The Economic and Social Review, Vol. 41, No. 1, Spring, 2010, pp. 77–85

When do Probit Residuals Sum to Zero?

DENIS CONNIFFE*
National University of Ireland, Maynooth

Abstract: Probit residuals need not sum to zero in general. However, if explanatory variables are qualitative the sum can be shown to be zero for many models. Indeed this remains true for binary dependent variable models other than Probit and Logit. Even if some explanatory variables are quantitative, residuals can sum to almost zero more often than might at first seem plausible.

I INTRODUCTION

This brief article is motivated by a comment and footnote in Green (2008, p. 778). He shows that in a Logit model the sum of the predicted probabilities must equal the sum of unit values of the binary dependent variable and goes on to note "… although regularly observed in practice, the result has not been verified for the Probit model". Some other authors, for example, Verbeek (2004) appeal to the similarity of the normal and logistic distributions to explain why Probit residuals may sum to almost zero. Section III of this paper

*I am grateful to an anonymous referee for helpful suggestions. I also thank the audience at the Irish Economic Association conference, April 2009, for useful comments and in particular Karl Whelan who drew attention to similarities between the Probit and non-linear modelling given categorical explanatory variables.

Current address: School of Economics, University College Dublin.

Paper delivered at the Twenty-Third Annual Conference of the Irish Economic Association, Blarney, Co Cork, April 24-26, 2009.

shows that for qualitative explanatory variables with nested structure the sum is exactly zero. This also holds for crossed structure with interactions specified in the model. For models which are correctly specified without interactions the result will hold almost exactly, but for models incorrectly omitting interactions the discrepancy is greater. Section IV examines why the residuals may often sum to near zero even with quantitative explanatory variables in a model.

II MODELING BINARY DEPENDENT VARIABLES

The usual approach in relating occurrence of a binary dependent variable Z_i to a vector of explanatory variables x_i is to write

$$\text{Prob}\ (Z_i = 1) = F(x_i' b),$$

where F is some continuous probability distribution. Common choices for F are the normal or the logistic or, less frequently, the Gumbel. The maximum likelihood equations for the parameter vector b can be written in various ways, but the form most appropriate for this paper is

$$\sum_{i=1}^{n}\left[Z_i - F(M_i)\right]\frac{f(M_i)}{F(M_i)\,[1-F(M_i)]}x_i = 0, \tag{1}$$

where n is the sample size, $M_i = x_i' b$ and f is the probability density. It is given, for example, in Verbeek (2004, Equation 7.13, p. 193). Assuming the model contains a constant, a say, one equation is

$$\sum_{i=1}^{n}\left[Z_i - F(M_i)\right]\frac{f(M_i)}{F(M_i)\,[1-F(M_i)]} = 0. \tag{2}$$

The predicted probability for observation i is $F(M_i)$[1] and so

$$\sum_{i=1}^{n}\left[Z_i - F(M_i)\right] = 0 \tag{3}$$

is required if the sum of predicted is to equal the sum, r say, of unit values. This will be the case if

[1] Strictly, the predicted probability is $F(\widetilde{M}_i)$ where $\widetilde{M}_i = x_i'\widetilde{b}$, but it is tidier to avoid circumflexes.

$$\frac{f(M_i)}{F(M_i)[1-F(M_i)]}$$

is constant so that (2) implies (3). It is easily seen that if F is the logistic distribution

$$F(M_i) = \Lambda(M_i) = \frac{1}{1+e^{-M_i}}$$

then $f(M_i) = F(M_i)[1 - F(M_i)]$. However, this seems specific to the logistic. Even choosing the uniform distribution, or linear probability model,

$$f(M_i) = 0 \text{ for } M_i < 0, \quad f(M_i) = M_i \text{ for } 0 \leq M_i \leq 1 \text{ and } f(M_i) = 0 \text{ for } M_i > 1,$$

does not generally give (3).[2] But why then is (3) frequently seen to hold exactly for Probit analysis and, if not exact, very often almost so?

