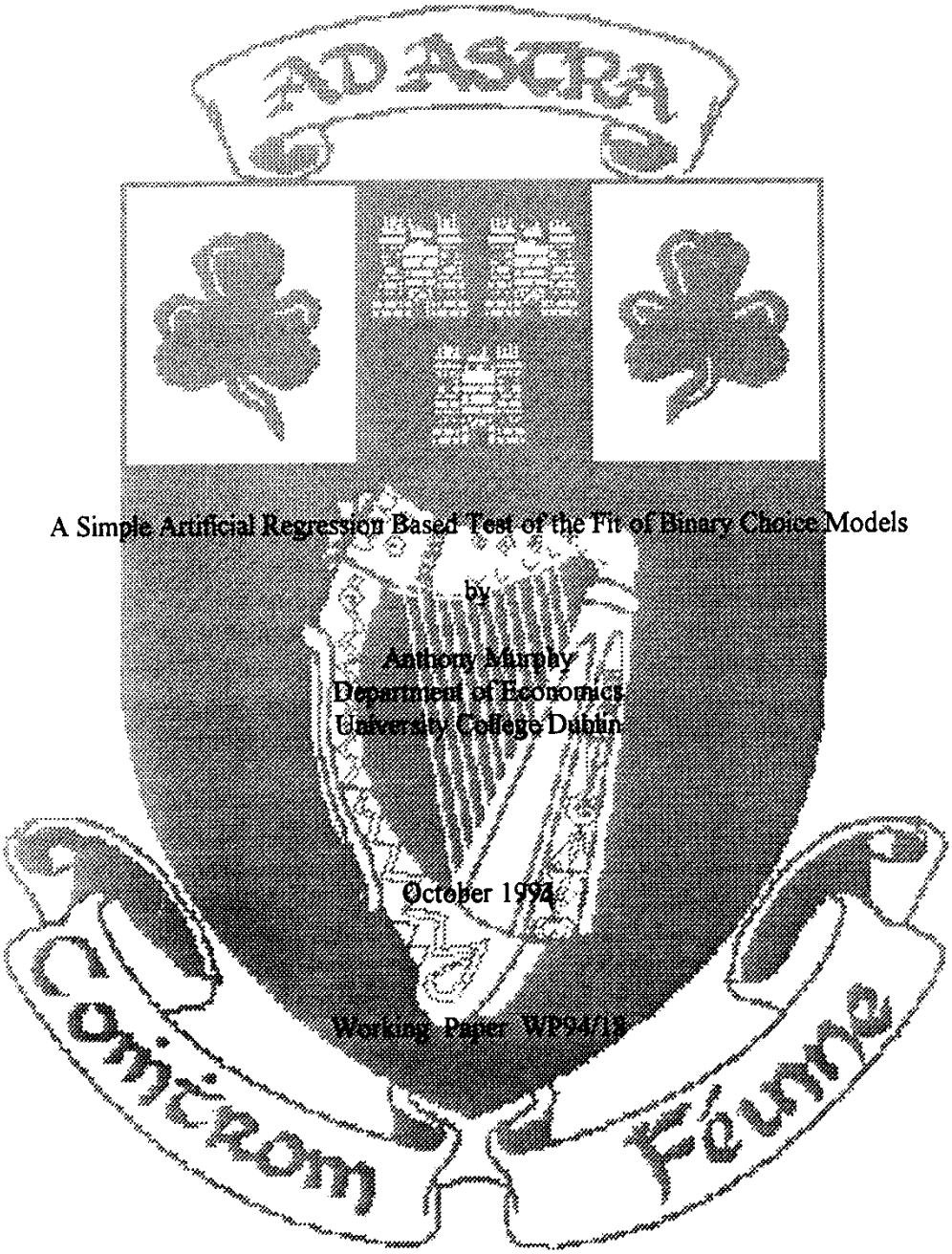
Title	A simple artificial regression based test of the fit of binary choice models
Authors(s)	Murphy, Anthony
Publication date	1994-10
Publication information	Murphy, Anthony. “A Simple Artificial Regression Based Test of the Fit of Binary Choice Models” (October, 1994).
Series	UCD Centre for Economic Research Working Paper Series, WP94/18
Publisher	University College Dublin. School of Economics
Item record/more information	http://hdl.handle.net/10197/1761
Notes	A hard copy is available in UCD Library at GEN 330.08 IR/UNI

CENTRE FOR ECONOMIC RESEARCH
DEPARTMENT OF ECONOMICS



A Simple Artificial Regression Based Test of the Fit of Binary Choice Models

by

Anthony Murphy
Department of Economics
University College Dublin

October 1994

Working Paper WP94/18

DEPARTMENT OF ECONOMICS
UNIVERSITY COLLEGE DUBLIN
BELFIELD DUBLIN 4

A SIMPLE ARTIFICIAL REGRESSION BASED TEST OF THE FIT OF BINARY CHOICE MODELS*

Anthony Murphy.

