 and/or is (ii) the best predictor of OLS- R^2 .

From a set of six widely used measures, our ranking is surprisingly clear: a measure suggested by McKelvey/Zavoina (1975) has the closest correspondence to OLS- R^2 . If we use a non-linear estimated conversion function of the

Pseudo- R^2 measures to the R^2 from the continuous case for predictions, a measure suggested by McFadden (1973) can be as good or better under certain conditions. The Aldrich/Nelson (1984) measure does poorly under our criteria, although its behavior is improved considerably under a normalization we propose. This normalization may be valuable as the resulting statistic performs almost as well as the McKelvey/Zavoina version, yet is easier to obtain.

Section 2 presents the measures used. Section 3 explains the experimental design and discusses the results. Section 4 contains a summary.

2. Some Pseudo- R^2 measures: An overview

A Pseudo- R^2 is a goodness-of-fit statistic when the endogenous variable is discrete, and is an analog to the R^2 -measure in the continuous data/linear regression case. Measures based on the Probit residuals, the likelihood-ratio test statistic and the prediction-realization table have been suggested. We concentrate here on a selection of widely used measures for individual data. Measures specifically developed for grouped data or prediction realization tables are reserved for future research.

Assume an underlying (continuous) response variable Y_i^* and the relationship

$$(1) \quad Y_i^* = \xi + \delta X_i + U_i, \quad i = 1, 2, \dots, N,$$

where X is an exogenous variable and ξ and δ are parameters; i is the index of the individual, U a random standard normal error term and N the sample size. Y_i^* is unobserved in practice, but will be simulated in our experiments. Instead of Y_i^* , a dummy variable Y_i defined as

$$(2) \quad Y_i = 1 \text{ if } Y_i^* > 0, \text{ and } Y_i = 0 \text{ otherwise}$$

is observed. Equations (1) and (2) imply

$$(3) \quad \begin{aligned} \Pr(Y_i = 1) &= \Pr(U_i > -\xi - \delta X_i) \\ &= F(\xi + \delta X_i), \end{aligned}$$

where F is the cumulative normal distribution function for U . The log-likelihood function is

$$(4) \quad L = \sum_{i=1}^N Y_i \log F(\cdot) + \sum_{i=1}^N (1 - Y_i) \log [1 - F(\cdot)].$$

A popular suggestion to measure goodness-of-fit has been the squared correlation coefficient between Y and F (Neter/Maynes, 1970; Lave, 1970; Efron, 1978; Morrison, 1972; Goldberger, 1973):

$$(5) \quad R_C^2 = \frac{[\text{cov}(Y, F)]^2}{\text{var}(Y) \cdot \text{var}(F)}$$

Morrison (1972) claimed that this measure has an upper bound below 1, whereas Goldberger (1973) showed that the upper bound is one. If we treat (Y_i, F_i) as joint random draws, following Goldberger $E(Y) = E(F)$ and $E(YF) = E(F^2)$. Hence, $\text{cov}(Y, F) = E(Y, F) - E(Y) \cdot E(F) = E(F^2) - E^2(F) = \text{var}(F)$. Thus,

$$(5') \quad R_C^2 = \frac{\text{var}(F)}{\text{var}(Y)}.$$

A version proposed by Lave (1970) is:

$$(5'') \quad R_L^2 = 1 - \frac{\sum_{i=1}^N (Y_i - F_i)^2}{\sum_{i=1}^N (Y_i - \bar{Y})^2},$$

where $\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$. Both R_C^2 and R_L^2 are implemented by replacing F_i with $\hat{F}_i$. Lave (1970, p. 321, footnote 4) also corrects for degrees of freedom (probably the same correction as Theil's corrected R^2), but we do not pursue this. While the estimated versions of (5), (5') and (5'') are in theory not the same for qualitative response models as is true in the standard regression case, the differences are tiny even in very small samples. Our Monte Carlo experiments exhibit no differences at all for sample sizes of 200 and 1000.

Other goodness-of-fit measures are based on the value of the likelihood function L and the likelihood-ratio test statistic

$$(6) \quad \text{LRT} = 2(l_M - l_0),$$

where l_M is the log-likelihood value of the model and l_0 is the log-likelihood value of a model with a constant only. A measure suggested by Aldrich/Nelson (1984) is

$$(7) \quad R_{AN}^2 = \text{LRT} / (\text{LRT} + N),$$

which will always be between zero and one and approximates the relationship between the corresponding likelihood ratio statistic and R^2 in the linear regression model. Equation (7) is also easy to calculate.

Unfortunately, the Aldrich/Nelson measure has an upper bound, which apparently has not been noted. If we set $l_M = 0$, the maximum value for l in (4), the Pseudo- R^2 is $-2l_0 / (N - 2l_0)$. However, $l_0 = N_0 \log(N_0/N) + N_1 \log(N_1/N)$, where N_0 is the number of reported 0's and N_1 is the number of reported 1's of the binary dependent variable. Hence, the limit is

$$(8) \quad R_{AN}^2 (\text{limit}) = -RS / (1 - RS)$$

and $RS = -2l_0/N$. If $N_0/N = 0.5$, the Pseudo- R^2 is 0.581 which is the maximum possible value of the Aldrich/Nelson measure. For $N_0/N = 0.1$ (or 0.9), we obtain

a maximum possible Pseudo- R^2 of 0.394! This suggests a normalization, for which we propose

$$(9) \quad R_{VZ}^2 = \frac{LRT}{LRT + N} \bigg/ \frac{-2l_0}{N-2l} \\ = \frac{-LRT(1 - RS)}{(LRT + N)RS}.$$

