A NEW MEASURE OF FIT FOR EQUATIONS WITH DICHOTOMOUS DEPENDENT VARIABLES

Arturo Estrella

A New Measure of Fit
For Equations With Dichotomous Dependent Variables

Arturo Estrella

Federal Reserve Bank of New York
33 Liberty Street
New York, New York 10045
Tel.: 212-720-5874
Fax: 212-720-1582
E-mail: arturo.estrella@frbny.sprint.com

ABSTRACT

The econometrics literature contains many alternative measures of goodness of fit, roughly analogous to R^2, for use with equations with dichotomous dependent variables. There is, however, no consensus as to the measures' relative merits or about which ones should be reported in empirical work. This paper proposes a new measure that possesses several useful properties that the other measures lack. The new measure may be interpreted intuitively in a similar way to R^2 in the linear regression context.

Key Words: model evaluation, probit, logit

1. INTRODUCTION

The econometrics literature contains many alternative measures of goodness of fit, roughly analogous to R^2, for use with equations with dichotomous dependent variables (DDVs). There is, however, no consensus as to the measures' relative merits and they are not consistently reported in empirical work. The upshot is that, whereas reported results of linear regressions invariably include the R^2 or adjusted R^2, results of DDV equations do not consistently contain any one measure of fit.

Among earlier proposals, moment-based measures generally fail to deal with the pervasive problem of heteroskedasticity or use restrictive linear approximations. Those based on likelihood ratio statistics are preferable because of their relationship to valid hypothesis tests. However, they are often obtained by arbitrary rescalings of functions of the likelihood ratio statistics.

This paper proposes a new measure that possesses several important properties that the current measures lack. The measure is constructed by imposing certain restrictions on its relationship with the underlying likelihood ratio statistic. These restrictions, including one expressed in terms of marginal increments in fit, are shown to be consistent with the formal properties of R^2 in the linear case and to provide consistently accurate signals as to statistical significance. The new measure may be interpreted intuitively in a similar way to R^2 in the linear regression context, even away from the endpoints of its range of values.

Section 2 introduces the DDV model and constructs the new measure of fit. In section 3, the new measure is compared with earlier proposals using various criteria.

Section 4 contains some remarks on model selection, section 5 then provides numerical illustrations based on a model for predicting recessions, and section 6 concludes.

2. THE DDV MODEL AND THE NEW MEASURE OF FIT

2.1 The DDV Model

This section presents the basic DDV model and discusses briefly both its estimation and tests of the hypothesis that all coefficients but the constant term are zero, which is usually associated with R^2. The exposition in this introduction is somewhat sketchy. Greater detail is provided, for example, in Maddala (1983) or Amemiya (1981).

The standard DDV model is defined in reference to a linear relationship of the form

$$y* = \beta'x + e, \tag{1}$$

where y* is unobservable, ß is a vector of k+1 coefficients, and x is a vector of values of k+1 independent variables, the first one of which is always unity. There is also an observable variable y, which has only two possible values and is related to y* in the following way:

$$y = 1 \quad if \ y*>0$$
$$y = 0 \quad otherwise.$$

The form of the estimated equation is

$$P(y=1|x) = F(\beta'x), \tag{2}$$

where F is the cumulative distribution function for -e. In practice, F is usually specified as normal or logistic, but any other continuous distribution function whose first two

2

derivatives exist and are well-behaved may be used. It suffices that F be twice continuously differentiable and that the resulting likelihood function be concave in ß.

The model is usually estimated by maximizing the likelihood function, which is defined as:

$$L = \prod_{\{y_j=1\}} F(\beta' x_j) \prod_{\{y_j=0\}} (1 - F(\beta' x_j))$$

Sufficient first and second order conditions for a maximum likelihood estimate b of ß are:

$$\partial L(b) / \partial \beta = 0 \quad and \quad \alpha' \left(\partial^2 L(b) / \partial \beta^2 \right) \alpha < 0$$

for all k+1-element vectors α. The resulting unconstrained maximum value of L will be denoted as L_u. The fitted values from equation (2) are

$$\hat{y} = P(y=1 | x) = F(b' x) .$$

Consider the hypothesis H_0 that all k coefficients of ß, other than the first, are zero. Then, if there are n observations, of which n_1 are such that y=1, L is maximized under H_0 when $F(\beta_0) = \bar{y} = n_1 / n$ and the maximum constrained value of L is $L_c = \bar{y}^{n_1} (1 - \bar{y})^{n - n_1}$. Note also that the average value of the log-likelihood function has a particularly simple form that depends only on $\bar{y}$:

$$A_c(\bar{y}) \equiv \log L_c / n = \bar{y} \log(\bar{y}) + (1 - \bar{y}) \log(1 - \bar{y}) \tag{3}$$

This fact will be useful in some of the subsequent analysis.

The hypothesis H_0 may be tested using the likelihood ratio $\lambda = L_c/L_u$. A well-known result (see, for example, Rao 1973) states that $-2 \log \lambda$ is asymptotically distributed under H_0 as a chi-squared variable with k degrees of freedom.

2.2 Motivation for the New Measure: R^2 and the Classical Test Statistics

In the standard linear model with normally distributed errors, there is a simple exact relationship between R^2 and the likelihood ratio statistic. The relationship is best expressed in terms of the average value of the statistic, namely,

$$A_{LR} = (2/n)(\log L_u - \log L_c),$$

which takes on values between zero (when there is no fit) and infinity (when the fit is perfect). Then

$$R^2 = 1 - \left(L_c/L_u\right)^{2/n} = 1 - \exp(-A_{LR}). \tag{4}$$

See, e.g., Anderson (1958). R^2 may be seen as a nonlinear rescaling of the average likelihood ratio statistic that takes on values in the unit interval. The endpoints of the interval correspond in a straightforward way to "no fit" and to a "perfect fit", respectively.

