Veall, Zimmermann:

Pseudo-R2 Measures for Some Common Limited Dependent Variable Models

Sonderforschungsbereich 386, Paper 18 (1996)

Online unter: http://epub.ub.uni-muenchen.de/

Pseudo-R^2 Measures for Some Common Limited Dependent Variable Models

Michael R. Veall
Department of Economics
McMaster University
Hamilton, Ontario, Canada

Klaus F. Zimmermann
SELAPO, University of Munich,
Munich, Germany
and
CEPR, London

January, 1996

ABSTRACT: A large number of different Pseudo-R^2 measures for some common limited dependent variable models are surveyed. Measures include those based solely on the maximized likelihoods with and without the restriction that slope coefficients are zero, those which require further calculations based on parameter estimates of the coefficients and variances and those that are based solely on whether the qualitative predictions of the model are correct or not. The theme of the survey is that while there is no obvious criterion for choosing which Pseudo-R^2 to use, if the estimation is in the context of an underlying latent dependent variable model, a case can be made for basing the choice on the strength of the numerical relationship to the OLS-R^2 in the latent dependent variable. As such an OLS-R^2 can be known in a Monte Carlo simulation, we summarize Monte Carlo results for some important latent dependent variable models (binary probit, ordinal probit and Tobit) and find that a Pseudo-R^2 measure due to McKelvey and Zavoina scores consistently well under our criterion. We also very briefly discuss Pseudo-R^2 measures for count data, for duration models and for prediction-realization tables.

Acknowledgements: We thank John DeNew for excellent research assistance. Useful comments were provided by Deb Fretz, Colin Cameron, Frank Windmeijer, the participants at seminars at McGill and McMaster University, three anonymous referees and the editor. The authors are responsible for all errors and shortcomings. The former author thanks the Alexander von Humboldt Foundation and the Social Sciences and Humanities Research Council of Canada. The second author acknowledges support by the German Science Foundation.

1. Introduction

This survey reviews some of the many R^2-type measures (or Pseudo-R^2's) that have been proposed for estimated limited dependent variable models. (A limited dependent variable model is a model where the observed dependent variable is constrained, such as in the binary probit model where it must be either zero or one, or in the Tobit model, where it is constrained to exceed zero.) The surveys of limited dependent variable models by Amemiya (1981) and Dhrymes (1986), as well as the standard reference by Maddala (1983), all briefly discuss goodness of fit and mention one or two possible Pseudo-R^2's, but none give a motivation as to why such measures might be calculated. (Some of the measures were initially introduced with little or no justification and in some cases are hard to motivate now.) In the next section, we discuss a number of possible motivations and show how each leads to a class of Pseudo-R^2's. While none of these is beyond challenge, we emphasize one, the McKelvey-Zavoina R^2, that in some situations seems most conducive to comparability across different types of empirical models.

As an example of this kind of comparability, consider a situation where the researcher is estimating a model with annual individual income as the dependent variable and some number of independent variables. The data, which are by individual, may be provided in three ways: (i) the complete data (ii) the complete data except that the dependent variable data is censored at $50,000 (so that one knows which individuals are earning more than $50,000 but not how much more) or (iii) the complete data except that the dependent variable is only reported by category, such as where a zero corresponds to "less than or equal to $50,000" and a one corresponds to "more than $50,000" in the binary category case. Mode (ii) or (iii) might perhaps be adopted due to confidentiality concerns. It might be desirable if the R^2 from OLS on sample (i) were as close as possible to the Pseudo-R^2 from a Tobit type regression if the data were provided as sample (ii) or the Pseudo-R^2 from a binary probit regression if the data were provided as sample (iii). This same kind of comparability might be used more generally to make rough comparisons across empirical models, where in some cases the dependent variable is observed continuously and in others it is limited.

We shall discuss other types of justifications in the next section, and while it is not the purpose of this survey to convince the reader that our favoured justification, or any other, is the "right" one, we do note that the above approach is consistent with the way practitioners use R^2 in the OLS context. Our view is that most empirical researchers are explicitly or implicitly making rough comparisons of "goodness of fit" across similar empirical models with similar samples, where the research experience in the area is far more important than any statistical criteria. For example, a researcher estimating macroeconometric OLS regressions using data from different countries might expect R^2's in the .8 or .9 range. If one of the country regressions has an R^2 of .4, this is a sign that special attention is required; there may even be an error. However in a different situation, practitioners using microdata on labour supply may expect R^2's of around .1. A regression with an R^2 of .02 might require further scrutiny while an R^2 of .4 would be suspiciously large. Our favoured approach is simply to choose a Pseudo-R^2 in the limited dependent variable context that will be as comparable as possible with the accumulated experience from R^2 in OLS regression. Emphasis on comparability also leads to another important theme of this survey: using the same data and the same model, there can be large numerical differences between different measures, even in large samples. For example, we shall discuss entirely typical cases with 1000 observations where one Pseudo-R^2, the McFadden R^2, will be about .25 while another, the McKelvey Zavoina R^2, will be about .5.

Unfortunately some confusion has arisen as to whether this R^2 should be calculated conditionally or unconditionally upon the realized discrete outcomes. For example, the manual of the popular computer package LIMDEP (Greene, 1995) only proposes one Pseudo-R^2, the McKelvey-Zavoina measure, but calculates it conditionally upon the discrete outcomes. We argue that this leads to a seriously biased measure and the unconditional fitted values should be used instead, a modification that is easily made.

Section 2 of this survey extends the introduction and discusses possible motivations for the use of a Pseudo-R^2 and how each reason leads to a particular class of Pseudo-R^2's. Section 3 shows how each type of Pseudo-R^2 applies to the binary dependent variable case and discusses the various Pseudo-R^2's and their performance according to various critera. Section 4 considers cases where the discrete dependent variable may take more than two values. In Section 5 we discuss the case where the dependent variable is continuous but limited, such as the Tobit model. Section 6 discusses Pseudo-R^2 measures based only on the prediction/realization table. Section 7 summarizes and concludes.