One may wish to follow McFadden (1973, p. 121) and choose

$$(10) \quad R_{MF}^2 = 1 - l_M / l_0.$$

This measure approaches 0 if l_M is close to l_0 , and 1 if $l_M = 0$, the maximum possible log-likelihood value. It also is clearly related to the chi-square statistic for testing the hypothesis that the coefficients of the exogenous variables (without the constant) are zero. For these reasons, this measure is recommended in many econometric textbooks like Fomby/Hill/Johnson (1984, p.352), Judge/Griffiths/Hill/Lütkepohl/Lee (1985, p.767), Dhrymes (1986, p.1585), and Greene (1990, p.602). Also Hauser (1977) shows that R_{MF}^2 has a clear meaning in an information-theoretic context in that it measures the percentage of uncertainty in the data which can be explained by the regressors.

A similar concept is (see Maddala, 1983, p. 39)

$$(11) \quad R_M^2 = 1 - (L_0 / L_M)^{2/N},$$

where the upper case L's denote the likelihood as opposed to the log likelihood. As the maximum value for L is 1, the R_M^2 from equation (11) must lie in the range

$$(12) \quad 0 \leq R_M^2 \leq 1 - (L_0)^{2/N}.$$

Cragg/Uhler(1970) have, therefore, suggested a normalization to the [0,1]-range by

$$(13) \quad R_{CU}^2 = \frac{1 - (L_0 / L_M)^{2/N}}{1 - (L_0 / L_{max})^{2/N}}$$

and $L_{max} = 1$.

Finally, McKelvey/Zavoina (1975, p. 111) have suggested an alternative measure which is not commonly used, even though it has attractive properties. They propose a Pseudo- R^2 that gives an estimate of the R^2 of the underlying regression model: namely the portion of the original variance of the dependent variable explained by the probit analysis. Because the error term in equation (1) is

standard normal, an estimate of the sum of squared disturbances is N . The explained sum of squares (EV) can be computed as usual by

$$(14) \quad EV = \sum_{i=1}^N (\hat{Y}_i^* - \bar{Y}^*)^2,$$

where $\hat{Y}_i^*$ is the conditional expectation of Y_i^* from equation (1) based on the probit estimates of equation (3) and

$$(15) \quad \bar{Y}^* = \frac{1}{N} \sum_{i=1}^N Y_i^* = \frac{1}{N} \sum_{i=1}^N \hat{Y}_i^*.$$

From this follows a Pseudo- R^2

$$(16) \quad R_{MZ}^2 = \frac{\sum_{i=1}^N (\hat{Y}_i^* - \bar{Y}^*)^2}{\sum_{i=1}^N (\hat{Y}_i^* - \bar{Y}^*)^2 + N}.$$

The estimated sum of squared disturbances N is not a sum of squared residuals as would be usual in an R^2 measure, but the difference will not affect (16) asymptotically.

All measures should vary in the $[0,1]$ -interval, but little is known about how fast they approach 0 or 1. Consequently, these Pseudo- R^2 's have very different values on the same data and some may not be good estimates of the underlying R^2 . We study this issue in the next section.

3. Experimental design and results

To evaluate the various proposed Pseudo- R^2 's, we carried out a number of Monte Carlo experiments. We compare the Pseudo- R^2 's and attempt to select the one that is "best" able to mimic the R^2 of the continuous case. Continuous data on the endogenous variable are simulated based on a sequence of true models generating R^2 's in the $[0,1]$ -range. We then explore which Pseudo- R^2 , calculated from the probit of the corresponding classified binary data, is (i) closest to R^2 and/or is (ii) the best predictor of R^2 .

The Monte Carlo study consisted of 100 experiments, each with sample size 1000. The design followed equation (1). The X 's were initially generated as standard normal and were thereafter kept fixed. Also, the U 's were drawn from a standard normal distribution for each individual in each sample. We, therefore, generated 100 times 1000 random errors. Given a value for parameter δ , a continuous variable Y_i^* was generated for each individual in the sample, setting $\xi=0$ for simplicity. These continuous data were transformed to binary data according to equation (2). Then 100 probit models were estimated (using Y) as well as the corresponding 100 OLS regressions (using Y^*). Then 100 estimates of R^2 's for the continuous case as well as 100 estimates for each suggested Pseudo- R^2 were available. The experiments were carried out using GAUSS.

To control for the size of R^2 , 21 clusters of these experiments were carried out by varying δ in the $[0,4]$ -range by increments of 0.2. We therefore collected

2100 values of the Pseudo- R^2 's, along with the corresponding values of the estimated "true" R^2 . The sample size of 1000 for each experiment is normal in many probit applications. To check the reliability of the findings, we also repeated the whole procedure with a sample size of 200. We therefore are able to compare results for small and large samples.

Results are analyzed by graphs and OLS regressions. Figures A1-A14, in an Appendix available from the authors on request, contain the outcomes of the Monte Carlo experiments for the small and the large sample for all seven Pseudo- R^2 measures although as mentioned the Lave and correlation coefficient measures are always virtually identical numerically. For each measure, the simulated Pseudo- R^2 has been graphed against the simulated (continuous) reference R^2 . In general, there is a close correspondence between both measures, with a tighter relationship apparent for the larger sample than for the smaller sample.