III DUMMY EXPLANATORY VARIABLES

Regressions with a binary dependent variable often involve explanatory variables that are qualitative, or categorical, in nature and some analyses may have only such explanatory variables. Then (3) can be shown to hold for a much wider class of distributions than the logistic. Suppose a qualitative explanatory factor has k categories, so that it is modeled by k-1 dummy variables. Observations for a particular category have values of unity in the corresponding dummy variable and zero in all other dummy variables. (Of course the intercept, identified with the first category, has the usual associated variable which is always unity.) Equations (1) now becomes

$$\sum_{(1)} [Z_i - F(M_1)] \frac{f(M_1)}{F(M_1)[1-F(M_1)]} = 0$$

$$\sum_{(2)} [Z_i - F(M_2)] \frac{f(M_2)}{F(M_2)[1-F(M_2)]} = 0$$

. . . .

[2] Of course, the OLS estimate of the linear probability model, which ignores heteroscedasticity, does make residuals sum to zero.

. . . .

$$\sum_{(k-1)} [Z_i - F(M_{k-1})] \frac{f(M_{k-1})}{F(M_{k-1})[1-F(M_{k-1})]} = 0,$$

where (p) implies summation over observations where the pth dummy variable equals unity and M_p is the constant $a + b_p$. Adding equations and subtracting from (2) gives

$$\sum_{(0)} [Z_i - F(M_0)] \frac{f(M_0)}{F(M_0)[1-F(M_0)]} = 0,$$

where (0) implies summation over observations in the first category and $M_p = a$. Since each M_p is a constant, the sum of $Z - F$ values within each category equals zero and so (3) holds.

Several Qualitative Variables with Hierarchical Structure

Now there are sets of observations that will have unity values simultaneously in several dummy variables, but these occur in a nested manner. For example, if comparing the employability of immigrants to Ireland to that of nationals, unity values of a dummy variable could designate immigrants. But a further dummy variable could distinguish non-EU from EU immigrants. So if someone had a unity in the non-EU dummy, they must also have had a unity in the immigrant dummy. For variables in the primary level of the hierarchy (immigrant v national in the example) there are again the equations

$$\sum_{(p)} [Z_i - F(M_p)] \frac{f(M_p)}{F(M_p)[1-F(M_p)]} = 0 \tag{4}$$

and again adding equations and subtracting from (2) gives

$$\sum_{(0)} [Z_i - F(M_0)] \frac{f(M_0)}{F(M_0)[1-F(M_0)]} = 0.$$

However, all the M_p within each Equation (4) are not equal because the secondary levels will change within primary levels. In the immigrant example there can be an additional non-EU effect for that subgroup so that $M = a + b_{IM} + c_{NEU}$ compared to $M = a + b_{IM}$ for EU immigrants. But clearly there will also be a summation that equates to zero for each group of observations defined by

unity in both the primary and secondary level dummies. In the example the non-EU immigrants is such a group. For a two level hierarchy

$$\sum_{(p(q))} [Z_i - F(M_{p(q)})] \quad \frac{f(M_{p(q)})}{F(M_{p(q)})[1 - F(M_{p(q)})]} = 0 \tag{5}$$

where (p(q)) implies summation over observations where the qth secondary level dummy variable within the pth primary level is unity and $M_{p(q)}$ is constant. Summation over the secondary level categories and subtraction from (4) gives

$$\sum_{(p(0))} [Z_i - F(M_{p(0)})] \frac{f(M_{p(0)})}{F(M_{p(0)})[1 - F(M_{p(0)})]} = 0.$$

All the observations are now in groups within which (5) holds with $M_{p(q)}$ constant and so (3) holds. The argument clearly extends to any level of hierarchy.

Qualitative Variables with Crossed Structure and Interactions in Models

The data structure may feature qualitative variables in a crossed rather than nested structure. For example, immigrants and nationals may be also be classified by gender. Suppose one qualitative explanatory factor has k categories and another has m categories. If the full model with interactions is fitted there are $km - 1$ dummy variables, of which $k - 1$ correspond to the main effect of factor 1, $m - 1$ for the main effect of factor 2 and $(k - 1)(m - 1)$ for their interaction with these being the products of the main effect dummies. For observations where a particular interaction dummy takes the value unity