Furthermore, differences in the statistic are related to differences in R^2 in an intuitive way. Specifically,

$$dR^2/(1-R^2) = dA_{LR}. \tag{5}$$

The left hand side of this equation is a marginal R^2, that is, it is the change in explained variation as a proportion of the variation that remains to be explained. The equation

4

indicates that for a small change in A_{LR}, the change also represents the marginal R^2. This relationship holds for any starting value of R^2.

The above relationship also holds locally for the average values of the other classical test statistics -- Lagrange multiplier and Wald -- in a neighborhood of the null H_0. Let W and LM, respectively, denote the values of the Wald and Lagrange multiplier statistics in the linear case, and define their average values as $A_W = W/n$ and $A_{LM} = LM/n$. It may be shown (see similar equations in Evans and Savin 1982, for example) that

$$R^2 = A_{LM} = A_W / (1 + A_W) = 1 - e^{-A_{LR}}. \tag{6}$$

From these identities, it may be verified that

$$R^2 = 0 \quad for \quad A_i = 0,$$
$$\lim R^2 = 1 \quad as \quad A_i \to B_i, \quad and \tag{7}$$
$$dR^2 / (1 - R^2) = dA_i \quad (dR^2/dA_i = 1) \quad for \quad A_i = 0,$$

where B_i is the upper bound for the corresponding statistic, that is, $B_i = 1$ for the Lagrange multiplier and ∞ for the likelihood ratio and Wald statistics.

2.3 Construction of the New Measure of Fit

Let us turn now to the DDV case. With a DDV, the approach of equation (4) fails because the average likelihood ratio statistic is bounded. Let A (with no subscript) denote the average likelihood ratio statistic in the DDV case, that is,

$$A = (2/n) (\log L_u - \log L_c).$$

The upper bound for A corresponds to the case where the fit is perfect and $L_u = 1$. The upper bound is thus given by

$$B = -(2/n) \log L_c = -2A_c(\bar{y}),$$

where A_c is as given in equation (3) and is only a function of the average value of the DDV. As $\bar{y}$ approaches one of the extreme values of 0 or 1, B approaches zero. B is symmetrical in $\bar{y}$ and attains its maximum value when $\bar{y} = \frac{1}{2}$, for which

$$B_{max} = \log 4 = 1.386.$$

The analysis of section 2.2 may be used to motivate, in the DDV case, a set of requirements for an R^2 analogue that possesses many of that measure's useful features. The following requirements are suggested by the analysis.

(i) The measure should take on values in the unit interval and have the right interpretation at the endpoints, i.e., zero corresponds to no fit and one corresponds to a perfect fit.

(ii) The measure should be based on a valid test statistic for the hypothesis that the coefficients of all explanatory variables, save the constant, are zero.

(iii) The derivative of the measure with respect to the test statistic should accord with the corresponding derivative in the linear case.

Requirement (i) is straightforward. The range of values is in principle arbitrary, but (i) corresponds to the conventional values of R^2 in the linear case.

Requirement (ii) is theoretically appealing and has the practical intent of avoiding conflicting signals between the measure of fit and the related statistical test. Of course, the linear case R^2 may be derived in a least-squares context without reference to distributional

assumptions. Nevertheless, R^2 has the desirable property that, given the number of observations and regressors, the ranking of models generated by R^2 agrees exactly with the ranking generated by the corresponding F test.

Requirement (iii) is essential if the value of the measure of fit is to be in itself meaningful. The significance of the extreme values of zero and one is clear. However, through experience, practitioners also tend to develop a sense of what an interior value of R^2 means. Generally, for instance, 0.25 may be interpreted as "modest", 0.5 as "strong" and 0.75 as "very strong". If (iii) is not imposed, the possibility of such interpretations is lost.

To see this, consider the measure $(A/B)^\gamma$, where γ is a positive number and A and B are defined as in this section. This measure satisfies requirements (i) and (ii), regardless of the value of γ, and has the appropriate interpretation at 0 and 1. Now suppose that $A/B = \frac{1}{2}$ and let $\gamma = 2$, 1 or $\frac{1}{2}$. The resulting values of the measure of fit are then 0.25, 0.5 or 0.71, respectively, which cover the whole gamut from "modest" to "very strong". It is thus clear that the scaling of interior points, in this case accomplished through γ, cannot be completely arbitrary if the measure is to be interpretable.

We now construct a measure of fit that satisfies (i)-(iii). In line with (ii), the measure is based on A, the average likelihood ratio statistic in the DDV case. In order to achieve the appropriate scaling, we specify (iii) as a differential equation with boundary conditions defined by (i).

The expression of (iii) as a differential equation rests primarily on an analogy with the relationship between marginal R^2 and the LM statistic in the linear case. The LM

statistic is a natural model because, like A, it is bounded above. Using equation (6), marginal R^2 in the linear case may be expressed in terms of the average LM statistic as:

$$dR^2/(1-R^2) = dA_{LM}/(1-A_{LM}) .$$

(8)

That is, marginal R^2 increases at a rate inversely proportional to the distance between the current value of the statistic and its upper bound. This distance is the unexplained fraction of the variance of the dependent variable.

In the DDV case, a measure based on the statistic A may be constructed by analogy, using the fact that $0 \le A/B \le 1$. The measure is designed so that the marginal increment in fit is inversely proportional to the fraction of the "information content" of y (see Theil 1971, section 12.6) that is still unexplained, where this fraction is interpreted as $1-A/B$. Hence, define ϕ_0 implicitly by the differential equation:

$$d\phi_0/(1-\phi_0) = dA/(1-A/B) .$$

(9)

With the initial condition $\phi_0(0)=0$, the solution to the equation is

$$\phi_0 = 1-(1-A/B)^B = 1-(logL_u/logL_c)^{-(2/n)logL_c}.$$

This solution also satisfies the conditions $\phi_0(B)=1$ and $\phi_0'(0)=1$.