Figures 1 and 2 summarize these graphical findings by graphing the cubic curves for all Pseudo- R^2 measures against the reference R^2 's for all 21 clusters. It can be seen that the graphs are very similar for the two sample sizes. The McKelvey/Zavoina measure is very close to the 45° line but all the other Pseudo- R^2 's tend to be smaller than the reference R^2 , with the exception of the normalized Aldrich/Nelson measure at the lower end of the range. This is most apparent for the unnormalized Aldrich/Nelson measure (where the effect of its upper bound is seen clearly) but is also obvious for the commonly-used McFadden and Lave/correlation coefficient measures. For the latter measures the difference is largest in the middle of the range. The difference is less for the normalized Aldrich/Nelson measure and the Cragg/Uhler measure and for these the difference tends to be greatest at the upper end of the range.

We have approximated the relationships using simple OLS regressions of reference R^2 's on the Pseudo- R^2 's with a linear, squared and cubed term with parameters α , β and γ , respectively. Results are given in Tables A1 and A2 of the Appendix available on request. The descriptive power of the estimated equations is quite high. Because the Pseudo- R^2 's are commonly small in application (e.g. cross-section data), we also performed these regressions using only the first seven clusters.

Table 1 summarizes the findings by evaluating the predictive quality of all of these measures according to two loss functions. Part 1 of the table gives the average sum of squared errors from the rule of using each Pseudo- R^2 as a direct predictor of the continuous data R^2 . Part 2 gives the average sum of squared errors from the rule of predicting reference R^2 from Pseudo- R^2 using the cubic polynomial regressions just described. All results are given for the small and the large sample size.

In the case of the first loss function, as the graphs would suggest the McKelvey/Zavoina measure clearly places first, followed by Aldrich/Nelson normalized, Cragg/Uhler, the Lave and correlation coefficient measures, McFadden and Aldrich/Nelson. For the full set of experiments in the large sample, the loss function for the Aldrich/Nelson normalized is 14.4 times larger than for McKelvey/Zavoina, with the corresponding ratio for Cragg/Uhler 30.1, for the correlation coefficient or Lave 29.4, for McFadden 196.0, and for Aldrich/Nelson

475.0. The order is almost identical for the small sample and the lower 7 clusters although the differences are smaller. Note that the error ratio is only 1.02 for the Aldrich/Nelson normalized measure for the small sample and 1.5 in the case of the large sample. Given that most applications with cross-sectional data will be in the lower range, this is a useful finding. Finally, it should be noted that for the full range of clusters the McKelvey/Zavoina loss is 6 times smaller for sample size 1000 than for sample size 200. The reduction is considerably larger than in the other cases: the ratio for Aldrich/Nelson normalized is 1.47, for Cragg/Uhler 1.29, for Lave or correlation coefficient 1.16, for McFadden 1.09, and for Aldrich/Nelson 1.05.

Using the cubic regressions for predictions, the large differences disappear, and the ranks are somewhat different: the McKelvey/Zavoina measure still places first (with the exception of the case of the 7 lower clusters and the larger sample where McFadden places first), followed by McFadden, Cragg/Uhler, Aldrich/Nelson, Aldrich/Nelson normalized, and Lave or correlation coefficient. For the full range of clusters the McKelvey/Zavoina loss is 4.98 times smaller for sample size 1000 than for sample size 200 and similar reductions are attained by the other measures, e.g. McFadden 4.93, Cragg/Uhler 4.90, Aldrich/Nelson normalized 4.81, Aldrich/Nelson 4.78, Lave and correlation coefficient 4.75.

Finally, Tables 2-5 summarize the relevant points of the cubic regressions estimated for the four distinct cases (small and large samples, full 21 or 7 lowest clusters). These tables may give applied researchers some idea as to what a given Pseudo- R^2 value may imply about the "true" underlying (continuous) R^2 and therefore facilitates comparisons between studies. For example, for a small sample ($S=200$) and the full range of clusters, an observed McFadden Pseudo- R^2 of 0.4 suggests an underlying R^2 of 0.673. Similarly, an observed Aldrich/Nelson of 0.3 would yield an estimate of 0.548 in the large sample case ($S=1000$).

4. Summary

There is no consensus in the literature as to *which* Pseudo- R^2 should be used in applications of limited-dependent variable models and as a result, different studies use different measures, making comparability difficult. We study this issue by Monte Carlo experimentation for the binary probit case and find that the Aldrich/Nelson, McFadden and correlation coefficient measures tend to be smaller than alternative Pseudo- R^2 's. In recommending a single measure to use in new studies, we argue that a Pseudo- R^2 should also have comparability with OLS- R^2 . Our criterion is therefore that a Pseudo- R^2 should be closely related to the (unobserved) OLS- R^2 for the underlying continuous data. The results are surprisingly clear: the measure suggested by McKelvey/Zavoina performs the best under our criterion. However, in the more likely case of low Pseudo- R^2 's, a normalization of the Aldrich/Nelson measure which we suggest is almost as good as the McKelvey/Zavoina, and is in general easier to calculate. We also show that if the underlying R^2 is predicted using cubic regressions given the Pseudo- R^2 , all measures perform much better. The tables for our cases may give an applied researcher some idea as to what a given Pseudo- R^2 value may imply about the underlying R^2 .