Department of Economics,
University College Dublin,
Dublin 4, Ireland.

July 1994

Abstract: A simple artificial regression based test of the fit of binary choice models is derived. The test statistic is likely to have reasonable small sample propoerties since it is not based on the outer product gradient form of the conditional moment test.

Keywords: Binary Choice Models, Fit, Conditional Moment Test, Outer Product Gradient Form.

JEL No: C35.

Introduction

Very often one wishes to examine the fit of binary choice models. Informal tests generally involve comparing actual and predicted aggregate choice shares when the sample dataset is divided into disjoint cells or strata which are not based on the choices made. Horowitz (1985) formalizes this procedure and developed a test of the fit of discrete choice models[1]. However his derivations are rather complicated.

In this paper a simple artificial regression based version of the Horowitz test of the fit of binary choice models is derived. Following a suggestion in Pagan and Vella (1989), the conditional moment framework of Newey (1985) and Tauchen (1985) is used. This approach greatly simplifies the derivation of the test statistic. However, unlike Pagan and Vella, the outer product gradient form of the test statistic is not used ie. the information matrix is not approximated by the outer product of the matrix of contributions to the score. Instead it is calculated as the expectation of this outer product. Davidson and MacKinnon

---

*This paper was written up whilst attending the Warwick Summer Research Workshop. The Workshop was supported by the Economic and Social research Council (UK) and the Human Capital and Mobility Programme of the European Community.

[1] Heckman (1984) derives a χ^2 test of fit when the sample is divided into disjoint cells using intervals of the range of y where y may be grouped etc. Andrews (1988a, 1988b) derives χ^2 tests of fit under very general conditions.

(1984b) and Engle (1984) use a similar approach[2].

A Binary Choice Model

Consider a binary choice model. With a random sample of N individuals, the log likelihood is:

$$l(\theta) = \sum_i \{y_i \ln p_i + (1-y_i)\ln(1-p_i)\} \tag{1}$$

where the subscript i denotes individuals, y_i is an indicator variable ie. y_i is 1 if i is successful and 0 otherwise and p_i is the probability of success, which depends on the K dimensional parameter vector θ. The score is:

$$\begin{aligned} s(\theta) &= \frac{\delta l(\theta)}{\delta\theta} = \sum_i s_i(\theta) \\ &= \sum_i \frac{y_i - p_i}{p_i(1-p_i)} \frac{\delta p_i}{\delta\theta} \end{aligned} \tag{2}$$

and the information matrix is:

$$\begin{aligned} I(\theta) &= \lim_{N\to\infty} E \frac{1}{N} \frac{\delta l}{\delta\theta} \frac{\delta l}{\delta\theta'} = \lim_{N\to\infty} E \frac{1}{N} \sum_i s_i(\theta) s_i(\theta)' \\ &= \lim_{N\to\infty} \sum_i \frac{1}{p_i(1-p_i)} \frac{\delta p_i}{\delta\theta} \frac{\delta p_i}{\delta\theta'} \end{aligned} \tag{3}$$

The observed score and information matrix are:

$$\begin{aligned} s(\hat\theta) &= \sum_i s_i(\hat\theta) = \sum_i \frac{y_i - \hat p_i}{\hat p_i(1-\hat p_i)} \frac{\delta \hat p_i}{\delta\hat\theta} = 0 \\ I(\hat\theta) &= \frac{1}{N} \sum_i \frac{1}{\hat p_i(1-\hat p_i)} \frac{\delta \hat p_i}{\delta\hat\theta} \frac{\delta \hat p_i}{\delta\hat\theta'} \end{aligned} \tag{4}$$

where $\hat p_i$ is estimated probability of success which depends on the estimated parameter vector $\hat\theta$. The model is assumed to be correctly specified and the

[2] Many papers have shown that the outer product gradient (OPG) from of the Lagrange Multiplier (LM) and other test statistics, such as the information matrix, have poor small sample properties. See Chesher and Spady (1991), Davidson and MacKinnon (1983, 1984a, 1993), Godfrey (1988), MacKinnon (1992) and Orme (1990) for example.

information matrix is assumed to be non-singular. Under standard regularity conditions the asymptotic distribution of $\hat{\theta}$ is:

$$\sqrt{N}(\hat{\theta}-\theta) \underset{a}{=} I(\theta)^{-1}\frac{1}{\sqrt{N}}\sum_i s_i(\theta)$$
$$\underset{a}{\sim} N(0,I(\theta)^{-1}) \tag{5}$$

Prediction Error

Let the sample be stratified into C disjoint cells and let d_{ic} be an indicator of whether or not individual i is in cell c. The number of individuals in cell c is N_c. Let $n_c = N_c/N$ be the share of observations in cell c. To avoid unnecessary detail assume that the sample and population cell shares are the same[3].