2. Motivation and Criteria for Pseudo-R^2's

R^2 measures cannot be used for diagnostic tests of the basic assumptions of the model, either in continuous or limited dependent variable contexts. (Pagan and Vella (1989), Smith and Peters (1990) and the papers in the special issue of the Journal of Econometrics edited by Blundell and summarized in Blundell (1987) all discuss diagnostics that apply to limited dependent variable models.) Nonetheless one is tempted to conclude that R^2 measures must have some use in econometrics, if only because they are so widely reported in the OLS case and almost as frequently so in the limited dependent variable case. There is a certain irony in that Pseudo-R^2 measures are seldom justified and commonly reported yet limited dependent variable diagnostics have well-known importance but are seldom reported, the latter a "sorry state of affairs" as Pagan and Vella note (1989, p. 530).

In the Introduction, we sketched a brief overall motivation for choosing Pseudo-R^2 measures that would maximize comparability across similar empirical models, some with continuous and some with limited dependent variables. To consider the basis of that comparability, consider the three properties given by Dhrymes (1986) for R^2 in the OLS case that he feels could be desirably extended to a Pseudo-R^2:

i. it stands in a one-to-one relation to the F-statistic for testing the hypothesis that the coefficients of the bona fide explanatory variables are zero;

ii. it is a measure of the reduction of the variability of the dependent variable through the bona fide explanatory variables;

iii. it is the square of the simple correlation coefficient between predicted and actual values of the dependent variable within the sample.

(Kvålseth (1985) gives a more complete set of interpretations attributable to OLS-R^2 under the assumption the model contains an intercept.) As Dhrymes (1986) notes, no single Pseudo-R^2 has all three properties. However, while the three properties are highly related, each can be used as a motivation for a class of Pseudo-R^2 measures.

Property (i), corresponding to what Magee (1990) calls the "significance of fit approach" is based on the OLS relationship:

(1) $R^2 = (k-1)F/((k-1)F + N - k)$

where k is the number of explanatory variables including the intercept, F is the F-statistic of the null hypothesis that the non-intercept variables are zero and N is the number of observations. There are similar relationships involving, instead of F, the likelihood ratio or other chi-square statistics of the same null hypothesis. Magee points out that a large class of R^2-type measures can be created by simply exporting these relationships to any estimation context so that an R^2 measure can be created for a binary probit, for example, by simply calculating the appropriate F-statistic and using formula (1). This means there is an R^2 measure available any time there are estimated coefficients and an estimated variance-covariance matrix. In addition, the McFadden R^2, probably the most commonly used Pseudo-R^2, can be related to this framework and it has a separate (Kullback-Leibler) information theoretic justification due to Hauser (1977), an approach recently considered by Cameron and Windmeijer (1993b) in their generalization of the McFadden measure to cover a wider variety of situations. We shall discuss this further next section.

The significance-of-fit approach has the great advantage of being able to provide an R^2 in almost any situation, including other contexts that have nothing to do with limited dependent variables. The main problem with using property (i) boils down to a single question which we cannot answer satisfactorily: if such a one-to-one relationship is the basis of the use of R^2, why not just use the F-statistic itself or its prob value?

Unlike property (i), properties (ii) and (iii) from Dhrymes (1986) do not seem to lead to R^2-type measures for very general contexts, but both seem to lead to useful approaches in some limited dependent variable contexts. Limited dependent variables are normally modelled as functions of underlying continuous variables that are not observed (as in the example in the Introduction, where in the binary probit case the continuous variable income is reduced to a (0,1) variable). Property (iii) corresponds most closely to what we shall call the "correlation approach". This yields the class of measures that use correlation coefficients between the actual outcomes of the discrete dependent variables and their "predictions", which will be estimated probabilities from the underlying continous dependent variable model. (Section 6 discusses the case where predictions must be zero/one and are not probabilities.) Hence these measures may be appropriate if the implicit loss function is in the difference between the outcome and the estimated probability that that outcome will occur, with the R^2 measure the estimated predictive gain from using the explanatory variables.

A measure based on (iii) and the correlation of actual values and their predicted probabilities could not be expected to be comparable to an R^2 for the OLS case (where the implicit loss is in the difference between an outcome measured on a continuous scale and its prediction on that same scale). However the loss function could be in the unexplained variability of the underlying continuous dependent variable and hence, adapting property (ii), a Pseudo-R^2 could be a measure of the reduction of the variability of the latent dependent variable through the bona fide explanatory variables. While the disadvantage of this "explained variation" approach is that the Pseudo-R^2 measure becomes rooted in an unobservable, it may nonetheless be of value to have an estimate of goodness-of-fit for the latent continous variable. (Indeed in cases such as in the Introduction example with the categorical data, the continuous but unobserved variable income is the target of the analysis.) A clear advantage is the kind of comparability of R^2 measures across models estimated by different techniques as described in the Introduction.

The criterion that the Pseudo-R^2 be as close as possible to what OLS-R^2 would be on the underlying latent variable model has been our favoured criterion (Veall and Zimmermann, 1990a, 1990b) but has also been used as a criterion by Hagle and Mitchell (1992), Laitila (1993) and Windmeijer (1995) in their studies of particular Pseudo-R^2's. A danger is that the unobserved latent variable in the limited dependent variable case may not be comparable with the continuous dependent variable. To revisit our Introduction example a final time, in the binary probit case the latent variable underlying the (0,1) variable might not be income but instead any monotonic transformation of income that led to a model linear in the explanatory variables with a normal disturbance. Nonetheless if these assumptions hold where they can be tested in the "comparable" continuous dependent variable cases, it may be reasonable to assume that they hold in the limited dependent variable cases as well.

Before we turn to actual measures in the next section, two of the more mundane criteria should be mentioned. One is that an R^2 measure is typically bounded between zero and one and this is common to almost all the measures we study. (None may exceed one; we shall indicate the few which some may under some circumstances be less than zero.) The second is that R^2 should tend to increase (or at least not decrease) as more explanatory variables are added, a property of all the measures we shall discuss.

3. Pseudo-R^2´s for the Case of a Binary Dependent Variable

Suppose the dependent variable holds only two values: e.g. either 0 or 1, as commonly assumed in the binary logit or binary probit model. A typical approach postulates an underlying continuous variable Y_i^* :

(2) $Y_i^* = x_i' \beta + U_i$, i = 1, 2,..., N

where x_i' is a row vector of the values of the explanatory variables at observation i, including a one for an intercept term, β is a vector of parameters and U_i is a random error term, typically assumed to be independently and identically distributed. We assume that Y_i^* is not observed but instead we observe

(3) $Y_i = 1$ if $Y_i^* > 0$ and $Y_i = 0$ otherwise

As is well known, (2) and (3) imply a log-likelihood function

(4) $$\ell = \sum_{i=1}^{N} Y_i \log\left[1 - H(-x_i'\beta)\right] + (1 - Y_i)\log H(-x_i'\beta)$$

where H is the cumulative distribution for U. If H is standard normal, the model is called a binary probit. If H is logistic, the model is called a binary logit. For future reference we define

(5) $$\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$$

and

(6) $$\overline{Y}^* = \frac{1}{N}\sum_{i=1}^{N} \hat{Y}_i^*$$

where $\hat{Y}_i^* = x_i'\hat{\beta}$ evaluated at the maximum likelihood estimates based on (4).