Table 1

Loss Function Evaluation of Pseudo-R's

Measure	N	Loss: <u>Observed</u> Pseudo v.s. Reference			Loss: <u>Predicted</u> Pseudo v.s. Reference		
		<u>SSE</u> N	Error Ratio	Rank	<u>SSE</u> N	Error Ratio	Rank
McKelvey/Zaveina							
S= 200	2100	0.0012	1.0000	1	0.0012	1.0000	1
	700	0.0019	1.0000	1	0.0017	1.0000	1
S= 1000	2100	0.0002	1.0000	1	0.0002	1.0000	1
	700	0.0003	1.0000	1	0.0003	1.0008	2
Cragg/Uhler							
S= 200	2100	0.0093	7.7218	3	0.0012	1.0229	3
	700	0.0045	2.3887	3	0.0017	1.0118	3
S= 1000	2100	0.0072	30.0757	3	0.0002	1.0404	3
	700	0.0026	8.2326	3	0.0003	1.0082	3
McFadden							
S= 200	2100	0.0511	42.3372	6	0.0012	1.0039	2
	700	0.0264	14.0167	6	0.0017	1.0110	2
S= 1000	2100	0.0469	195.9583	6	0.0002	1.0138	2
	700	0.0232	74.4436	7	0.0003	1.0000	1
Lave							
S= 200	2100	0.0355	29.4127	5	0.0013	1.1034	6
	700	0.0164	8.6798	5	0.0018	1.0866	6
S= 1000	2100	0.0307	128.3722	5	0.0003	1.1575	6
	700	0.0131	41.9672	5	0.0003	1.0729	6
Aldrich/Nelson							
S= 200	2100	0.1189	98.5128	7	0.0012	1.0355	4
	700	0.0267	14.1900	7	0.0017	1.0143	4
S= 1000	2100	0.1137	474.9572	7	0.0003	1.0781	5
	700	0.0229	73.7776	6	0.0003	1.0129	4
Aldrich/Nelson normalized							
S= 200	2100	0.0050	4.1897	2	0.0012	1.0367	5
	700	0.0019	1.0196	2	0.0017	1.0158	5
S= 1000	2100	0.0034	14.3700	2	0.0003	1.0741	4
	700	0.0005	1.5131	2	0.0003	1.0136	5
Correlation Coefficient							
S= 200	2100	0.0355	29.3849	4	0.0013	1.1037	7
	700	0.0183	8.6680	4	0.0018	1.0867	7
S= 1000	2100	0.0307	128.3385	4	0.0003	1.1577	7
	700	0.0130	41.9539	4	0.0003	1.0730	7

Ratios are derived from the comparison of each Pseudo's SSE/N to that of the lowest SSE/N of all. Rankings are made accordingly, with the lowest SSE/N ranked highest. S is sample size with S= 200 and S= 1000, respectively. N refers to number of observations in the regressions and is based on 21 clusters of 100 replications each or on the 7 lowest clusters to check for stability of the findings.

Table 2
Estimates of "Underlying" R^2 for a given Pseudo- R^2 :
Small Sample (200) and Full Range of Clusters (21)

Transformation of "Observed" Pseudo- R^2 into estimated
 "Underlying" R^2 for each separate Pseudo- R^2 measure^a

"Observed" Pseudo- R^2 Value	McKelvey/ Zavoina	Cragg/ Uhler	McFadden	Lave	Aldrich/ Nelson	Aldrich/ Nelson normalized	Correlation Coefficient
0.1	0.102	0.116	0.204	0.161	0.163	0.090	0.161
0.2	0.203	0.237	0.385	0.318	0.357	0.192	0.318
0.3	0.305	0.360	0.541	0.468	0.562	0.304	0.468
0.4	0.405	0.482	0.673	0.606	0.760	0.422	0.606
0.5	0.505	0.599	0.780	0.728	0.934	0.541	0.728
0.6	0.604	0.710	0.863	0.828	—	0.659	0.828
0.7	0.702	0.811	0.921	0.903	—	0.772	0.903
0.8	0.798	0.900	0.955	0.949	—	0.876	0.948
0.9	0.893	0.972	0.983	0.960	—	0.968	0.960

^a Therefore, an "observed" McKelvey/Zavoina Pseudo- R^2 of 0.2 would yield 0.203 as an estimate of the "underlying" R^2 . Similarly, an "observed" Lave Pseudo- R^2 of 0.7 would yield 0.903 as an estimate of the "underlying" R^2 .—: Values are declining or out of range.

Table 3
Estimates of "Underlying" R^2 for a given Pseudo- R^2 :
Small Sample (200) and Lower Range of Clusters (7)

Transformation of "Observed" Pseudo- R^2 into estimated
 "Underlying" R^2 for each separate Pseudo- R^2 measure^a

"Observed" Pseudo- R^2 Value	McKelvey/ Zavoina	Cragg/ Uhler	McFadden	Lave	Aldrich/ Nelson	Aldrich/ Nelson normalized	Correlation Coefficient
0.1	0.101	0.120	0.210	0.163	0.170	0.094	0.163
0.2	0.206	0.245	0.392	0.326	0.363	0.200	0.326
0.3	0.310	0.367	0.533	0.470	0.542	0.312	0.470
0.4	0.410	0.479	0.621	0.580	0.672	0.423	0.580
0.5	0.501	0.573	0.643	0.639	0.719	0.526	0.639
0.6	0.579	0.641	—	—	—	0.613	—

^a Therefore, an "observed" McKelvey/Zavoina Pseudo- R^2 of 0.2 would yield 0.206 as an estimate of the "underlying" R^2 . Similarly, an "observed" Lave Pseudo- R^2 of 0.5 would yield 0.639 as an estimate of the "underlying" R^2 .—: Values are declining or out of range. Range is limited to 0.6 as only appropriate to use this table for lower R^2 values.