$$\sum_{(p,\,q)} [Z_i - F(M_{pq})] \frac{f(M_{pq})}{F(M_{pq})[1 - F(M_{pq})]} = 0, \tag{6}$$

where (p, q) implies summation over observations where the product of the pth dummy variable for factor 1 with the qth for factor 2 equals unity. M_{pq} is constant made up of an effect of factor 1, factor 2 and an interaction. Corresponding to unity values of the pth dummy variable

$$\sum_{(p)} [Z_i - F(M_p)] \quad \frac{f(M_p)}{F(M_p)[1 - F(M_p)]} = 0, \tag{7}$$

which contains the summations (6) for all q and where M_p changes with q. However, subtracting the terms (6) gives

$$\sum_{(p,0)} [Z_i - F(M_{p0})] \frac{f(M_{p0})}{F(M_{p0})[1-F(M_{p0})]} = 0, \tag{8}$$

where M_{p0} is the intercept plus the coefficient on the pth dummy for factor 1. Also, by difference of (7) from (2),

$$\sum_{(0,0)} [Z_i - F(M_{00})] \frac{f(M_{00})}{F(M_{00})[1-F(M_{00})]} = 0$$

where M_{00} is the intercept. So all observations are divisible into groups within which (6) holds with constant M_{pq}. So (3) follows. The argument obviously extends to multiple factors if all possible interactions of whatever order are specified in the model. In practice, however, it is unusual for all possible interactions to be so included.

Qualitative Variables with Crossed Structure Without Interactions Specified

These findings may no longer apply if interactions are not included in models. For the two factor case it is still true that

$$\sum_{(p)} [Z_i - F(M_p)] \frac{f(M_p)}{F(M_p)[1-F(M_p)]} = 0$$

and

$$\sum_{(q)} [Z_i - F(M_q)] \frac{f(M_q)}{F(M_q)[1-F(M_q)]} = 0.$$

But the summation (6) is contained in both the above and is no longer constrained to zero. The resulting situation is easily illustrated by the immigrant/gender example. The immigrant and gender effects are estimable from the simultaneous equations

$$\sum_{(IM)} [Z_i - F(M_{IM})] \frac{f(M_{IM})}{F(M_{IM})[1-F(M_{IM})]} = 0, \tag{9}$$

where $M_{IM} = a + b_{IM} + b_F$ if the immigrant is female and $M_{IM} = a + b_{IM}$ if the immigrant is male, and

$$\sum_{(F)} [Z_i - F(M_F)] \frac{f(M_F)}{F(M_F)[1-F(M_F)]} = 0, \tag{10}$$

where $M_F = a + b_{IM} + b_F$ if the female is an immigrant and $M_{IM} = a + b_F$ otherwise.

Denoting the ML estimators from (9) and (10) by $\tilde{a}$, $\tilde{b}_{IM}$ and $\tilde{b}_F$, the summation terms common to both (9) and (10) are

$$\frac{f(\tilde{M}_{IM,F})}{F(\tilde{M}_{IM,F})[1-F(\tilde{M}_{IM,F})]} \sum_{(IM,F)} [Z_i - F(\tilde{M}_{IM,F})] \tag{11}$$

where $\tilde{M}_{IM,F} = \tilde{a} + \tilde{b}_{IM} + \tilde{b}_F$. If (11) is zero, divisibility into groups summing to zero and with constant M again occurs and (3) follows. Now if there really is no interaction and the sample size is very large (11) should be nearly zero. This is because (11) is the derivative of the log-likelihood with respect to the interaction parameter, with that parameter set to zero and the nuisance parameters a, b_{IM} and b_F replaced by their estimates under the null hypothesis of no interaction.[3] Then (3) may hold or very nearly do so. But if there is an interaction present, which is not represented in the model, (3) can be expected to differ from zero. Then the discrepancy of the residual sum from zero can be larger, although still not great for Probit analysis for reasons to be discussed in the next Section.