The similarity between equations (8) and (9) suggests that, in a well-defined way, requirement (iii) holds over the entire range $0 \le \phi_0 \le 1$. Also, note that if B is "replaced by" infinity in the formula for ϕ_0,

8

$$\lim_{B \to \infty} 1 - (1 - A/B)^B = 1 - \exp(-A) \; ,$$

which is the exact expression for R^2 in the linear case (equation 4), where A is unbounded.

The applicability of the measure ϕ_0 extends beyond the DDV case. Although several special features of the DDV model have been used to motivate the need for such a measure and to specify the requirements that it should meet, the foregoing derivation uses only a few basic facts about the statistical problem. First, the model is estimated by maximum likelihood. Second, the hypothesis that all the coefficients except for the constant term are zero is tested using a likelihood ratio test. Third, the average likelihood ratio statistic has a finite upper bound. Under these conditions (which are also met, for instance, in polychotomous dependent variable cases such as multinomial logit or probit), ϕ_0 may be used as a measure of fit.

3. COMPARISON OF THE NEW MEASURE WITH EARLIER PROPOSALS

3.1 A Summary of Previous Proposals

Many measures of fit have been proposed in the context of the DDV model. The following is a fairly comprehensive list of measures contained in the literature, with an indication of the original proponents.

$$\phi_1 = 1 - \log L_u / \log L_c \qquad\qquad \text{McFadden (1974)}$$

$$\phi_2 = 1 - \left(L_c / L_u \right)^{2/n} \qquad\qquad \text{Cragg-Uhler (1970)}$$

$$\phi_3 = \frac{1 - \left(L_c/L_u\right)^{2/n}}{1 - L_c^{2/n}}$$

Cragg-Uhler (1970)

$$\phi_4 = \frac{2\left(\log L_u - \log L_c\right)}{2\left(\log L_u - \log L_c\right) + n}$$

Aldrich-Nelson (1989)

$$\phi_5 = \frac{2\left(\log L_u - \log L_c\right)}{2\left(\log L_u - \log L_c\right) + n} \cdot \frac{2\log L_c - n}{2\log L_c}$$

Veall-Zimmermann (1992)

$$\phi_6 = \rho^2\left(y, \hat{y}\right)$$

Morrison (1972), Goldberger (1973)

$$\phi_7 = 1 - E\left(y - \hat{y}\right)^2 / E\left(y - \bar{y}\right)^2$$

Efron (1978)

$$\phi_8 = \rho^2\left(y, x\right)$$

Davidson-McKinnon (1993)

The notation in these formulas has been introduced earlier with the exception of the correlation coefficient, which is denoted ρ.

Measures one, two, three and six are discussed in the Maddala (1983) survey, whereas Amemiya (1981) considers numbers one, six and seven. Dhrymes (1986) contains an extensive discussion of the first measure only, and Magee (1990) suggests a general procedure that is used to define measures two, three, four and seven.

Measures four and five are analyzed in recent surveys by Veall and Zimmermann (1992) and Windmeijer (1995). Those papers focus on the fit of equation (1), involving the latent dependent variable y^*, and both endorse a proposal by McKelvey and Zavoina (1975) based on the fitted values from (1): $E\left(\hat{y}* - \bar{\hat{y}}*\right)^2 / \left(E\left(\hat{y}* - \bar{\hat{y}}*\right)^2 + 1\right)$. However, both surveys also conclude that this measure is not very useful in assessing the fit of the

probability equation (2), which is our focus here. Finally, measure eight may be inferred from a test proposed by Davidson and McKinnon (1993, p. 525) in the context of an artificial regression.

Measures one to five are based on the maximum likelihood statistics. Measures six to eight, in contrast, are based on first and second moments of the actual and fitted values of the dependent variable y and of the explanatory variables x. Thus, it is appropriate to analyze each of the two subsets separately, as in the next two subsections. It will be argued that each of the foregoing measures lacks at least one of the important properties (i)-(iii) that an R^2 analogue should have.

3.2 Comparison of New Measure with Likelihood-Based Alternatives

The use of the average likelihood (A) and its upper bound (B), defined for the DDV model in section 2.3, simplifies the analysis of the likelihood-based measures of fit ϕ_0-ϕ_5, as well as their relationships. In this section, we begin by examining whether each of the measures satisfies conditions (i) and (iii) of section 2.3. The results are summarized in table 1, which presents for each of the measures: an expression in terms of A and B, the values of the measure at the endpoints, and the derivative of the measure with respect to A ($d\phi_i / dA$) when A=0. Condition (i) suggests that the endpoint values should be zero and one, respectively, and condition (iii) suggests that the derivative should be one.

Closed-form relationships exist among some of the measures. For instance, it may be shown that $\phi_1 < \phi_0$ if and only if B > 1 and that $\phi_4 < \phi_2$ and $\phi_5 > \phi_3$ for 0 < A < B.

More importantly, table 1 demonstrates the following conclusions. First, all the measures have a value of zero when A=0. Second, measures two and four fail the upper

11

bound criterion. They have upper bounds that depend on B, but which in all cases are much less than one. Third, measures one, three and five violate the requirement that the derivative at A=0 be unity. Measure one may meet this requirement, but only with the data-driven condition that B is exactly one ($\bar{y}$=.1997). Thus, only measure zero satisfies requirements (i)-(iii).