Amemiya (1981) surveys such models and proposes some goodness-of-fit measures in R^2 form. We have added a few to his list and grouped them according to Dhrymes`s (1986) interpretations (i) - (iii). While it is possible to estimate such models by OLS in some cases as an approximation to probit or logit regression, Cox and Wermuth (1992) point out that in any case where this is feasible, R^2 is restricted from above and unlikely to be useful. Hence we focus on probit and logit methods.

Dhrymes's interpretation (i) suggests generating a Pseudo-R^2 using a one-to-one relationship to the F-statistic (or some similar statistic) for testing the hypothesis that the coefficients of the explanatory variables, besides the intercept, are zero. Magee (1990) suggests using formula (1) as a possible rule for generating a Pseudo-R^2 in a wide variety of circumstances. As he points out, it is also possible (and more common) in this context to use the corresponding likelihood ratio statistic:

(7) $LRT = 2(\ell_M - \ell_0)$,

where ℓ_M is the log-likelihood value of the model and ℓ_0 is the log-likelihood value if the non-intercept coefficients are restricted to zero. It is also helpful to define

(8) $LRT^* = 2(\ell_{MAX} - \ell_0)$

where ℓ_{MAX} is the maximum possible likelihood (i.e. a perfect fit) and in this case is 0. Some possible Pseudo-R^2 measures based on F and LRT are contained in Table 1.

Turning to the table, it should be clear that all the measures lie between zero and 1. All the likelihood based measures cannot fall as right hand side variables are added to the model; the others may fall but the probability of this happening vanishes as N increases. Considering the Significance-of-fit Class first, Magee's R^2_{MA} has already been discussed, while as Magee points out, Aldrich and Nelson's R^2_{AN} can be seen as being based on the OLS relationship between R^2 and the Wald statistic (which equals N multiplied by the ratio of the explained to unexplained sums of squares, given the model contains a constant), except that the likelihood ratio statistic is used instead of the Wald statistic. Veall and Zimmermann (1990a, 1992a) point out that R^2_{AN} has an upper bound far less than one. For example the maximum value of this measure is .581, occuring when the observed dependent variable is zero and one in exactly equal proportions. If the proportion of zeroes is either .1 or .9, the bound shrinks to .394. Veall and Zimmermann propose R^2_{ANN}, the Aldrich Nelson measure normalized, which has upper bound one whenever the observed dependent variable is discrete.

Table 1

Pseudo-R^2 Measures in the Binary Dependent Variable Case

Measure	Reference
<u>Significance of Fit Class</u>	
$R^2_{MA} = \frac{(k-1)F}{(k-1)F+N-k}$	Magee (1990)
R^2_{AN} = LRT/(LRT + N)	Aldrich and Nelson (1984)
$R^2_{ANN} = \frac{LRT}{LRT+N} / \frac{LRT^*}{LRT^*+N} = \frac{LRT}{LRT+N} / \frac{-2l_0}{N-2l_0}$	Veall and Zimmermann (1990a, 1992a)
$R^2_{MF} = \frac{LRT}{LRT^*} = \frac{(\ell_M - \ell_0)}{(\ell_{MAX} - \ell_0)} = 1 - \frac{\ell_M}{\ell_0}$	McFadden (1973, p. 121)
$R^2_M = 1 - \exp(-LRT/N)$	Maddala (1983, p. 39)
$R^2_{CU} = \frac{1 - \exp(-LRT/N)}{1 - \exp(-LRT^*/N)}$	Cragg and Uhler (1970)
<u>Explained Variation Class</u>	
$R^2_{MZ} = \frac{\sum_{i=1}^{N} (\hat{Y}_i^* - \bar{Y}^*)^2}{\sum_{i=1}^{N} (\hat{Y}_i^* - \bar{Y}^*)^2 + N\hat{\sigma}^2}$	McKelvey and Zavoina (1975)

Table 1 Cont'd

<u>Correlation Class</u>	
$R_C^2 = \frac{[\mathrm{cov}(Y,H)]^2}{\mathrm{var}(Y) \cdot \mathrm{var}(H)} = \frac{\mathrm{var}(H)}{\mathrm{var}(Y)}$	Neter and Maynes (1970), Morrison (1972), Goldberger (1973) and Efron (1978)
$R_L^2 = 1 - \frac{\sum_{i=1}^{N} (Y_i - H_i)^2}{\sum_{i=1}^{N} (Y_i - \bar{Y})^2}$	Lave (1970)

The McFadden R^2_{MF} is probably the most popular Pseudo-R^2 and for example is the only Pseudo-R^2 provided by the computer package STATA (1995). The various expressions for it in Table 1 show how it is the achieved gain in the log-likelihood due to the explanatory variables relative to the maximum possible achievable gain, where the third equality follows as $\ell_{MAX} = 0$ in logit or probit models. It is sometimes reported in these contexts as the "likelihood ratio index". Hauser (1977) discusses this measure in a Kullback-Leibler divergence information-theoretic context. Merkle and Zimmermann (1992) and Cameron and Windmeijer (1993b) emphasize that the first two expressions for this Pseudo-R^2 in the Table can be used in other situations where ℓ_{MAX} may not necessarily be zero or one. These authors also point out that this measure can be seen as being based on the "deviance decomposition" in the same way that R^2 in OLS can be based on the variance decomposition of the total sum of squares into the explained and unexplained sum of squares. (The deviance decomposition decomposes ℓ_{MAX} - ℓ_0, total achievable likelihood gain starting from the constant-only model into the explained portion, $\ell_M - \ell_0$, and the unexplained portion, $\ell_{MAX} - \ell_M$.)

As Magee (1990) makes clear, the Maddala R^2_M is based on the OLS formula for R^2 in terms of LRT, assuming a normal likelihood. However it has an upper bound less than one; the last measure on the first panel of Table 1, the Cragg and Uhler R^2_{CU}, is the Maddala measure normalized so that its upper bound is one in any discrete dependent variable case such as this one.