In cell c the average "error" m_c is:

$$m_c(\hat{\theta}) = \frac{1}{N_c}\sum_i d_{ic}(y_i-\hat{p}_i) = \frac{1}{N}\sum_i n_c d_{ic}(y_i-\hat{p}_i)$$
$$= \frac{1}{N}\sum_i m_{ic}(\hat{\theta}) \tag{6}$$

Intuitively, if the model fits "well" then the average prediction error in all C cells should be "small". However the distribution of the m_c's is required in order to derive a test statistic.

Stack the m_c's into the C dimensional column vector m. Then:

$$\sqrt{N}\, m(\hat{\theta}) = \frac{1}{\sqrt{N}}\sum_i m_i(\hat{\theta})$$
$$\underset{a}{=} \frac{1}{\sqrt{N}}\sum_i m_i(\theta) + M(\theta)\sqrt{N}(\hat{\theta}-\theta) \tag{7}$$

where:

$$M(\theta) = \lim_{N\to\infty} E\frac{1}{N}\sum_i \frac{\delta m_i(\theta)}{\delta\theta'}$$

The cj'th element of the C by K dimensional matrix M is:

---

[3] Strictly the sample n_c must converge to the population n_c so that one may use the population value in all the asymptotic results.

$$\lim_{N\to\infty} E\frac{1}{N}\sum_i \frac{\delta m_{ic}(\theta)}{\delta\theta_j} = -\lim_{N\to\infty} E\frac{1}{N}\sum_i n_c d_{ic}\frac{\delta p_i}{\delta\theta_j}$$

The Asymptotic Distribution of $m(\hat{\theta})$

Substituting (5) into (7) yields:

$$\sqrt{N}m(\hat{\theta}) \underset{a}{=} \frac{1}{\sqrt{N}}\sum_i m_i(\theta) + M(\theta)I(\theta)^{-1}\frac{1}{\sqrt{N}}\sum_i s_i(\theta) \tag{8}$$

from which one can derive the asymptotic distribution of $m(\hat{\theta})$:

$$\sqrt{N}m(\hat{\theta}) \underset{a}{\sim} N(0,V) \tag{9}$$

where:

$$V = V_{mm} + M(\theta)I(\theta)^{-1}M(\theta)' + M(\theta)\ I(\theta)^{-1}V_{sm} + V_{ms}I(\theta)^{-1}M(\theta)'$$

and:

$$V_{mm} = \lim_{N\to\infty} E\frac{1}{N}\sum_i m_i(\theta)m_i(\theta)'$$

$$V_{sm} = \lim_{N\to\infty} E\frac{1}{N}\sum_i s_i(\theta)m_i(\theta)'$$

This expression may be simplified further using the generalised information matrix identity. Since the expectation of $m(\theta)$ is zero, one can show that:

$$M(\theta) = -V_{ms}$$

Thus the asymptotic variance-covariance matrix of $m(\hat{\theta})$ is:

$$V = V_{mm} - V_{ms}I(\theta)^{-1}V_{sm} \tag{10}$$

and the natural test statistic of the fit of the model is:

$$\tau = Nm(\hat{\theta})'\hat{V}^{-1}m(\hat{\theta}) \tag{11}$$

where $\hat{V}$ is a consistent estimate of V. Under standard regularity conditions, τ is distributed as χ^2 with C degrees of freedom under the null hypothesis that the

model is correctly specified within each cell.

Artificial Regression Based Test Statistic

The test statistic may be simply calculated using an artificial regression since:

$$m(\hat{\theta}) = \frac{1}{N}\sum_i m_i(\hat{\theta}) = \frac{1}{N}\sum_i \hat{r}_i\hat{z}_i \tag{12a}$$

consistent estimates of V_{mm} and V_{ms} are:

$$\hat{V}_{mm} = \frac{1}{N}\sum_i \hat{z}_i\hat{z}_i'$$

$$\hat{V}_{ms} = \frac{1}{N}\sum_i \hat{z}_i\hat{x}_i' \tag{12b}$$

and a consistent estimate of the information matrix is:

$$I(\hat{\theta}) = \frac{1}{N}\sum_i \hat{x}_i\hat{x}_i' \tag{12c}$$

where:

$$\hat{r}_i = \frac{y_i - \hat{p}_i}{\sqrt{\hat{p}_i(1-\hat{p}_i)}}$$

$$\hat{x}_i = \frac{1}{\sqrt{\hat{p}_i(1-\hat{p}_i)}}\frac{\delta p_i}{\delta\hat{\theta}} \tag{13}$$

