Redundancy, Unemployment and Manpower Policy: A Comment

T. Lancaster

The Economic Journal, Volume 86, Issue 342 (Jun., 1976), 335-338.

Stable URL:
http://links.jstor.org/sici?sici=0013-0133%28197606%2986%3A342%3C335%3ARUAMPA%3E2.0.CO%3B2-U

The Economic Journal, **86** (*June* 1976), 335–338
Printed in Great Britain

REDUNDANCY, UNEMPLOYMENT AND MANPOWER POLICY: A COMMENT

In a recent article in the ECONOMIC JOURNAL (1972) Mackay and Reid present a study of the effect of a set of variables upon the duration of unemployment of over 600 redundant people. They provide least squares coefficients in the regression of duration of unemployment on variables including age, skill, amount of redundancy pay, etc.; they give estimated standard errors; and they provide, and comment upon, the coefficient of determination, R^2. The R^2 reported is 0·25 and this "relatively low" value they attempt to explain in several ways, one of which is that "The model is trying to explain a complete pattern of human behaviour, and economics as a discipline, and indeed any other of the social sciences, possesses only rudimentary tools for the task. An attempt to marry economics to other disciplines might yield improvements but the most enterprising and ambitious approach so far adopted has produced rather disappointing results." The authors cite several other studies which have produced no greater values for R^2.

This naturally prompts the question, what sort of R^2 values might one expect on data of this type? The answer depends upon the joint probability distribution of the durations of unemployment of each of the individuals in the sample, so what can be said about this, prior to the scrutiny of the data? Mackay and Reid do not state their statistical assumptions but it appears from the procedure they adopt that they believe the durations of unemployment of the persons they studied to be uncorrelated and perhaps also independent. When a large number of men with similar qualifications are competing for a small number of jobs in a limited geographical area the assumption of independence would be unreasonable, but it is fairly plausible for Mackay and Reid's data. Let us adopt it. Assuming independence we need only consider the form of the probability distribution of duration of unemployment – let us call this D – for some particular person. According to the model of Mackay and Reid the expected value of D is a parametrically linear combination of the set of regressors, age, skill, and so on indicated above. What else can be said about the distribution? Finding a new job can be reasonably thought of as a chance or stochastic process, and an enlightening way of thinking about the probabilistic law governing the duration of this process is by asking how the probability that a person will find a new job changes as time, measured from the start of his search, elapses.

The simplest model, and thus a natural starting point for a statistical study, is that in which this probability is independent of time. That is, an unemployed person moves through time with a constant probability, π, per unit time period, of re-entering employment. In this case the form of the probability distribution of D, the duration of the spell of unemployment of a particular person, is the Exponential, with mean $\mu = 1/\pi$ where of course π and μ will vary from person to person. Now while there are certainly objections to be raised *a priori* to this model, and we shall point out some of these later, it may be that the correct model is not too far removed from the Exponential.[1] That is, the probability of finding a new job may change sufficiently slowly that, at least for, say, post-war British and American data, the Exponential model is not too bad a fit.

Let us then suppose that each duration is an observation on an Exponential variate with mean, μ, *correctly* specified as a function of Mackay and Reid's regressor variables. That is, the variation in μ from person to person is correctly described by Mackay and Reid's regression function including their list of regressor variables. It is characteristic of a chance process like this that persons with longer expected durations of unemployment also have more variable durations of unemployment. Of men who can expect to find a new job in 2 days almost all will be back in employment within 8, but of those who can expect to find a new job only in 10 days a few will find one almost right away but some will still be searching after a month – the variance in duration will be much greater for the second group. In fact for the Exponential distribution the variance is equal to the square of the mean, or, in regression terminology, the error variance is equal to the square of the regression function, $\sigma^2 = \mu^2$. Note that in contrast to the standard regression model there is no *independent* error variance parameter whose relative magnitude could, perhaps, be taken as a measure of the success or failure of one's modelling.