Turning to the second panel of Table 1, there is a single measure in what we call the <u>Explained Variation Class</u>. In the McKelvey and Zavoina R^2_{MZ}, the numerator (which is also the first term of the denominator) is an estimate of what the explained sum of squares would be based on the conditional expectation of the latent variable. $N\hat{\sigma}^2$ is an estimate of the unexplained variation so the measure can be interpreted as an estimate of the explained sum of squares divided by an estimate of the sum of the explained and unexplained sum of squares. Note that in discrete dependent variable models, $\hat{\sigma}^2$ is not estimated but is set by normalization, for example to one in the binary probit case and $\pi^2/6$ in the binary logit case.

In calculating R^2_{MZ}, it is important that we use $\hat{Y}_i^* = x_i'\hat{\beta}$ and <u>not</u> condition on the realized value of Y_i which would mean using $Y_{i/Y} = x_i'\hat{\beta} + \hat{\lambda}_i$, where $\hat{\lambda}_i$ is the inverse Mill's ratio. The manual for the computer package LIMDEP (Greene, 1995) proposes the latter approach. It can be shown that this results in a serious upward bias in R^2_{MZ}. For example, we have shown theoretically (and confirmed by simulation) that if $\hat{\beta} = 0$, the LIMDEP version of R^2_{MZ} will be almost .4 when the true $R^2_{MZ} = 0$. The bias is smaller for nonzero $\hat{\beta}$ but remains serious. The correct R^2_{MZ} is easily calculated in LIMDEP: simply omit "+ LAMBDA" in the CREATE statement on p. 421 (and the "Hold" in the PROBIT statement is unnecessary).

Now consider the third panel of Table 1. For the first measure in the <u>Correlation Class</u>, the correlation coefficient R^2_C, the equality is established (following Goldberger, 1973) by noting that if we treat (Y_i, H_i) as joint random draws (with H_i the value of the cumulative distribution function for observation i), E(Y) = E(H) and E(YH) = E(H^2). Hence cov(Y,H) = E(YH) - E(Y)E(H) = E(H^2) - $(E(H))^2$ = var(H). The second measure, Lave's R^2_L, is based on a decomposition using similar rules. Both are implemented using the sample Y_i and the estimated values of H_i from the model. While these two measures can be different, empirically the differences are tiny even in very small samples. Experiments in Veall and Zimmermann (1990a) exhibit such small numerical differences for sample sizes of 200 and 1000, that one line on a graph does for both measures (as it will in this paper as well).

Veall and Zimmermann (1990. 1992a, 1994a), Hagle and Mitchell (1992) and Windmeijer (1995) all investigate the properties of various Pseudo-R^2's using Monte Carlo experiments. All three conduct simulations to determine how closely most of the Pseudo-R^2's correspond to the OLS-R^2 on the underlying latent variable model. We reproduce the results of one of the Veall and Zimmermann experiments here. The latent variable model consists of an intercept and one standard normal explanatory variable, with the true intercept coefficient set at zero and the slope coefficient set at 21 different values to move the underlying R^2 through the range from zero to one. Simulated Y_i^*'s are converted to simulated observations Y_i using (3). At each of the 21 settings 100 experiments were conducted, with binary probit models estimated on each data set. Figure 1 graphs the average OLS or "Reference" R^2 against the average or "Predicted" Pseudo-R^2, for the 1000 observations case. The graph gives some idea as to how to compare different measures: for example it can be seen that an OLS-R^2 of .5 on the latent variable model corresponds to an Aldrich-Nelson or McFadden R^2 each of about .25 and a normalized Aldrich-Nelson or McKelvey-Zavoina R^2 each of about .5. The McKelvey-Zavoina line is very close to the 45^0 line, indicating that it estimates the underlying OLS-R^2 without bias, as one might expect from its formulation which is designed to estimate the latent variable R^2. Of the other measures, the likelihood-based normalized Aldrich-Nelson R^2 does best under this criterion and then in increasing order of downward bias, the Cragg-Uhler R^2 and, in a tie, R^2_L and R^2_C. The downward bias of the Aldrich-Nelson R^2 and the McFadden R^2 is greater still. Veall and Zimmermann (1990a) also include scatter diagrams and cubic polynomial regressions of the OLS-R^2 as a function of the various Pseudo-R^2's which show that the degree of variability in all these measures is very close and that with the right polynomial transformation, all could be used to estimate the underlying OLS-R^2 very accurately. Although R^2_{MZ} still provided the most accurate estimate, its real advantage under this criterion is simply the convenience that its relationship with the latent variable OLS-R^2 is so close to the 45^0 line.

The research of Hagle and Mitchell (1992) and Windmeijer (1995) is entirely consistent with these findings and adds the following insights:

(a) While Veall and Zimmermann (1990a) find that the choice of sample size of either 200 or 1000 makes little difference, Hagle and Mitchell (1992) find that with a sample size of only 100, the sample variance of R^2_{MZ} is somewhat larger than that of R^2_{ANN}.

(b) Windmeijer (1995) finds that R^2_{MZ} and R^2_{MF} are the only measures relatively insensitive to changes in the value of the intercept of the underlying latent model, which can change the proportion of zeros and ones observed in Y.

(c) Windmeijer (1995) also finds that R^2_{MZ} scores best (although not very well) using the criterion of closeness to the squared correlation of the actual and predicted probabilities. (We have our objections to this criterion. See Veall and Zimmermann (1995). It is true that if the squared correlation of the actual and predicted probabilities is one, this could indicate the correct model had been chosen; on the other hand this correlation could be one even for models that fit very poorly as it makes no allowance for unexplained variation that may be left. For example, the correlation could be one even if the model omits a normally distributed variable that is orthogonal to the others.)

(d) Hagle and Mitchell (1992) also consider the case of misspecification, where the method of estimation does not match the probability distribution of the errors. If the binary logit method is applied instead of the binary probit even though the true disturbance is normal, there is almost no consequence in terms of the performance of the Pseudo-R^2's. However if either probit or logit analysis is done when the true error is skewed or bimodal, the effects on the R^2 measures are large, with the results favouring the choice of R^2_{MF}.