Table 4
Estimates of "Underlying" R^2 for a given Pseudo- R^2 :
Large Sample (1000) and Full Range of Clusters (21)

Transformation of "Observed" Pseudo- R^2 into estimated
 "Underlying" R^2 for each separate Pseudo- R^2 measure^a

"Observed" Pseudo- R^2 Value	McKelvey/ Zavoina	Cragg/ Uhler	McFadden	Lave	Aldrich/ Nelson	Aldrich/ Nelson normalized	Correlation Coefficient
0.1	0.098	0.112	0.200	0.154	0.157	0.087	0.154
0.2	0.197	0.230	0.376	0.306	0.345	0.186	0.306
0.3	0.298	0.349	0.530	0.453	0.548	0.295	0.453
0.4	0.399	0.469	0.661	0.590	0.749	0.410	0.590
0.5	0.500	0.586	0.769	0.712	0.933	0.528	0.712
0.6	0.601	0.698	0.855	0.816	—	0.647	0.816
0.7	0.701	0.802	0.919	0.896	—	0.762	0.896
0.8	0.800	0.895	0.961	0.949	—	0.871	0.949
0.9	0.898	0.975	0.981	0.970	—	0.970	0.970

^a Therefore, an "observed" McKelvey/Zavoina Pseudo- R^2 of 0.2 would yield 0.197 as an estimate of the "underlying" R^2 . Similarly, an "observed" Lave Pseudo- R^2 of 0.7 would yield 0.896 as an estimate of the "underlying" R^2 . —: Values are declining or out of range.

Table 5
Estimates of "Underlying" R^2 for a given Pseudo- R^2 :
Large Sample (1000) and Lower Range of Clusters (7)

Transformation of "Observed" Pseudo- R^2 into estimated
 "Underlying" R^2 for each separate Pseudo- R^2 measure^a

"Observed" Pseudo- R^2 Value	McKelvey/ Zavoina	Cragg/ Uhler	McFadden	Lave	Aldrich/ Nelson	Aldrich/ Nelson normalized	Correlation Coefficient
0.1	0.097	0.115	0.200	0.153	0.163	0.091	0.153
0.2	0.198	0.233	0.379	0.308	0.346	0.191	0.308
0.3	0.300	0.352	0.528	0.454	0.538	0.298	0.454
0.4	0.401	0.467	0.637	0.580	0.725	0.408	0.580
0.5	0.498	0.576	0.699	0.676	0.892	0.520	0.676
0.6	0.591	0.673	0.704	0.730	—	0.630	0.730

^a Therefore, an "observed" McKelvey/Zavoina Pseudo- R^2 of 0.2 would yield 0.198 as an estimate of the "underlying" R^2 . Similarly, an "observed" Lave Pseudo- R^2 of 0.7 would yield 0.733 as an estimate of the "underlying" R^2 . —: Values are declining or out of range. Range is limited to 0.6 as only appropriate to use this table for lower R^2 values.

References

- ALDRICH, J.H., AND F.D. NELSON (1984): *Linear Probability, Logit, and Probit Models*. Beverly Hills: Sage University Press.
- AMEMIYA, T. (1981): "Qualitative Response Models: A Survey," *Journal of Economic Literature*, 19, 1483-1536.
- CRAGG, J.G., AND R. UHLER (1970): "The Demand for Automobiles," *Canadian Journal of Economics*, 3, 386-406.
- DHRYMES, P.J. (1986), "Limited Dependent Variables," in *Handbook of Econometrics*, Vol. III, ed. by Z. Griliches and M.D. Intriligator. Amsterdam: North-Holland.
- EFRON B. (1978): "Regression and ANOVA with Zero-One Data: Measures of Residual Variation," *Journal of American Statistical Association*, 73, 113-121.
- FOMBY, T.B., R.C. HILL AND S.R. JOHNSON (1984): *Advanced Econometric Methods*. New York: Springer-Verlag.
- GOLDBERGER, A.S. (1973): "Correlations Between Binary Choices and Probabilistic Predictions," *Journal of the American Statistical Association*, 68, 84.
- GREENE, W.H. (1990): *Econometric Analysis*. New York: Macmillan Publishing Company.
- HAUSER, J.R. (1977), "Testing the Accuracy, Usefulness and Significance of Probabilistic Choice Models: An Information Theoretic Approach," *Operations Research*, 26, 406-421.
- JUDGE, G.G., W.E. GRIFFITHS, R.C. HILL, H. LÜTKEPOHL AND T.-C. LEE (1985): *The Theory and Practice of Econometrics*. New York: John Wiley and Sons.
- LAVE, C.A. (1970): "The Demand for Urban Mass Transportation," *Review of Economics and Statistics*, 37, 320-323.
- MADDALA, G.S. (1983): *Limited-Dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press.
- MCFADDEN, D. (1973): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka. New York: Academic Press, 105-142.
- MCKELVEY, R., AND W. ZAVOINA (1975): "A Statistical Model for the Analysis of Ordinal Level Dependent Variables," *Journal of Mathematical Sociology*, 4, 103-120.
- MORRISON, D.G. (1972): "Upper Bounds for Correlation Between Binary Outcomes and Probabilistic Predictions," *Journal of the American Statistical Association*, 67, 68-70.
- NETER, J., AND E.S. MAYNES (1970): "Correlation Coefficient with a 0,1 Dependent Variable," *Journal of the American Statistical Association*, 65, 501-509.