$$\hat{z}_i = \sqrt{\hat{p}_i(1-\hat{p}_i)}\,w_i$$

and the ic'th element of w_i is $n_c d_{ic}$. The scalar $\hat{r}_i$ is a scaled residual ie. it has a zero mean and a unit variance. $\hat{x}_i$ and $\hat{z}_i$ are K and C dimensional vectors. Thus a consistent estimate of $\hat{V}$ is:

$$\hat{V} = \hat{V}_{m'm} - \hat{V}_{ms}I(\hat{\theta})^{-1}\hat{V}_{sm}$$

$$= \frac{1}{N}\sum_i \hat{z}_i\hat{z}_i' - \frac{1}{N}\sum_i \hat{z}_i\hat{x}_i'\left(\frac{1}{N}\sum_i \hat{x}_i\hat{x}_i'\right)^{-1}\frac{1}{N}\sum_i \hat{x}_i\hat{z}_i' \tag{14}$$

and the test statistic for the fit of the model may be expressed as:

$$\tau = \sum_i \hat{r}_i\hat{z}_i'(\sum_i \hat{z}_i\hat{z}_i' - \sum_i \hat{z}_i\hat{x}_i'(\sum_i \hat{x}_i\hat{x}_i')^{-1}\sum_i \hat{x}_i\hat{z}_i')^{-1}\sum_i \hat{z}_i\hat{r}_i \qquad (15)$$

This is simply the explained sum of squares from the regression of $\hat{r}_i$ on $\hat{x}_i$ and $\hat{z}_i$ since the likelihood equation is:

$$s(\hat{\theta}) = \sum_i s_i(\hat{\theta}) = \sum_i \hat{r}_i\hat{x}_i = 0$$

Conclusion

A test of the fit of any binary choice model is very simply calculated using an artificial regression. The derivation is simpler than in Horowitz (1985) and the test statistic is likely to have reasonable small sample properties since the outer product gradient form of the conditional moment test is not used.

The approach used in this paper may be extended to testing the fit of general discrete choice models. However the form of the artificial regression is different.[4] The generalized inverse of $\hat{V}$ must be used, since it is generally singular, or else the fit in some arbitrary choice category is ignored .

References

Andrews, D.W. (1988a). "Chi-Square Diagnostic Tests for Econometric Models", *Journal of Econometrics,* 37, 135-56.

Andrews, D.W. (1988b). "Chi-Square Diagnostic Tests for Econometric Models Theory", *Econometrica*, 56, 1419-53.

Chesher, A. and R. Spady (1991). "Asymptotic Expansions of the Information Matrix Test Statistic", *Econometrica*, 59, 787-815.

Davidson, R. and J.G. MacKinnon (1983). "Small Sample Properties of Alternative Forms of the Lagrange Multiplier Test", *Economics Letters,* 12, 269-75.

Davidson, R. and J.G. MacKinnon (1984a). "Model Specification Tests Based on Artificial Linear Regressions", *International Economic Review,* 25, 485-502.

Davidson, R. and J.G. MacKinnon (1984). "Convenient Specification Tests for the Logit and Probit Models", *Journal of Econometrics,* 25, 241-62.

Davidson, R. and J.G. MacKinnon (1993). *Estimation and Inference in Econometrics*, Oxford, Oxford University Press.

Engle, R.F. (1984). "Wald, Likelihood Ratio and Lagrange Multiplier Tests in Econometrics" Ch. 13 in *Handbook of Econometrics,* Vol. II, eds. Z. Griliches and M.D. Intriligator, 775-826. Amsterdam, North-Holland.

---

[4] See Murphy (1994) for details of the artifical regression used in general discrete choice models.

Godfrey, L.G. (1988). *Mis-Specification Tests in Econometrics,* Cambridge, Cambridge University Press.

Heckman, J. (1984). "The χ^2 Goodness of Fit Statistics for Models with Parameters Estimated from Microdata," *Econometrica*, 52, 1543-47.

Horowitz, J. (1985). "Testing Probabilistic Discrete Choice Models of Travel Demand by Comparing Predicted and Observed Aggregate Choice Shares", *Transport Research B*, 19, 17-38.

MacKinnon, J.G. (1992). "Model Specification Tests and Artificial Regressions", *Journal of Economic Literature*, XXX, 102-46.

Murphy, A. (1994). "Artificial Regression Based Mis-Specification Tests for Discrete Choice Models", mimeo.

Newey, W.K. (1985). "Maximum Likelihood Specification Testing and Conditional Moments Tests", *Econometrica*, 53, 1047-70.

Orme, C. (1990). "The Small Sample Properties of the Information Matrix Test", *Journal of Econometrics*, 46, 309-31.

Pagan, A.R. and F. Vella (1989). "Diagnostic Tests for Models Based on Individual Data: A Survey", *Journal of Applied Econometrics,* 4, S29-59.

Tauchen, G.E. (1985). "Diagnostic Testing and Evaluation of Maximum Likelihood Models", *Journal of Econometrics,* 30, 415-43.