(e) Windmeijer (1995) also emphasizes misspecification in the choice of right hand side variables (and hence the issue of selection of those variables). He sets:

(9) $Y_i^* = \alpha + x_{1i} + x_{2i} + x_{3i} + x_{4i} + x_{5i} - \epsilon_i$

with various ways of generating the different x´s and also a nonincluded variable x_6. All the measures are calculated for the simple regression for Y_i with an intercept and x_1, calculated again with x_2 added to the regression, again with x_3 added and so on up to x_6. The plot of the various measures as variables are added shows that R^2_{MZ} is closest to the underlying OLS-R^2 and hence might be the most useful in model selection.

4. Pseudo-R^2's with Discrete Dependent Variables with More Than Two Outcomes

Sometime models with discrete dependent variables have more than two outcomes. These models include ordinal probit and ordinal logit, where the outcomes are ordered (e.g. no employment, part-time employment, full-time employment) and multinomial probit and logit models, where there is no such ordering (e.g. choice of heating by gas, oil or electricity). For the unordered approaches, only the significance-of-fit measures apply and there is no research on which Pseudo-R^2 is best. Maddala (1983) describes the overall multinomial probit/logit model and with respect to Pseudo-R^2, the approach of Magee will work and, as noted, Hauser (1977) and Cameron and Windmeijer (1993b) emphasize the information theoretic support for R^2_{MF}.) For the ordered approaches, R^2_{MZ} is available but the correlation approach is not usually applied.

Veall and Zimmermann (1992a) conduct a Monte Carlo experiment for the example of an ordinal probit along the lines of the binary probit Monte Carlo analysis described in the previous section. Ordinal probit and logit are as in (2) except instead of (3), $Y_{ik} = 1$ if $\alpha_{k-1} < Y_i^* < \alpha_k$, where $k = 1, \ldots, K$ and $Y_{ik} = 0$ otherwise. K is the number of categories (2 in the binary case) and the α's are usually unobserved and must be estimated by the researcher, subject to a normalization. The log-likelihood function can be found in Maddala (1983), for example. With respect to Pseudo-R^2 measures, the principal conclusions are the same as for the binary probit case, specifically that the McKelvey and Zavoina measure is closest to the latent variable OLS with the normalized Aldrich-Nelson R^2 second and the Cragg-Uhler R^2 third. Moreover as the number of categories is increased to three and to four, the downward bias of most measures lessens except for the McFadden R^2 which becomes worse from the perspective of this criterion. This seems unusual, as adding outcome categories seems to be making the data a closer approximation to continuous data and perhaps we would expect that a Pseudo-R^2 would converge to the continuous data R^2.

Integer variables such as number of children, are sometimes modelled using a count data approach. (See Winkelmann and Zimmermann, 1995 for a survey.) Again only the significance of fit measures are applicable. Merkle and Zimmermann (1992) and Cameron and Windmeijer (1993a,b) suggest the measure R^2_{DV} based on the deviance, as described in the previous section. For ordinal logit and probit and multinomial logit and probit, $R^2_{DV} = R^2_{MF}$. For Poisson models, Cameron and Windmeijer (1993b) show that the LRT/LRT* version of R^2_{MF} as in Table 1 becomes:

$$(10) \qquad R^2_{DV} = \frac{\sum_{i=1}^{N} \left\{ Y_i \log(\hat{U}_i \; / \; \bar{Y}) - (\hat{\mu}_i - \bar{Y}) \right\}}{\sum_{i=1}^{N} Y_i \log(Y_i \; / \; \bar{Y})}$$

where $\hat{\mu}_i = \exp(X_i \hat{\beta})$. They also calculate deviance-based measures for the negative binomial case and other generalized linear models based on the Bernoulli, Gamma and inverse Gaussian. Merkle and Zimmermann (1992) and Cameron and Windmeijer (1993a) also propose

$$(11) \qquad R^2_x = 1 - \sum_{i=1}^{N} \frac{\dfrac{(Y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}}{N \operatorname{var}(Y_i)/\bar{Y}}$$

5. Pseudo-R^2's in the Tobit Case

Sometimes the dependent variable is continuous over some range but limited either from above or below or both. The standard model is as in (2), except that instead of (3), $Y_i = Y_i^*$ if $Y_i^* > 0$ and U is virtually always assumed to have the normal distribution. In his survey of such "Tobit models" (after Tobin, 1958), Amemiya (1984) made no mention of goodness-of-fit measures, in contrast to his earlier 1981 survey of qualitative response models. However, in such cases the likelihood based Aldrich-Nelson and Maddala R^2's are still valid, as are others based on the Magee significance of fit principle. However the McFadden measure is invalid because it relies on the log-likelihood having a maximum of zero, which is not the case when the limited dependent variable is even partially continuous. The normalizations of the Aldrich-Nelson or Maddala measures are also no longer valid. Greene (1981) shows that simply using OLS on the entire data set, censored and uncensored, leads to a downward bias in the R^2.

While McKelvey and Zavoina (1975) did not consider the Tobit case, their Pseudo-R^2 measure is nonetheless valid. (See Veall and Zimmermann, 1990b, 1994) Laitila (1993) provides a formal proof (and also extends the measure to the case where there is only data on non-limit observations, commonly known as truncated regression). However as Laitila (1993) and Veall and Zimmermann (1990b, 1994b) point out, in the Tobit case there is an estimate of σ^2 and no need to set this value by normalization. One of the few measures suggested in the literature specifically for this case is from Dhrymes (1986, p.1603):

$$(12) \qquad R_D^2 = \left[\underset{>0}{\text{corr}}(\hat{Y}_i^A, Y_i)\right]^2$$

where $\hat{Y}_i^A = \hat{Y}_i^* + \hat{\sigma}\phi(x_i\hat{\beta}/\hat{\sigma})/\Phi(-x_i\hat{\beta}/\hat{\sigma})$ and $\hat{Y}_i = X_i\hat{\beta}$, with Φ the cumulative standard normal distribution function and ϕ the standard normal density function.

The symbol ">0" means that the correlation is only taken over the positive observations; $\hat{Y}_i^A$ is the expectation of Y_i conditional on (i) x_i and (ii) $Y_i > 0$, evaluated at the Tobit maximum likelihood estimates $\hat{\beta}$ and $\hat{\sigma}$. Veall and Zimmermann (1994b) also calculate

$$(13) \qquad R^2_{cont} = [\underset{>0}{corr}(\hat{Y}_i^*, Y_i)]^2$$

where there has been no adjustment in Y_i^* for the sample selectivity, as well as a number of other measures which are weighted averages of the others (designed to capture the idea that the Tobit likelihood function can be partitioned into an "OLS part" and a "Probit part").