Evidence that ϕ_0 is preferable to measures one to five is provided by a comparison of these measures with the odds ratio in the context of a two-by-two contingency table. Suppose that there is only one independent variable x in (1) and that, like y, x is dichotomous. The information contained in a given sample may be summarized in a two-by-two table of frequencies of the form:

	y=1	y=0
x=1	n_{11}	n_{10}
x=0	n_{01}	n_{00}

The odds ratio $r = n_{11}n_{00}/(n_{10}n_{01})$ is an index of the degree of independence of x and y. Independence corresponds to $r=1$, whereas positive and negative relationships are indicated by r greater than and less than one, respectively. Under the null of independence, r has a hypergeometric distribution with parameters $n_{11}+n_{10}$, $n_{01}+n_{00}$ and $n_{11}+n_{01}$ (Lehmann 1986, section 4.6). Given the sample size (n) and the marginal distributions of x and y ($\bar{x}=(n_{11}+n_{10})/n$ and $\bar{y}=(n_{11}+n_{01})/n$), the value of r is sufficient to determine all the frequencies in the table. In turn, the information in the table

is sufficient to calculate the likelihood-based measures of fit if the distribution function F in
(2) is specified.

Let F be logistic. The estimate of the slope coefficient β_1 is then $\log(r)$ and, for
any given set of marginal distributions of x and y, all the likelihood-based measures zero to
five are monotonically increasing in the odds ratio. For instance, suppose that a sample of
500 observations is drawn from a population with $\bar{x}=\bar{y}=.5$. If the odds ratio is
alternatively 1, 3, 30, and 300, then $\phi_0=0$, .07, .48, and .81, respectively. However,
when the marginal distributions are allowed to vary, the odds ratio and the measures of fit
may not move in tandem.

For various marginal distributions of x and y and a sample of size 500, table 2
presents the critical value of the odds ratio at the .00001 level, as well as the values of
measures zero through five that correspond to each critical value of the odds ratio. We
focus on a low significance level because, otherwise, the corresponding values of all the
measures of fit would tend to be quite low.

As $\bar{y}$ changes with $\bar{x}$ held constant, measures one, three and five sometimes move
with the odds ratio, but sometimes move in the opposite direction, as when $\bar{x}=.05$. The
range of values of these measures is fairly broad, particularly considering that the
significance level of the critical odds ratios is constant. Measure zero, in contrast, remains
virtually constant, which is consistent with the constant significance level and supports the
proposed interpretation of this measure. The same conclusion may be reached with regard
to measures two and four in the context of table 2. Note, however, that those measures
fare well only in a neighborhood of $A=0$, where, like ϕ_0, they meet certain conditions

13

with regard to their level and first derivative (see table 1). As A approaches B, measures two and four fail requirement (i) of section 2.3 and cannot approach unity.

Another comparison of the likelihood-based measures follows from an important and useful feature of the standard R^2: that it may be transformed into an F statistic by $F = R^2 / (1-R^2) \cdot (n-k-1) / k$, where n is the number of observations and k the number of explanatory variables. From this expression, a level of significance may be calculated for the statistical test that all coefficients but the constant term are zero (see Maddala 1977). If any of the ϕ_i measures in the DDV model is interpreted as an R^2, the same procedure may be applied purely formally to obtain an F-test significance level.

Consider a data set with n observations, a DDV y, and k explanatory variables (not necessarily dichotomous). Let $X = -2 \log \lambda$ be the chi-squared statistic described in section 2.1 for testing whether all k coefficients in the DDV model are zero. From X (and k), a level of significance α_X for the chi-squared test may be calculated. In addition, X (together with n and $\overline{y}$) may be used to calculate any of the measures of fit ϕ_0 through ϕ_5 that are based on maximum likelihood estimates. Now interpret the value of one of these measures as the R^2 from a linear regression and perform an F test formally by calculating $F_i = \phi_i / (1-\phi_i) \cdot (n-k-1) / k$ and its significance level α_F.

If the interpretation of ϕ_i as an R^2 is plausible, the implicit α_F will be roughly equal to α_X. Results for n=500, k=2 and significance levels of .01, .05 and .10 are presented in table 3 for measures zero through five. The table shows cases in which $\overline{y} = .05$ and .5 because the results for measure one are quite sensitive to this parameter.

The results in table 3 are fairly striking in support of ϕ_0. The statistical

significance implied by interpreting this measure as an R^2 is almost identical to the actual

significance measure of the appropriate chi-squared test. However, the results for both ϕ_3

and ϕ_5 are consistently very low, with the high levels of these measures overstating the

significance of the relationships. For ϕ_1, the results depend on the value of $\overline{y}$. When

$\overline{y} = .05$, measure one overstates the significance substantially, although not as much as

measures three and five. When $\overline{y} = .5$, on the other hand, the values of measure one tend

to be low and the significance levels too high. This pattern suggests that there may be an

intermediate case in which ϕ_1 does relatively well. In fact, for $\overline{y} = .1997$, ϕ_1 is identical to

ϕ_0, and both perform quite well. In table 3, as in table 2, ϕ_2 and ϕ_4 perform about as well

as ϕ_0. Again note, however, that measures two and four are subject to upper bounds that

are considerably less than one.

The foregoing results assume specific values of n, k and $\overline{y}$, and only sensitivity to

the latter has been examined here. Nevertheless, experimentation with other values of n

and k shows that the results are robust. For small samples or many independent variables,

the results for measure zero may deteriorate a bit, but they are still good and incomparably

better than those obtained with the other measures.

3.3 Comparison of New Measure with Moment-Based Measures of Fit

We proceed to compare measures six to eight, as defined in section 3.1, with ϕ_0.

The moment-based measures make some intuitive sense and can be sometimes helpful in

evaluating the results of estimates with DDVs. However, they are based on second

moments (sums of squares), whereas the maximum likelihood framework that produces the

15

estimates does not depend on such statistics in the DDV case. Hence, the use of the moment-based functions as measures of fit is inconsistent in principle with the use of maximum likelihood for estimation. The moment-based measures tend to ignore either or both the heteroskedasticity and nonlinearity problems of the basic linear probability model.