- WHITE, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817-838.
- ZIMMERMANN, K.F. (1986): "Multivariate Micro Prediction Analysis of Qualitative Models," in *Methods of Operations Research 53*, ed. by M.J. Beckmann, K.-W. Gaede, K. Ritter and H. Schneeweiss. Königstein/Taunus: Verlag Anton Hain, 689-696.

Appendix

Figure A1:McKelvey and Zavoina's R^2_{MZ} (equation 16) : Small Sample **Figure A2:**McKelvey and Zavoina's R^2_{MZ} (equation 16) : Large Sample

Figure A3:
Cragg and Uhler's R^2_{CU} (equation 13): Small Sample

Figure A4:
Cragg and Uhler's R^2_{CU} (equation 13): Large Sample

Figure A5:
McFadden's R^2_{MF} (equation 10): Small Sample

Figure A6:
McFadden's R^2_{MF} (equation 10): Large Sample

Figure A7:Lave's R_L^2 (equation 5''): Small Sample **Figure A8:**Lave's R_L^2 (equation 5''): Large Sample

Figure A9:Aldrich and Nelson's R_{AN}^2 (equation 7): Small Sample **Figure A10:**Aldrich and Nelson's R_{AN}^2 (equation 7): Large Sample

Figure A11:

Aldrich and Nelson's R^2 Normalized R_{VZ}^2 (equation 9): Small Sample

Figure A12:

Aldrich and Nelson's R^2 Normalized R_{VZ}^2 (equation 9): Large Sample

Figure A13:

Correlation Coefficient (equation 5): Small Sample

Figure A14:

Correlation Coefficient (equation 5): Large Sample

Table A1

OLS Regressions: Reference R^2 's and Pseudo- R^2 's in the Small Sample Case

Measure	N	α	β	γ	R^2	SEE
McKelvey/Zavoina	2100	1.0178 (43.67)	0.0032 (0.05)	-0.0355 (0.84)	0.9988	0.0342
	700	0.9695 (24.47)	0.4397 (2.54)	-0.7469 (4.13)	0.9915	0.0408
Cragg/Uhler	2100	1.1269 (44.60)	0.3896 (5.19)	-0.4910 (8.84)	0.9988	0.0346
	700	1.1427 (24.49)	0.6636 (2.75)	-1.3136 (4.40)	0.9914	0.0411
McFadden	2100	2.1626 (81.20)	-1.1914 (13.29)	-0.0255 (0.34)	0.9988	0.0343
	700	2.1958 (33.35)	-0.7604 (1.60)	-2.1172 (2.59)	0.9914	0.0411
Lave	2100	1.6089 (58.91)	0.0625 (0.71)	-0.7395 (10.37)	0.9987	0.0359
	700	1.5854 (26.91)	0.7675 (2.01)	-2.7667 (4.67)	0.9905	0.0426
Aldrich/Nelson	2100	1.4209 (32.77)	2.4171 (11.02)	-3.0431 (11.02)	0.9987	0.0348
	700	1.4726 (19.75)	2.8865 (4.66)	-5.9151 (4.81)	0.9913	0.0411
Aldrich/Nelson normalized	2100	0.8225 (32.58)	0.8184 (11.04)	-0.5964 (11.01)	0.9987	0.0348
	700	0.8547 (19.72)	0.9862 (4.62)	-1.1463 (4.77)	0.9913	0.0411
Correlation Coefficient	2100	1.6077 (58.87)	0.0849 (0.73)	-0.7407 (10.39)	0.9987	0.0359
	700	1.5844 (26.91)	0.7686 (2.01)	-2.7650 (4.67)	0.9905	0.0426

Absolute t-values in parentheses are heteroskedasticity-corrected following White (1980). Sample size: 200. Estimates are for 21 clusters (2100 observations) and 7 clusters (700 observations) with 100 replications each. N is number of observations in the regression. SEE is the standard error of the equation.

Table A2

OLS Regressions: Reference R^2 's and Pseudo- R^2 's in the Large Sample Case

Measure	N	α	β	γ	R^2	SEE
McKelvey/Zavoina	2100	0.9899 (94.27)	0.0978 (3.41)	-0.0747 (3.81)	0.9997	0.0153
	700	0.9483 (48.62)	0.2764 (2.95)	-0.3587 (3.31)	0.9984	0.0173
Cragg/Uhler	2100	1.0873 (96.48)	0.3901 (11.31)	-0.4377 (18.84)	0.9997	0.0156
	700	1.1150 (48.76)	0.3817 (2.97)	-0.6189 (3.57)	0.9984	0.0173
McFadden	2100	2.1108 (179.34)	-1.1576 (27.73)	0.0253 (0.69)	0.9997	0.0154
	700	2.0746 (61.95)	-0.6056 (2.25)	-1.4925 (2.88)	0.9984	0.0173
Lave	2100	1.5278 (126.08)	0.1597 (3.87)	-0.7327 (21.21)	0.9997	0.0165
	700	1.4892 (50.59)	0.6173 (2.96)	-1.7845 (5.04)	0.9983	0.0179
Aldrich/Nelson	2100	1.3618 (68.62)	2.3564 (22.98)	-2.6945 (20.66)	0.9997	0.0159
	700	1.4744 (41.32)	1.7360 (5.44)	-2.2345 (3.27)	0.9983	0.0174
Aldrich/Nelson normalized	2100	0.7909 (68.66)	0.7950 (23.02)	-0.5285 (20.73)	0.9997	0.0159
	700	0.8561 (41.30)	0.5860 (5.45)	-0.4384 (3.27)	0.9983	0.0174
Correlation Coefficient	2100	1.5278 (126.05)	0.1600 (3.87)	-0.7327 (21.21)	0.9997	0.0165
	700	1.4892 (50.59)	0.6173 (2.96)	-1.7845 (5.04)	0.9983	0.0179