Veall and Zimmermann (1994b) also provide a Monte Carlo analysis of some Tobit R^2's. Here the use of the "closeness to OLS-R^2" criterion seems particularly useful as it is common to compare OLS and Tobit results in empirical contexts. With samples of 50 and 200 and with degrees of censoring set at 25%, 50% and 75%, the simple McKelvey-Zavoina measure scores by far the best, with the closest challengers some minor modifications based on the McKelvey-Zavoina principle. One of the weighted average measures proposed by Veall and Zimmermann performs acceptably in most cases, and a Magee significance of fit R^2 based on (1) does relatively well when there are only 50 observations and low censoring. Measures that only use part of the sample, such as R_D^2 or R^2_{CONT}, do not do well.

Laitila (1992) also does some Monte Carlo work in the Tobit context, investigating the performance of R^2_{MZ}. He points out that the measure can be estimated entirely with estimates of β, σ and the variance-covariance matrix of the x's and also can be calculated using estimates from other methods besides Maximum Likelihood. He uses estimates from Powell's (1986) method. He also shows that R^2_{MZ} is strongly related to the latent variable OLS-R^2 , and shows that both change similarly as a regressor is added.

Another case with a continuous but limited dependent variable is censored survival data. For Cox's proportional hazard model, Kent and O'Quigley (1988) propose a McKelvey-Zavoina type measure:

(35) $R^2_{KO} = A/(A+1)$

where A is the sum of squared fitted values. Its properties have not been studied in simulation analysis.

6. R^2 Measures from Prediction / Realization Tables.

A completely different style of Pseudo-R^2 is based on predictions and realizations. The prediction-realization table has entries of the form p_{ij} , the fraction of times the realization was outcome I when the model predicted outcome j. We define $p_{\bullet j}$ as the fraction of times alternative j is predicted, $p_{i\bullet}$ as the fraction of times alternative i occurs, p_{mj} is the fraction of times the most common outcome occurs given that outcome j was predicted (that is $p_{mj} = \max_i (p_{ij}))$ and $p_{m\bullet}$ is the most common outcome (that is $p_{m\bullet} = \max_i(p_{i\bullet}))$. There are a number of ways to convert the information into R^2-type measures. Table 2 gives the measures for the binary case although most of the measures can be generalized straightforwardly.

All the measures have an upper bound of one. Some may take negative values but only if the predictive power of the model is worse than random. The "fraction of correct predictions"

C is so commonly used that we have given no reference. This measure can be extremely misleading. For example, C could be .94 for a model of predicting refrigerator ownership, which seems like good performance until one is told that 98% of households own refrigerators and hence a C of .98 is attained by predicting all households own refrigerators.

McFadden, Puig and Kirschner's σ will be positive for a model with any predictive power; σ can be 0 (if p_{ij} = 0,25 for all i,j) and if the predictions are worse than random, σ can even be as small as -1. The maximum value of σ is $1-p^2_{\cdot 1} - p^2_{\cdot 2}$; hence σ_n is a normalized measure. δ is the determinant of the actual 2x2 matrix divided by the determinant of the "perfect fit" 2x2 matrix.

Table 2

R^2 Measures for Binary Prediction/Realization Tables.

Measure	Reference (common name in quotation marks if applicable)
$C = p_{11} + p_{22}.$	no reference, "fraction of correct predictions"
$\sigma = p_{11} + p_{22} - p^2_{\bullet 1} - p^2_{\bullet 2},$	McFadden, Puig and Kirschner (1977)
$\sigma_n = \sigma/(1 - p^2_{\bullet 1} - p^2_{\bullet 2}).$	Veall and Zimmermann (1992b)
$\delta = (p_{11}p_{22} - p_{12}p_{21}) / [(p_{11} + p_{12})(p_{21} + p_{22})]$	Veall and Zimmermann (1992b)
$\Phi^2 = (p_{11}p_{22} - p_{12}p_{21})^2 / (p_{1\bullet}\, p_{2\bullet}\, p_{\bullet 1}\, p_{\bullet 2})$	Bishop, Fienberg and Holland (1975), ("Pearson's Φ"; square root is "Tschuprov's T")
$\lambda = (p_{m1} + p_{m2} - p_{m\bullet}) / (1 - p_{m\bullet}),$	Goodman and Kruskal (1954)
$\lambda' = \frac{p_{11} + p_{22} - p_{m\bullet}}{1 - p_{m\bullet}}$	slight modification of Goodman and Kruskal (1954)
$\tau = (\sum_j \sum_i p_{ij}^2/p_j - \sum p_i^2)/(1 - \sum p_i^2),\ i,\ j = 1,\ 2$	Bishop, Fienberg and Holland (1975)
$Q = (p_{11}p_{22} - p_{12}p_{21})/(p_{11}p_{22} + p_{12}p_{21})$	Bishop, Fienberg and Holland (1975) ("Yule's Q")
$Y = [(p_{11}p_{22})^{0.5} - (p_{12}p_{21})^{0.5}] / [(p_{11}p_{22})^{0.5} + (p_{12}p_{21})^{0.5}]$	Bishop, Fienberg and Holland (1975) ("Yule's Y")
$\kappa = (p_{11} + p_{22} - p_{1\bullet}p_{\bullet 1} - p_{2\bullet}p_{\bullet 2}) / (1 - p_{1\bullet}p_{\bullet 1} - p_{2\bullet}p_{\bullet 2})$	Bishop, Fienberg and Holland (1975)

If prediction is random, $\delta = 0$, but $\delta = 0$ also as soon as a diagonal element equals zero regardless of how well the model fits. Again it is possible for δ to be as small as -1.

Bishop, Fienberg and Holland (1975) is a standard reference and considers "measures of association" for contingency tables. Nominal measures of association between predictions and realizations can be based on variants of the goodness-of-fit χ^2 (GFχ^2) measuring nonindependence. For instance, Pearson's ϕ^2 is GFχ^2 divided by the sample size. We will investigate the square root of Pearson's ϕ^2, which in this case is equivalent to Tschuprov's T. T is in the range [0,1] and T=0 for independence of predictions and realizations. Note that T=1 if $p_{12} = p_{21} = 0$ or $p_{11} = p_{22} = 0$. Therefore, the measure scores a prediction process that is never correct to be as good as one that is always correct. If the predictive power of the qualitative choice model under consideration is at least as good as random, this is not an issue.