To see this, we perform a Monte Carlo experiment in which a likelihood-based measure (ϕ_0) and a moment-based measure (ϕ_7) are used for model selection. The object is to distinguish the true model, which contains a single variable x_1, from a model with a single variable x_2 that is imperfectly correlated with the first. Thus, suppose

$y* = c + b x_1 + u_1$ with $x_1, u_1 \sim N(0,1)$ and x_1 and u_1 uncorrelated. The DDV y is based on whether y* is positive. Suppose further that there are two other variables $x_2, u_2 \sim N(0,1)$ such that $x_1 = \rho x_2 + \sqrt{1-\rho^2} u_2$ with x_2 and u_2 uncorrelated. The correlation between x_1 and x_2 is ρ, which is set at .9 in the simulations that follow.

The strategy of the Monte Carlo experiments is to simulate a sample of size 500, to estimate probit equations based on x_1 and x_2, respectively, and to observe the ranking of the two models according to each of the two measures of fit. One thousand iterations are performed and the probability of selecting the wrong model with each of the two measures is noted. The results are presented in table 4. Note that to make the results in this table more easily interpretable, combinations of the parameters c and b are selected so that the means of y and of ϕ_0 have specific reference values. For each of these combinations, the table indicates the probability $p(\phi_i)$ of an error (selecting model 2) using measure i=0 or 7 and the results of a Monte Carlo t-test of whether the probability of an error is higher with ϕ_7 than with ϕ_0.

Not surprisingly, since the parameters are estimated by maximum likelihood, the likelihood-based measure has a tendency to perform better. When the fit of the model is particularly good ($\phi_0=.5$), both measures are very accurate. However, when the fit is not as good and the mean y is away from .5, the likelihood-based measure clearly outperforms the moment-based.

4. SOME REMARKS ON MODEL SELECTION

As a criterion for empirical model selection, R^2 has the drawback that it can only increase when additional variables are introduced. For this reason, empirical researchers frequently use and report $\overline{R}^2$, which increases with an additional variable only if the improvement in fit is sufficient to overcome a loss in degrees of freedom. Adjusted R^2 is defined implicitly (see Maddala 1977) by:

$$1 - \overline{R}^2 = (1 - R^2) \cdot (n-1) / (n-k-1) ,$$

where k is the number of explanatory variables. Like R^2, ϕ_0 is nondecreasing with the introduction of additional explanatory variables. By analogy, we can define an adjusted ϕ_0 by

$$\phi_0^b = 1 - (1 - \phi_0) \cdot (n-1) / (n-k-1) .$$

An alternative degrees-of-freedom adjustment may be obtained by employing the form of the Akaike information criterion (AIC) within the formula for ϕ_0. Citing earlier research that shows that no one adjustment for degrees of freedom dominates other possible

17

choices, Amemiya (1981) expresses a preference for the AIC ($= -\log L_u + k$) because of its simplicity. The resulting adjusted ϕ_0 is in this case:

$$\phi_0^a = 1 - \left((\log L_u - k) / (\log L_c)\right)^{-\frac{2}{n}\log L_c}.$$

The AIC adjustment tends to impose a more severe penalty for additional variables than the $\overline{R}^2$-type adjustment. For example, when $\phi_0 = 0.5$ and $n = 120$, the inclusion of one additional variable tends to reduce ϕ_0^a by 0.014 or more, depending on the value of $\overline{y}$, whereas ϕ_0^b is reduced only by about 0.005. The stiffer penalty associated with the AIC may be desirable, since, in the linear model, the condition for an additional variable to increase $\overline{R}^2$ is only that its t statistic exceeds one.

The foregoing suggestions for use of the measure of fit in model selection seem plausible, but more research is required to determine whether the modifications are useful in practice. Such research, which is beyond the scope of this paper, might use Monte Carlo simulations to examine the effectiveness of the adjustments in the context of the selection of nested and nonnested models.

5. NUMERICAL ILLUSTRATIONS USING A MODEL FOR PREDICTING RECESSIONS

This section illustrates the application of the new and previously suggested measures to an equation that predicts whether or not an economy will be in a recession 12 months ahead. The dependent variable y in this equation equals 1 if the economy is in a recession, and it equals 0 otherwise. The equation includes, in addition to a constant term, a single

explanatory variable (SPREAD) representing a yield curve spread: the difference between

a 10-year government bond and a 3-month government bill. Specifically,

$$P(y_t=1 \mid SPREAD_{t-12}) = F(\beta_0 + \beta_1 SPREAD_{t-12}),$$

where the distribution F is taken to be normal.

The equation is estimated for France, Germany, Italy, the United Kingdom and the

United States with monthly data from January 1973 to December 1994. The interest rate

data is as of the end of the month. The recession DDV is based on business cycle dating

by the National Bureau of Economic Research (NBER) for the United States and by the

Columbia Center for Business Cycle Research for the European countries. For France and

Italy, data for the full sample were not available. A similar set of estimates and a

discussion of the underlying economics may be found in Estrella and Mishkin (1996) (see

also Estrella and Hardouvelis (1991) for a U.S. application over a longer sample period).

Results are shown in table 5. The ordering across countries produced by the nine

basic measures is identical, which is attributable to the large differences in fit that make

distinctions easy. That the ordering is reasonable may be confirmed by examining figure

1, which plots the fitted probabilities from the equation against the shaded recessionary

periods (when y=1). Of course, visual inspection of the figure can only provide a

subjective assessment of the fit. Nevertheless, the numerical results confirm the patterns

suggested by the discussion of the earlier sections, as inspection of the numbers in table 5

indicates.