Absolute t-values in parentheses are heteroskedasticity-corrected following White (1980). Sample size: 1000. Estimates are for 21 clusters (2100 observations) and 7 clusters (700 observations) with 100 replications each. N is number of observations in the regression. SEE is the standard error of the equation.

Discussion Paper Series, CentER, Tilburg University, The Netherlands:

(For previous papers please consult previous discussion papers.)

No.	Author(s)	Title
8935	C. Dang and D. Talman	A New Triangulation of the Unit Simplex for Computing Economic Equilibria
8936	Th. Nijman and M. Verbeek	The Nonresponse Bias in the Analysis of the Determinants of Total Annual Expenditures of Households Based on Panel Data
8937	A.P. Barten	The Estimation of Mixed Demand Systems
8938	G. Marini	Monetary Shocks and the Nominal Interest Rate
8939	W. Güth and E. van Damme	Equilibrium Selection in the Spence Signaling Game
8940	G. Marini and P. Scaramozzino	Monopolistic Competition, Expected Inflation and Contract Length
8941	J.K. Dagsvik	The Generalized Extreme Value Random Utility Model for Continuous Choice
8942	M.F.J. Steel	Weak Exogeneity in Misspecified Sequential Models
8943	A. Roell	Dual Capacity Trading and the Quality of the Market
8944	C. Hsiao	Identification and Estimation of Dichotomous Latent Variables Models Using Panel Data
8945	R.P. Gilles	Equilibrium in a Pure Exchange Economy with an Arbitrary Communication Structure
8946	W.B. MacLeod and J.M. Malcomson	Efficient Specific Investments, Incomplete Contracts, and the Role of Market Alternatives
8947	A. van Soest and A. Kapteyn	The Impact of Minimum Wage Regulations on Employment and the Wage Rate Distribution
8948	P. Kooreman and B. Melenberg	Maximum Score Estimation in the Ordered Response Model
8949	C. Dang	The D_3 -Triangulation for Simplicial Deformation Algorithms for Computing Solutions of Nonlinear Equations
8950	M. Cripps	Dealer Behaviour and Price Volatility in Asset Markets
8951	T. Wansbeek and A. Kapteyn	Simple Estimators for Dynamic Panel Data Models with Errors in Variables

No.	Author(s)	Title
8952	Y. Dai, G. van der Laan, D. Talman and Y. Yamamoto	A Simplicial Algorithm for the Nonlinear Stationary Point Problem on an Unbounded Polyhedron
8953	F. van der Ploeg	Risk Aversion, Intertemporal Substitution and Consumption: The CARA-LQ Problem
8954	A. Kapteyn, S. van de Geer, H. van de Stadt and T. Wansbeek	Interdependent Preferences: An Econometric Analysis
8955	L. Zou	Ownership Structure and Efficiency: An Incentive Mechanism Approach
8956	P. Kooreman and A. Kapteyn	On the Empirical Implementation of Some Game Theoretic Models of Household Labor Supply
8957	E. van Damme	Signaling and Forward Induction in a Market Entry Context
9001	A. van Soest, P. Kooreman and A. Kapteyn	Coherency and Regularity of Demand Systems with Equality and Inequality Constraints
9002	J.R. Magnus and B. Pesaran	Forecasting, Misspecification and Unit Roots: The Case of AR(1) Versus ARMA(1,1)
9003	J. Driffill and C. Schultz	Wage Setting and Stabilization Policy in a Game with Renegotiation
9004	M. McAleer, M.H. Pesaran and A. Bera	Alternative Approaches to Testing Non-Nested Models with Autocorrelated Disturbances: An Application to Models of U.S. Unemployment
9005	Th. ten Raa and M.F.J. Steel	A Stochastic Analysis of an Input-Output Model: Comment
9006	M. McAleer and C.R. McKenzie	Keynesian and New Classical Models of Unemployment Revisited
9007	J. Osiewalski and M.F.J. Steel	Semi-Conjugate Prior Densities in Multivariate t Regression Models
9008	G.W. Imbens	Duration Models with Time-Varying Coefficients
9009	G.W. Imbens	An Efficient Method of Moments Estimator for Discrete Choice Models with Choice-Based Sampling
9010	P. Deschamps	Expectations and Intertemporal Separability in an Empirical Model of Consumption and Investment under Uncertainty