IV RESIDUAL SUM WITH QUANTITATIVE REGRESSORS INCLUDED

First consider a quantitative variable without qualitative variables in the model. While (3) will not follow from (2) except for F logistic, it may be 'nearly' true for Probit analysis. Then $F(M_i) = \Phi(M_i)$, where Φ denotes the standard normal distribution. The quantity

$$\frac{\phi(M_i)}{\Phi(M_i)[1-\Phi(M_i)]} = \Psi(M_i), \text{ say,}$$

is symmetric about zero with minimum there and it increases only slowly for modest departures from zero. To two decimal places $\Psi(0) = 1.60$, $\Psi(.05) = 1.60$, $\Psi(.1) = 1.60$, $\Psi(.15) = 1.60$, $\Psi(.2) = 1.60$, $\Psi(.25) = 1.61$, $\Psi(.3) = 1.61$, $\Psi(.35) = 1.62$, $\Psi(.4) = 1.63$. So for values of M_i likely to produce a good mix of ones and zeros, $\Phi(M_i)$ could be regarded as approximately constant. Of course, this is not true for larger departures from zero. For example, $\Psi(1) = 1.81$ and $\Psi(2) = 4.32$. But big positive M_i would make $\Phi(M_i)$ almost unity and probably ensure Z_i is unity, resulting in a negligible term in (3). Similarly, big negative M_i

[3] That is (11) is a particular case of the Score, or LM, test criterion.

would make $\Phi(M_i)$ negligible and probably make Z_i zero. So it is not surprising that sums of Probit residuals are often near zero. This is not necessarily true for arbitrary F. For example, for a Gumbel the function corresponding to Ψ is not even symmetric about zero.

When there are also qualitative factors in the model the equations of previous sections occur again. Take the case of two crossed factors with interaction. Now

$$\sum_{(p,q)} [Z_i - \Phi(M_{pq}(x_i))] \frac{\phi(M_{pq}(x_i))}{\Phi(M_{pq}(x_i))[1-\Phi(M_{pq}(x_i))]} = 0,$$

where $M_{pq}(x_i) = a + b_p + c_q + d_{pq} + gx_i$. This is constant within the (p, q) cell except for x which can vary only over the range within the cell. A quantitative variable is rarely orthogonal to all qualitative variables and will often have limited range within cells. This, plus the described behavior of $\Phi(M_i)$ will often result in residual sums little different from zero.

V CONCLUDING REMARKS

It should be noted that the arguments in Section III did not assume that $F(M) = \Phi(M)$ and so the findings are not limited to the Probit case. They would apply to binary variables generated from other underlying distributions. However, this may not be of great import since use of other distributions is very rare. The connection to non-linear least squares regression is perhaps more interesting. Non-linear LS fits the model

$$Y_i = W(x_i' b) + e_i,$$

where Y_i is usually continuous, by minimising

$$\sum_{i=1}^{n} \left[Y_i - W(x_i' b)\right]^2,$$

leading to the equations

$$\sum_{i=1}^{n} \left[Y_i - W(x_i' b)\right] W'(x_i' b) x_i = 0.$$

Again if the model contains an intercept one x variable is a vector of units giving

$$\sum_{i=1}^{n} \left[Y_i - W(x_i' b)\right] W'(x_i' b) = 0.$$

The analogies with Equations (1) and (2) are clear and although in general the sum of residuals

$$\sum_{i=1}^{n}\left[Y_i - W(x_i' b)\right]$$

is non-zero, if the explanatory variables are qualitative the arguments of Section III will again apply. However, for multifactor situations specification of all interactions to the highest order is again required to ensure residuals sum to exactly zero. Angrist and Pischke (2008) describe such models as 'saturated', but again they are infrequent in practice. Indeed some classical experimental designs do not replicate the factor combinations and depend on assumptions of the non-existence of higher order interactions to permit estimation of factor effects and standard errors.

Perhaps it should be said that being able to explain when and why residuals sum to zero in non-linear models is perhaps of questionable practical importance. But surely there is some merit in explaining the reasons for a phenomenon noted in the literature on Probit analysis.

REFERENCES

ANGRIST, J. and J-S. PISCHKE, 2008. Princeton: Princeton University Press.
GREEN, W. H., 2008. *Econometric Analysis,* 6th Edition, New Jersey: Pearson.
VERBEEK, M., 2004. *A Guide to Modern Econometrics*, 2nd Edition, Chichester: Wiley.