Measures six to eight, which are based on moments, produce results that are very similar to each other. These results are also generally similar to those obtained with ϕ_2, except in the case of the United States in which $\bar{y}$, and hence the upper bound for ϕ_2, are low. The moment-based measures are also generally lower than those corresponding to the preferred measure ϕ_0. This latter difference is particularly large in the case of Germany, for which the failure of the moment-based measures to account for heteroskedasticity results in giving "too much weight" to the outliers in the second-moment calculation.

The alternatives based on maximum likelihood exhibit the behavior that would be expected from the analysis of sections 2 and 3. For example, measures two and four, whose values cannot exceed 0.75 and 0.58, respectively, produce relatively low estimates. In contrast, measures three and five, which are obtained by blowing up measures two and four, respectively, in a somewhat arbitrary way, are consistently higher than all the others.

Measure one makes the smallest distinction between the fits for Germany and the United States, about 8 percentage points as compared with 20 for ϕ_0. B is not far from one for the United States, but is essentially at its maximum value (1.386) for Germany. Hence, the nonlinear adjustment included in ϕ_0, but not in ϕ_1, has a much more noticeable effect for Germany.

The entry in the table labeled MZ corresponds to the McKelvey-Zavoina measure of fit of the latent equation, as defined in section 3.1. These results are consistently on the high side. Finally, the values of the two measures of section 4 that adjust for degrees of freedom are also presented in the table.

20

6. CONCLUSIONS

This paper proposes a new measure of fit for equations with dichotomous dependent variables that has various desirable properties that earlier proposals lack. Of the measures examined in this paper, the new proposal is the only one that conforms with classical R^2 in terms of both its range and its relationship with the underlying test statistics.

The new measure is like an R^2 in that it is contained in the unit interval and has suitable interpretations at the endpoints of the interval. In addition, unlike the earlier proposals, its marginal relationship with the average likelihood ratio statistic is closely in line with similar relationships between R^2 and the classical tests in the linear model. Thus, the behavior of this new measure in the interior of the unit interval is more intuitively interpretable and is consistent with the statistical significance of hypothesis tests normally associated with R^2.

REFERENCES

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

Anderson, T.W. (1958), An Introduction to Multivariate Statistical Analysis, Wiley.

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

Davidson, R. and McKinnon, J. G. (1993), Estimation and Inference in Econometrics, Oxford.

Dhrymes, P. J. (1986), "Limited Dependent Variables", in Griliches, Zvi and Michael D. Intriligator, eds., Handbook of Econometrics, North Holland.

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

Estrella, A. and Hardouvelis, G. (1991), "The Term Structure as a Predictor of Real Economic Activity", Journal of Finance, 46, 555-576.

Estrella, A. and Mishkin, F. S. (in press), "The Predictive Power of the Term Structure of Interest Rates in Europe and the United States: Implications for the European Central Bank", European Economic Review.

Evans, G. B. A. and Savin, N. E. (1982), "Conflict Among the Criteria Revisited: The W, LR, and LM Tests", Econometrica, 50, 737-748.

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

Lehmann, E. L. (1986), Testing Statistical Hypotheses, Wiley.

Maddala, G. S. (1977), Econometrics, McGraw-Hill.

Maddala, G. S. (1983), Limited-Dependent and Qualitative Variables in Econometrics, Cambridge.

Magee, L. (1990), "R^2 Measures Based on Wald and Likelihood Ratio Joint Significance

    Tests", The American Statistician, 44, 250-253.

McFadden, D. (1974), "Conditional Logit Analysis of Qualitative Choice Behavior", in

    Zarembka, P. (ed.), Frontiers in Econometrics, Academic Press.

Morrison, D. G. (1972), "Upper Bounds for Correlations Between Binary Outcomes and

    Probabilistic Predictions", Journal of the American Statistical Association, 7, 68-70.

Rao, C. R. (1973), Linear Statistical Inference and Its Applications, Wiley.

Theil, H. (1971), Principles of Econometrics, Wiley.

Veall, M. R. and Zimmermann, K. F. (1992), "Pseudo-R^2's in the Ordinal Probit Model",

    Journal of Mathematical Sociology, 16, 333-342.

Windmeijer, F. A. G. (1995), "Goodness-of-Fit Measures in Binary Choice Models",

    Econometric Reviews, 14, 101-116.

Table 1. Properties of Likelihood-Based Measures

Measure	As function of A and B	Value at A=0	Value at A=B	Derivative at A=0
ϕ_0	$1-(1-A/B)^B$	0	1	1
ϕ_1	A/B	0	1	$1/B$
ϕ_2	$1-\exp(-A)$	0	$1-\exp(-B)$ $\leq .75$	1
ϕ_3	$\dfrac{1-\exp(-A)}{1-\exp(-B)}$	0	1	$(1-\exp(-B))^{-1}$ ≥ 1.33
ϕ_4	$A/(A+1)$	0	$B/(B+1)$ $\leq .581$	1
ϕ_5	$\dfrac{A/(A+1)}{B/(B+1)}$	0	1	$(B+1)/B$ ≥ 1.72

Table 2. Critical Values of Odds Ratio and Corresponding Measures of Fit

$\overline{x}$	$\overline{y}$	r*	ϕ_0	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5
.05	.05	9.87	.032	.079	.031	.095	.031	.107
.05	.1997	5.77	.034	.034	.033	.052	.033	.065
.05	.5	7.94	.034	.025	.034	.045	.033	.057
.5	.05	7.94	.035	.086	.034	.102	.033	.116
.5	.1997	2.71	.037	.037	.037	.058	.036	.072
.5	.5	2.18	.037	.027	.036	.049	.036	.062

Note: r* is the critical value of the odds ratio for a significance level of .00001. Results are based on 500 observations and the measures of fit are derived from a logit model.

