APPENDIX 7

Models with of Multiple Responses

1. Introduction

This chapter reviews some of the methodology which is being used in a statistical investigation of the assessment process associated with GNVQ courses. So far, a limited quantity of data has been acquired; and this has been subjected to a preliminary analysis. The analysis is serving both as a test of the methodology and as a test of the adequacy of the data and the means of collecting it.

2. The Process of Training and Assessment as a Transformation

In order to derive mathematical models which may be used in a statistical investigation of the assessment process, several simplified concepts which are amenable to mathematical representation must be established and adhered to. It should not to be imagined that, in using of such concepts for the purpose of model building, we are proposing an oversimplified view of a complex reality.

The typical GNVQ course may be envisaged as a processes of training and assessment which is applied to individual students. At the beginning of the course, each student is endowed with an educational potential. The course transforms the potential of the student into an assessment rating which, consists, ultimately, of a grade selected from a fourfold classification: *Fail, Pass, Merit* and *Distinction.* It is this transformation of the student's potential into their assessment rating which is the subject of an exercise in mathematical modelling. The transformation is of a random or stochastic nature.

The term random implies no disparagement of the training or of the assessment. It corresponds to the fact that there are numerous determinants of the assessment rating, active throughout the period of the training, which are not under the direct control of the agents of the processes—the teachers and the assessors; and it acknowledges the fact that, to some extent, the assessments themselves are arbitrary and subjective. In the ideal circumstances, the relevant statistical parameters of the transformation will be independent of the location such that the quality of the training and the nature of the assessments can be described as uniform throughout the training establishments. The object of the study is to determine the extent to which the reality of the GNVQ processes departs from such ideal conditions.

In order to evaluate the GNVQ processes, we propose to construct a mathematical model of the statistical relationship linking the potentials of the students to their assessment ratings which are respectively the subjects and the products of the transformation. There is no unique way of describing the potentials of the students; and one of the objects of the study is to identify ways which are both mathematically tractable and statistically robust.

There is some leeway also in the matter of how one chooses to represent the assessment ratings and examination results. The problem of representing the assessments would be greatly simplified if a cardinal index of performance on a scale of 0 to 100 were available for each candidate and if it could be assumed that the index varies continuously and evenly within this range. It may be possible to synthesise an index of this nature by carefully combining the outcomes of the various elements of assessment which contribute to the ultimate grading. However, it seems likely that, for the foreseeable future, those who wish to investigate the GNVQ assessment process will have at their disposal only the summary gradings of the GNVQ units. From the point of view of statistical modelling, this represents the least tractable of the cases which can be envisaged; and therefore it is encouraging to recognise that, given sufficiently abundant data, it can be tackled directly without undue difficulty.

3. A Model with a Bounded Response

One of the mainstays of multivariate statistical modelling is the linear regression model. The model comprises an equation of the form

$$
\begin{aligned}
y &= \alpha + x_1\beta_1 + x_2\beta_2 + \cdots + x_k\beta_k + \varepsilon \\
&= \alpha + x'\beta + \varepsilon
\end{aligned} \tag{1}
$$

which explains the value of a dependent variable y in terms of k observable variables in $x' = [x_1, x_2, \ldots, x_k]$ and an unobservable random variable ε. The latter has a zero expected value $E(\varepsilon) = 0$, and, under standard assumptions, it has a variance or dispersion $V(\varepsilon) = \sigma^2$ which is fixed for all instances of the relationship. The elements of $\beta = [\beta_1, \beta_2, \ldots, \beta_k]'$ are described as the regression parameters, and α is known as the intercept parameter.

The linear regression model has been used by Armitage [2], [3] in an exercise designed to monitor the student assessments of the Open University. In that case, the variable y was an index of performance based on essay marks and examination marks, whilst the explanatory variables $x_1, x_2, \ldots, x_k$ comprised indices of the candidate's past performance, the level of their previous education etc. The linear combination $\xi = x'\beta = x_1\beta_1 + x_2\beta_2 + \cdots + x_k\beta_k$