Other measures of nominal association employ the proportional-reduction-in-error logic. The approach as applied here is to measure the percentage reduction in the probability of error achieved by the model predictions as opposed to blind guesses. We discuss a measure suggested originally by Goodman and Kruskal (1954), and we also make an obvious modification and suggest λ'. Both measures are in the [0,1] range. To motivate λ and λ', realize that without knowledge of the model, the best guess is to choose the category with the largest marginal probability of realizations ($p_{m\bullet}$). λ' is therefore the fraction of correct predictions minus the fraction of correct predictions by the naive rule that always predicts the most common outcome all divided by a denominator equal to one minus the number of correct predictions by the naive rule. This can be calculated simply in one's head in most instances so that if we know that 60 per cent of the population own houses and a model has a prediction success rate of 90 per cent then $\lambda' = (.9 - .6) / (1 - .6) = .75$. λ' differs from λ in that λ "gives credit" for incorrect predictions if $p_{mj} > p_{jj}$. This is a little like finding value in one of the current author's sports predictions: all one has to do is hear the prediction and bet on the other team. We prefer λ' in this regard (because we think econometric models should not be given credit for being wrong) but note that a consequence is that λ' can be negative (if the model is worse than random) but λ cannot. The measure τ also has a proportion of explained variance interpretation, which is described in Bishop et al. (1975, pp. 389-391).

Popular measures of association for ordinal data are Yule's Q and Y (see Bishop, Fienberg and Holland (1975), pp. 378-379)). Both measures vary in the [-1, 1] range, but indicate a poor model if they take negative values. Note that Q and Y can be 1 even if all observations are not on the main diagonal in the unlikely event that one of the off-diagonal elements is zero.

Measures of agreement [see Bishop, Fienberg and Holland (1975, pp. 397-998) for references] are additional alternatives. κ is a well-known measure of this type.

Veall and Zimmermann (1992b) again conduct a very limited Monte Carlo experiment using the same kind of latent variable models used in the experiments previously described, with the outcome classed as a binary variable (0 or 1) and the prediction classed as a 0 or 1 depending on which had the larger estimated probability. Six measures were very close, with σ_n (our normalization of the McFadden, Puig and Kirschner measure) the best by a little, and δ, κ, T, Y and λ virtually indistinguishable. λ' was not included in the initial study but for this survey we have reperformed the experiments and find it finishes second overall. All these seven measures tend to underpredict OLS-R^2 slightly when less than .5, then overpredict slightly when greater than .5. However, overall the performance is very good, although the error in predicting an underlying OLS-R^2 is obviously much larger when only prediction/ realization information is available than when complete output from say a probit estimation is available. The other measures (τ, Q, σ and C) do not perform well. A clearer choice might emerge in a more extensive set of experiments.

7. Summary and Conclusions

We have surveyed a large literature suggesting many alternative Pseudo-R^2 measures for a variety of cases where the dependent variable is limited in some way. These include the cases of both binary and nonbinary discrete dependent variables, continuous dependent variables with limit observations (most commonly modelled in the Tobit framework) and the case of discrete dependent variables when the only available information is the comparison of predictions and realizations.

Our survey indicates that different Pseudo-R^2's may have very different values on the same model and data. Therefore if researchers are comparing Pseudo-R^2 values from the same estimation technique on different models and data sets, it is obviously important to ensure it is the same Pseudo-R^2. It is also important that such comparisons be informed by the modelling context, just as researchers expect OLS-R^2 to be larger with aggregate time series data in levels than with cross section data.

In some cases, comparisons may be made between R^2 values on different models and data sets where the estimation techniques are not the same, say OLS-R^2 with a continuous dependent variable compared to a Pseudo-R^2 from a binary probit model or a Tobit model. While such comparisons should be treated cautiously, comparability may be possible if all the limited dependent variable models can be cast in a latent variable framework such as in Tobit, binary and ordinal probit and logit but not in multinomial probit on logit. So while in the general case we can suggest that if an R^2 measure is desired one can be obtained as a monotonic function of test statistics for "significance", in cases based on ordered response models we argue that the most useful Pseudo-R^2 is one that is most comparable to OLS-R^2 on the underlying latent variable model.

We have reviewed a number of Monte Carlo experiments from our own earlier research and by others and find that of many candidate R^2's, the McKelvey and Zavoina R^2_{MZ} scores best under the comparability criterion and hence may allow the best possible comparability across OLS, binary and ordinal probit and logit models and Tobit models. However there is some evidence that in binary probit and logit, R^2_{MZ} is more sensitive to misspecification in the error term than the more common McFadden R^2_{MF}. Also models of the multinomial probit or multinomial logit type do not lend themselves to comparison with OLS and for these cases only R^2_{MF} and a class of measures summarized by Magee (1990) seem worthwhile, although little is known as to which alternative is best. In the case of simple prediction-realization comparisons, limited simulation analysis suggests that a number of measures are very close but a normalization we propose of a measure due to McFadden, Puig and Kirschner performs a bit better than the others under our criterion. While it is straight forward to calculate, it is even easier to compute an obvious modification of Goodman and Kruskal's λ which also performs well under our criterion.

References

Aldrich, J. H. and Nelson, F. D. (1984): *Linear Probability, Logit, and probit Models*, Sage University Press, Beverly Hills.

Amemiya, T. (1981): "Qualitative Response Models: A Survey", *Journal of Economic Literature* 19, pp. 1483-1536.

Amemiya, T. (1984): "Tobit Models: A Survey", *Journal of Econometrics* 24, pp. 3-61.

Bishop, Y.M., Fienberg, S.E., Holland, P.W. (1975): *Discrete Multivariate Analysis: Theory and Practice* (MIT Press, Cambridge, MA).

Blundell, R. (editor) (1987): “Specification Testing in Limited and Discrete Dependent Variable Models”, *Journal of Econometrics*, 34.

Cameron, A.C. and Windmeijer, F.A.G. (1993a): "R-Squared Measures for Count Data Regression Models with Applications to Health Care Utilization", Dept. of Economics Working Paper 93-24, University of California at Davis.

Cameron, A.C. and Windmeijer, F.A.G. (1993b): "Deviance Based R-Squared Measures of Goodness of Fit for Generalized Linear Models", Working Papers in Economics and Econometrics.