Table 3. Comparison of Actual (χ^2) and Implicit (F) Significance Levels

$\overline{y}$	χ^2 Level	ϕ_0	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5
				F Level			
.05	.01	.010	.000	.010	.000	.010	.000
.05	.05	.049	.002	.051	.001	.051	.000
.05	.10	.100	.009	.101	.004	.101	.002
.5	.01	.010	.029	.010	.003	.010	.001
.5	.05	.050	.096	.051	.024	.051	.010
.5	.10	.101	.163	.101	.058	.101	.031

Table 4. Monte Carlo Comparisons of ϕ_0 and ϕ_7

Ey	$E\phi_0$	$p(\phi_0)$	$p(\phi_7)$	$p(\phi_7)\text{-}p(\phi_0)$	t statistic	p value
.05	.1	.064	.110	.046	5.21	.000
.05	.3	.004	.008	.004	2.00	.023
.05	.5	.000	.000	.000	na	na
.2	.1	.052	.060	.008	1.64	.051
.2	.3	.003	.002	-.001	-1.00	.159
.2	.5	.000	.000	.000	na	na
.5	.1	.061	.064	.003	0.73	.233
.5	.3	.002	.002	.000	na	na
.5	.5	.000	.000	.000	na	na

Table 5: Numerical Illustration of Measures of Fit

|	Statistic	France	Germany	Italy	UK	US
	n	192	252	216	252	252
	$\bar{y}$	0.573	0.488	0.444	0.484	0.179
	A	0.000	0.628	0.075	0.200	0.356
	B	1.365	1.386	1.374	1.385	0.938
Measures of fit:						
	ϕ_0	0.000	0.567	0.074	0.194	0.361
	ϕ_1	0.000	0.454	0.055	0.144	0.379
	ϕ_2	0.000	0.467	0.072	0.181	0.299
	ϕ_3	0.000	0.622	0.097	0.241	0.492
	ϕ_4	0.000	0.386	0.070	0.166	0.262
	ϕ_5	0.000	0.664	0.121	0.287	0.542
	ϕ_6	0.000	0.496	0.067	0.177	0.347
	ϕ_7	0.000	0.496	0.067	0.177	0.346
	ϕ_8	0.000	0.464	0.071	0.181	0.340
	MZ	0.000	0.704	0.119	0.283	0.514
	ϕ_0^a	-0.010	0.561	0.065	0.186	0.352
	ϕ_0^b	-0.005	0.565	0.070	0.191	0.358

Figure 1. Probability of Recession Using Yield Curve Spread (t-12). Monthly Data, January 1973 to December 1994. This figure provides a visual illustration of the fit of the equation for predicting recessions in each of the five countries in the sample. The predictor is the yield curve spread twelve months earlier. Shaded regions indicate recessions.
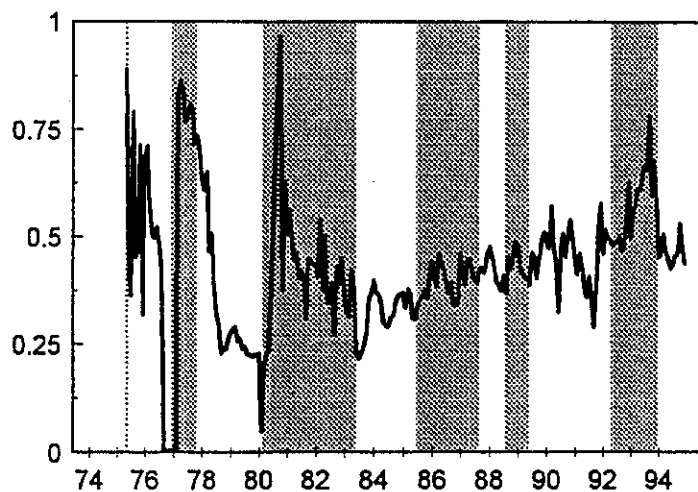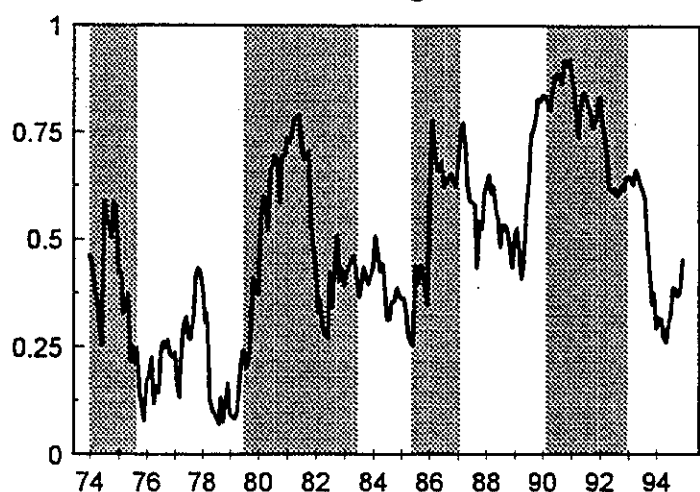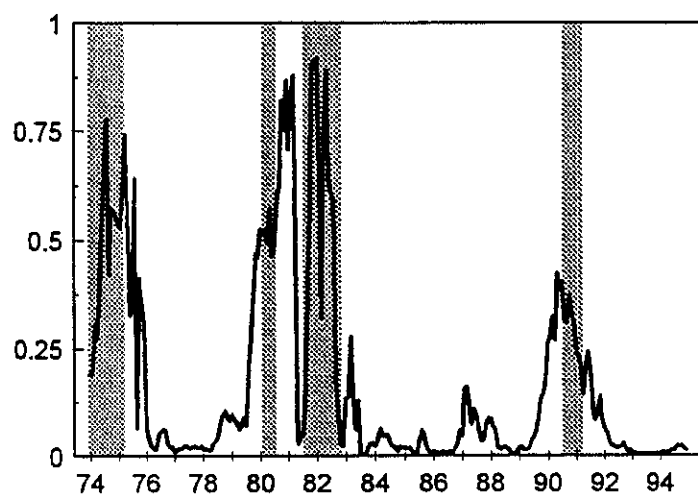
France
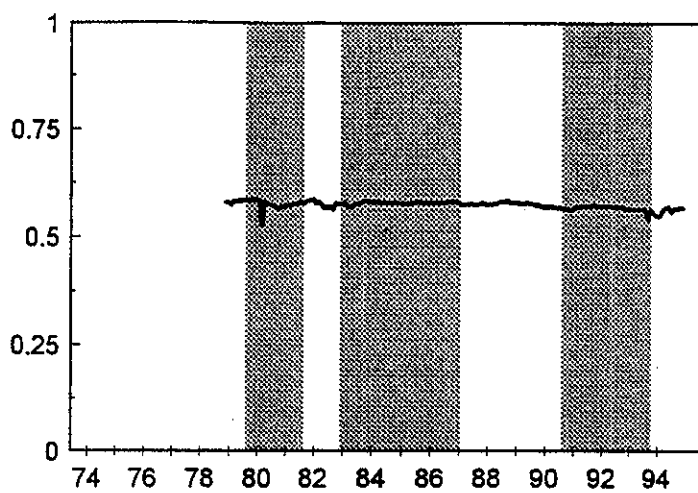
Germany
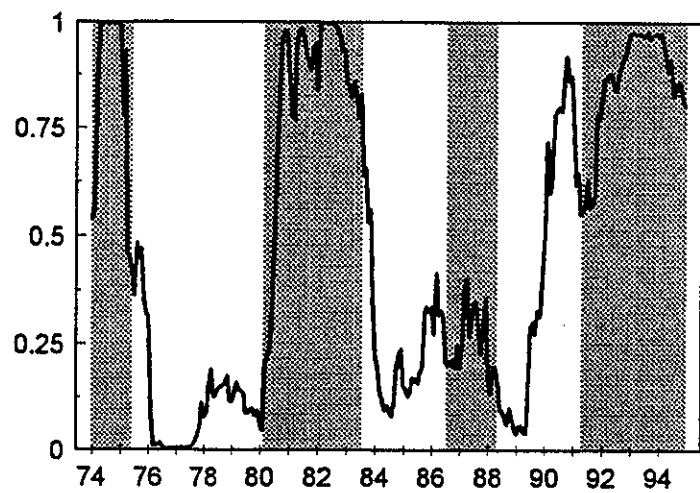
Italy