No.	Author(s)	Title
9011	W. Güth and E. van Damme	Gorby Games - A Game Theoretic Analysis of Disarmament Campaigns and the Defense Efficiency-Hypothesis
9012	A. Horsley and A. Wrobel	The Existence of an Equilibrium Density for Marginal Cost Prices, and the Solution to the Shifting-Peak Problem
9013	A. Horsley and A. Wrobel	The Closedness of the Free-Disposal Hull of a Production Set
9014	A. Horsley and A. Wrobel	The Continuity of the Equilibrium Price Density: The Case of Symmetric Joint Costs, and a Solution to the Shifting-Pattern Problem
9015	A. van den Elzen, G. van der Laan and D. Talman	An Adjustment Process for an Exchange Economy with Linear Production Technologies
9016	P. Deschamps	On Fractional Demand Systems and Budget Share Positivity
9017	B.J. Christensen and N.M. Kiefer	The Exact Likelihood Function for an Empirical Job Search Model
9018	M. Verbeek and Th. Nijman	Testing for Selectivity Bias in Panel Data Models
9019	J.R. Magnus and B. Pesaran	Evaluation of Moments of Ratios of Quadratic Forms in Normal Variables and Related Statistics
9020	A. Robson	Status, the Distribution of Wealth, Social and Private Attitudes to Risk
9021	J.R. Magnus and B. Pesaran	Evaluation of Moments of Quadratic Forms in Normal Variables
9022	K. Kamiya and D. Talman	Linear Stationary Point Problems
9023	W. Emons	Good Times, Bad Times, and Vertical Upstream Integration
9024	C. Dang	The D_2 -Triangulation for Simplicial Homotopy Algorithms for Computing Solutions of Nonlinear Equations
9025	K. Kamiya and D. Talman	Variable Dimension Simplicial Algorithm for Balanced Games

No.	Author(s)	Title
9026	P. Skott	Efficiency Wages, Mark-Up Pricing and Effective Demand
9027	C. Dang and D. Talman	The D_1 -Triangulation in Simplicial Variable Dimension Algorithms for Computing Solutions of Nonlinear Equations
9028	J. Bai, A.J. Jakeman and M. McAleer	Discrimination Between Nested Two- and Three-Parameter Distributions: An Application to Models of Air Pollution
9029	Th. van de Klundert	Crowding out and the Wealth of Nations
9030	Th. van de Klundert and R. Gradus	Optimal Government Debt under Distortionary Taxation
9031	A. Weber	The Credibility of Monetary Target Announcements: An Empirical Evaluation
9032	J. Osiewalski and M. Steel	Robust Bayesian Inference in Elliptical Regression Models
9033	C. R. Wichers	The Linear-Algebraic Structure of Least Squares
9034	C. de Vries	On the Relation between GARCH and Stable Processes
9035	M.R. Baye, D.W. Jansen and Q. Li	Aggregation and the "Random Objective" Justification for Disturbances in Complete Demand Systems
9036	J. Driffill	The Term Structure of Interest Rates: Structural Stability and Macroeconomic Policy Changes in the UK
9037	F. van der Ploeg	Budgetary Aspects of Economic and Monetary Integration in Europe
9038	A. Robson	Existence of Nash Equilibrium in Mixed Strategies for Games where Payoffs Need not Be Continuous in Pure Strategies
9039	A. Robson	An "Informationally Robust Equilibrium" for Two-Person Nonzero-Sum Games
9040	M.R. Baye, G. Tian and J. Zhou	The Existence of Pure-Strategy Nash Equilibrium in Games with Payoffs that are not Quasiconcave
9041	M. Burnovsky and I. Zang	"Costless" Indirect Regulation of Monopolies with Substantial Entry Cost
9042	P.J. Deschamps	Joint Tests for Regularity and Autocorrelation in Allocation Systems

No.	Author(s)	Title
9043	S. Chib, J. Osiewalski and M. Steel	Posterior Inference on the Degrees of Freedom Parameter in Multivariate-t Regression Models
9044	H.A. Keuzenkamp	The Probability Approach in Economic Method- ology: On the Relation between Haavelmo's Legacy and the Methodology of Economics
9045	I.M. Bomze and E.E.C. van Damme	A Dynamical Characterization of Evolution- arily Stable States
9046	E. van Damme	On Dominance Solvable Games and Equilibrium Selection Theories
9047	J. Driffill	Changes in Regime and the Term Structure: A Note
9048	A.J.J. Talman	General Equilibrium Programming
9049	H.A. Keuzenkamp and F. van der Ploeg	Saving, Investment, Government Finance and the Current Account: The Dutch Experience
9050	C. Dang and A.J.J. Talman	The D_1 -Triangulation in Simplicial Variable Dimension Algorithms on the Unit Simplex for Computing Fixed Points
9051	M. Baye, D. Kovenock and C. de Vries	The All-Pay Auction with Complete Information
9052	H. Carlsson and E. van Damme	Global Games and Equilibrium Selection
9053	M. Baye and D. Kovenock	How to Sell a Pickup Truck: "Beat-or-Pay" Advertisements as Facilitating Devices
9054	Th. van de Klundert	The Ultimate Consequences of the New Growth Theory; An Introduction to the Views of M. Fitzgerald Scott
9055	P. Kooreman	Nonparametric Bounds on the Regression Coefficients when an Explanatory Variable is Categorized
9056	R. Bartels and D.G. Fiebig	Integrating Direct Metering and Conditional Demand Analysis for Estimating End-Use Loads
9057	M.R. Veall and K.F. Zimmermann	Evaluating Pseudo- R^2 's for Binary Probit Models

PO. BOX 90153 5000 LE TILBURG THE NETHERLANDS

Bibliotheek K. U. Brabant

17 000 01117540 4