Cox, D.R. and Wermuth, N. (1992): "A Comment on the Coefficient of Determination for Binary Probit", *The American Statistican* 46, 1, pp. 1-4.

Cragg, J.G., and Uhler, R. (1970): "The Demand for Automobiles". *Canadian Journal of Economics* 3, pp.386-406.

Dhrymes, P.J. (1986): "Limited Dependent Variables", in Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics* Vol III, Amsterdam: North-Holland, pp. 1567-1631.

Efron B. (1978): "Regression and ANOVA with Zero-One Data: Measures of Residual Variation", *Journal of the American Statistical Association* 73, pp. 113-121.

Goldberger, A.S. (1973): "Correlations Between Binary Choices and Probabilistic Predictions," *Journal of the American Statistical Association* 68, p. 84.

Goodman, L.A. and Kruskal, W.H. (1954): "Measures of Association for Cross-Classifícation", *Journal of the American Statistical Association* 49, pp. 732-764.

Greene, W.H. (1981): "On the Asymptotic Bias of the Ordinary Least Squares Estimator of the Tobit Model", *Econometrica* 49, 2, pp. 505-513.

Greene, W.H. (1995): LIMDEP Version 7.0 User's Manual, Econometric Software Inc., Bellport, NY.

Hagle, T.M. and Mitchell II, G.E. (1992): "Goodness-of-fit Measures for Probit and Logit", *American Journal of Political Science* 36, 762-784.

Hauser, J.R. (1977): Testing the Accuracy, Usefulness and Significance of Probabilistic Choice Models: An Information Theoretic Approach", *Operations Research* 26, pp. 406-421.

Kent, J.T. and O´Quigley, J. (1988): "Measures of Dependence for Censored Survival Data", *Biometrika* 75, 3, pp.525-534.

Kvålseth, T.O. (1985): "Cautionary Note About R^2", *The American Statistician* 34, No. 4, pp.274-285.

Laitila, T. (1992): "A Pseudo-R^2 Measure for Limited and Qualitative Dependent Variable Models", mimeograph, University of Umeå, Sweden.

Laitila, T. (1993): "A Pseudo-R^2 Measure for Limited and Qualitative Dependent Variable Models", *Journal of Econometrics* 56, 341-356.

Lave, C. A. (1970): "The Demand for Urban Mass Transportation", *Review of Economics and Statistics* 52, pp. 320-323.

Maddala, G.S. (1983): *Limited-dependent and Qualitative Variables in Econometrics*, New York: Cambridge University Press.

Magee, L. J. (1990): "R^2 Measures Based on W and LR Joint Significance Test Statistics", *The American Statistician* 44, pp. 250-253.

McFadden, D. (1973): "Conditional Logit Analysis of Qualitative Choice Behavior", in Zarembka, P. (ed.), *Frontiers in Econometrics*, pp.105-142, Academic Press, New York.

McFadden, D., Puig, C., Kirschner, D. (1977): "Determinants of the Long-run Demand for Electricity", *Proceedings of the American Statistical Association* (Business and Economics Section), pp.109-117.

McKelvey, R., and Zavoina, W. (1975): "A Statistical Model for the Analysis of Ordinal Level Dependent Variables", *Journal of Mathematical Sociology* 4, pp. 103-120.

Merkle, L. and Zimmermann, K.F. (1992): "The Demographics of Labor Turnover: A Comparison of Ordinal Probit and Censored Count Data Models", *Recherches Economiques de Louvain* 58, pp.283-307.

Morrison, D. G. (1972): "Upper Bounds for Correlation Between Binary Outcomes and Probabilistic Predictions", *Journal of the American Statistical Association* 67, pp. 68-70.

Neter, J., and Maynes, E.S. (1970): "Correlation Coefficient with a 0,1 Dependent Variable," *Journal of the American Statistical Association* 65, pp. 501-509.

Pagan, A. and F. Vella (1989): “Diagnostic Tests for Models Based on Individual Data: A Survey”, *Journal of Applied Econometrics*, 4, pp. s29-s59.

Powell, J.L. (1986): "Symmetrically Trimmed Least Squares Estimation for Tobit Models", *Econometrica* 54, pp. 1435-1460.

Smith, R. and S. Peter (1991): “Distributional Specification Tests Against Semi-Parametric Alternatives”, *Journal of Econometrics*, 47, 175-194.

STATA (1995): Stata Reference Manual, Release 4, Computing Resource Center, Santa Monica California.

Tobin, J. (1958): "Estimation of Relationships for Limited Dependent Variables", *Econometrica* 26, pp. 24-36.

Veall, M.R., and Zimmermann, K.F. (1990a): "Pseudo-R^2's in the Ordinal Probit Model", Discussion Paper No. 90-15, University of Munich.

Veall, M.R. and Zimmermann, K.F. (1990b): “Goodness of Fit Measures in the Tobit Model”, Discussion Paper No. 90-23, University of Munich.

Veall, M.R. and Zimmermann, K.F. (1992a): "Pseudo-R^2´s in the Ordinal Probit Model", *Journal of Mathematical Sociology* 16, pp. 332-342.

Veall, M.R. and Zimmermann, K.F. (1992b): "Performance Measures from Prediction-Realization Tables", *Economics Letters* 39, pp.129-134.

Veall, M.R. and Zimmermann, K.F. (1994a): "Evaluating Pseudo-R^2`s for Binary Probit Models", *Quantity and Quality*, 28, pp. 151-164.

Veall, M.R. and Zimmermann, K.F. (1994b): "Goodness of fit measures in the Tobit Model", *Oxford Bulletin of Economics and Statistics*, 56, pp. 485-499.

Veall, M.R. and Zimmermann, K.F. (1995): “Comments on ‘Goodness-of-Fit Measures in Binary Choice Models’”, *Econometric Reviews*, 14, pp. 117-120.

Windmeijer, F.A.G. (1995): "Goodness-of-fit measures in Binary Choice Models", *Econometric Reviews*, 14, pp. 101-116.

Winkelmann, R. and Zimmermann, K.F. (1995): “Recent Developments in Count Data Modelling: Theory and Applications”, *Journal of Economic Surveys*, 9, pp. 1-24.

<u>Figure 1:</u>

Reference OLS R^2 and Predicted Pseudo-R^2's:
An Overview for the Binary Probit Case