United Kingdom

United States

FEDERAL RESERVE BANK OF NEW YORK
RESEARCH PAPERS
1997

The following papers were written by economists at the Federal Reserve Bank of New York either alone or in collaboration with outside economists. Single copies of up to six papers are available upon request from the Public Information Department, Federal Reserve Bank of New York, 33 Liberty Street, New York, NY 10045-0001 (212) 720-6134.

9701. Chakravarty, Sugato, and Asani Sarkar. "Traders' Broker Choice, Market Liquidity, and Market Structure." January 1997.

9702. Park, Sangkyun. "Option Value of Credit Lines as an Explanation of High Credit Card Rates." February 1997.

9703. Antzoulatos, Angelos. "On the Determinants and Resilience of Bond Flows to LDCs, 1990 - 1995: Evidence from Argentina, Brazil, and Mexico." February 1997.

9704. Higgins, Matthew, and Carol Osler. "Asset Market Hangovers and Economic Growth." February 1997.

9705. Chakravarty, Sugato, and Asani Sarkar. "Can Competition between Brokers Mitigate Agency Conflicts with Their Customers?" February 1997.

9706. Fleming, Michael, and Eli Remolona. "What Moves the Bond Market?" February 1997.

9707. Laubach, Thomas, and Adam Posen. "Disciplined Discretion: The German and Swiss Monetary Targeting Frameworks in Operation." March 1997.

9708. Bram, Jason, and Sydney Ludvigson. "Does Consumer Confidence Forecast Household Expenditure: A Sentiment Index Horse Race." March 1997.

9709. Demsetz, Rebecca, Marc Saidenberg, and Philip Strahan. "Agency Problems and Risk Taking at Banks." March 1997.

9710. Lopez, Jose. "Regulatory Evaluation of Value-at-Risk Models." March 1997.

9711. Cantor, Richard, Frank Packer, and Kevin Cole. "Split Ratings and the Pricing of Credit Risk." March 1997.

9712. Ludvigson, Sydney, and Christina Paxson. "Approximation Bias in Linearized Euler Equations." March 1997.

9713. Chakravarty, Sugato, Sarkar, Asani, and Lifan Wu. "Estimating the Adverse Selection Cost in Markets with Multiple Informed Traders." April 1997.

9714. Laubach, Thomas, and Adam Posen. "Some Comparative Evidence on the Effectiveness of Inflation Targeting." April 1997.

9715. Chakravarty, Sugato, and Asani Sarkar. "A General Model of Brokers' Trading, with Applications to Order Flow Internalization, Insider Trading and Off-Exchange Block Sales." May 1997.

To obtain more information about the Bank's *Research Papers* series and other publications and papers, visit our site on the World Wide Web (http://www.ny.frb.org). From the research publications page, you can view abstracts for *Research Papers* and *Staff Reports* and order the full-length, hard copy versions of them electronically. Interested readers can also view, download, and print any edition in the *Current Issues in Economics and Finance* series, as well as articles from the *Economic Policy Review.*