What are the implications of this for R^2? R^2 arises out of a partition of the total variance of a set of observations into the variance of the means and the residual variance. If we have a large sample of observations[2] so that the Least Squares estimates of the regression function can be replaced by the true values, the μ_i, then the partition is,

$$\frac{\Sigma(D_i-\bar{\mu})^2}{N} = \frac{\Sigma(\mu_i-\bar{\mu})^2}{N} + \frac{\Sigma(D_i-\mu_i)^2}{N}.$$

The first term on the right-hand side is the variance of the expected durations and the second term is, in expectation, the average of the "error variances". But for Exponential data each "error variance" is simply μ_i^2. Thus, in expectation, the r.h.s. is simply var μ + mean square μ. But R^2 is the ratio of the variance of the means to the total variance so that

$$R^2 = \frac{\text{var } \mu}{\text{var } \mu + \text{mean square } \mu}$$

or, rearranging using the rule that variance = mean square – squared mean,

$$R^2 = \frac{1}{2+(1/c^2)} \leqslant \tfrac{1}{2}, \tag{1}$$

where c is the coefficient of variation of the expected durations of unemployment among the members of the sample. That is, in large samples, the coefficient of determination will be close to a number which depends only on the diversity of

[1] Coincidently, a recent paper in the Economic Journal, Cripps and Tarling (1974), adopts this assumption in a study of the unemployment register.

[2] A proof instead of the following heuristic argument is available on request.

individuals in the sample with respect to their expected periods of unemployment, and which *cannot exceed a half*.[1] Due to the way in which random variation is built into the model, always at least half the realised variation in durations of unemployment will be attributable to chance.

Thus if each individual's unemployment is a "negative exponential" variate, a value of R^2 of less than one half is what one must expect. It is neither a reminder that more explanatory variables should ideally be included, if only one had the statistics to allow this, nor a demonstration that no economic model of this phenomenom can be altogether satisfactory. Indeed Mackay and Reid might well, as we have assumed, have listed *all* the systematic determinants of the variation of the duration of unemployment between the men they studied. The R^2 is irrelevant to the decision about whether the model is correctly specified. The relevant criterion is whether the residuals are consistent with the model assumed.

The result (1) tells one something about the R^2 to be expected in least squares linear regression analysis of duration of unemployment data and is offered here as a possible explanation of why investigators find R^2's of 0·25 or thereabouts. It says that if the data are Exponentially distributed, near enough, and even if the regression function is correctly specified, and if you run a Least Squares multiple regression and read off R^2, you can expect to find, if your sample is large,[2] a number less than a half. If one suspected, before one started, that one was dealing with Exponential data, or data from one of the families suggested in the final paragraph of this note, one would not use ordinary Least Squares and one would not compute R^2, but that is a different point.

Now the Exponential model is open to *a priori* criticism on at least two grounds. First it is widely believed that the experience of unemployment leads people to become less employable, so that for this reason the probability of finding a new job would, at least eventually, tend to fall. Secondly, there is some evidence[3] that, in response to the failure to find a job, people relax their job requirements. This would suggest a rising re-employment probability. More generally, the Exponential model suggests a fairly implausible passivity on the part of the unemployed which is inconsistent with the currently fashionable emphasis on rational search activity. These remarks indicate that one would want to study a likelihood function more flexible than the Exponential but including it as a particular case; the Weibull or Gamma families for example.[4] For these families the variance is proportional to the squared mean and the factor of proportionality will enter into the expression for R^2 analogous to (1).

[1] A value of R^2 of 0·25 would suggest a coefficient of variation of the expected durations of unemployment among the members of the sample of about 0·7.

[2] The sample does not in fact have to be very large for this effect to show up. A short simulation with 2 regressor variables and a constant, a sample size of 40, and means with a coefficient of variation of 0·56 produced, in 400 replications, only two R^2's greater than 0·5.

[3] Holt (1970).

[4] With such models it would be sensible to choose a parametrically exponential regression function and then to begin an efficient estimation procedure from the linear regression analysis of the logarithm of duration.

An alternative approach which might be worth considering is that recently developed by Cox (1972).

Incidentally, if any of the models suggested in this paragraph apply, the standard errors quoted by MacKay and Reid will be inconsistent estimates.

The upper bound may then be greater than or less than a half but always less than one. The general point remains, however; with data whose variances vary as the square of the mean one must expect "low" R^2 figures to be printed out by a Least Squares regression programme.

T. LANCASTER

University of Hull,

Date of receipt of final typescript: August 1975

REFERENCES

Cox, D. R. (1972). "Regression Models and Life Tables." *Journal of the Royal Statistical Society*, Vol. 34, No. 2.

Cripps, T. F. and Tarling, R. J. (1974). "An analysis of the Duration of Male Unemployment in Great Britain 1932–73." ECONOMIC JOURNAL, Vol. 84, No. 334, June.

Holt, C. C. (1970). In *Microeconomic Foundations of Employment and Inflation Theory*, ed. Phelps. Macmillan.

Mackay, D. I. and Reid, G. L. (1972). "Redundancy, Unemployment and Manpower Policy." ECONOMIC JOURNAL, Vol. 82, No. 328, December.