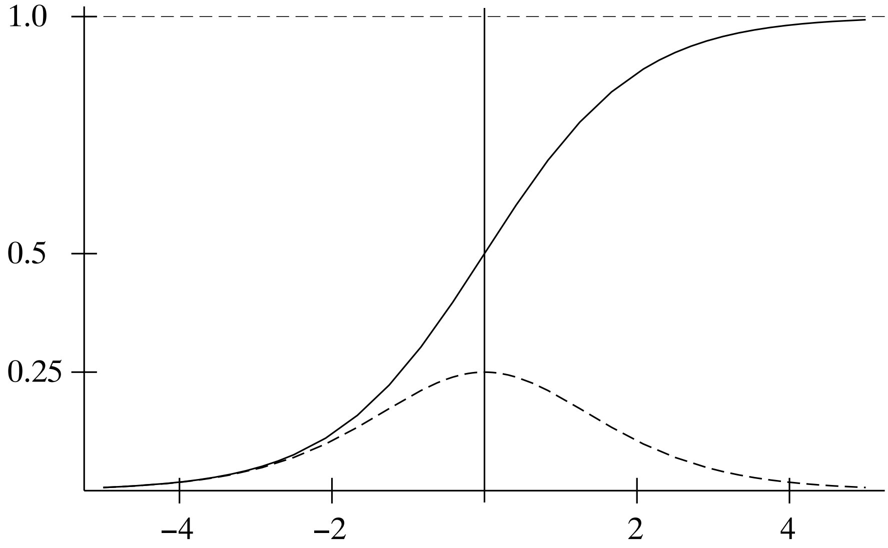

Figure 1. The logistic function $e^x/(1+e^x)$ and its derivative. For large negative values of x, the function and its derivative are close. In the case of the exponential function e^x, they coincide for all values of x.

of these variables represents what we are describing here as the student's potential. Under this construction, the random variable ε, which is liable to be described as the disturbance term, represents the vagaries of the processes of training and assessment. That is to say, it is the stochastic element in the transformation which maps the potential of the individual students into their assessment ratings.

The linear regression model is inappropriate even in the ideal circumstances where a cardinal index of performance is available on a scale of 0 to 100. The reason is that it imposes no explicit bound on the range of the dependent variable y which represents a student's assessment rating. This difficultly is easily overcome by applying the appropriate transformation to the cardinal index so as to convert its range to that of the entire real line. One such transformation is provided by the inverse of the so-called logistic function.

In its simplest form, the logistic function, which represents a mapping from the real line into the interval $[0,1]$, is given by

$$\pi(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{1+e^x}. \tag{2}$$

The second expression comes from multiplying top and bottom of the first expression by e^x. The logistic curve varies between a value of zero, which is approached as $x \to -\infty$, and a value of unity, which is approached as $x \to +\infty$. At the mid point, where $x = 0$, the value of the function is $\pi(0) = \frac{1}{2}$.

The inverse mapping $x = x(\pi)$ is easily derived. Consider

$$
\begin{aligned}
1 - \pi &= \frac{1 + e^x}{1 + e^x} - \frac{e^x}{1 + e^x} \\
&= \frac{1}{1 + e^x} = \frac{\pi}{e^x}.
\end{aligned} \tag{3}
$$

This is rearranged to give

$$
e^x = \frac{\pi}{1 - \pi}, \tag{4}
$$

whence the inverse function is found by taking natural logarithms:

$$
x(\pi) = \ln\left\{\frac{\pi}{1 - \pi}\right\}. \tag{5}
$$

The logistic curve needs to be elaborated before it can be fitted flexibly to a set of observations $y_1, \ldots, y_n$ tending to an upper asymptote. A more general form of the function, to replace that of (2), is

$$
\pi(x) = \frac{\gamma}{1 + e^{-\xi(x)}} = \frac{\gamma e^{\xi(x)}}{1 + e^{\xi(x)}}; \quad \xi(x) = \alpha + \beta x. \tag{6}
$$

Here γ is the upper asymptote of the function which, in the case of the cardinal measure of student performance, may be the mark of 100—or it may be something less if it is accepted that the upper limit of 100 is never attained in practice. The single explanatory variable x might stand for the past performance of a student, which also indicates the student's potential. The parameters β and α determine respectively the rate of ascent of the logistic function and the mid-point of its ascent, measured on the x-axis.

When the disturbance term is added to ξ, equation (6) becomes

$$
y = \frac{\gamma}{1 + e^{-\lambda}} = \frac{\gamma e^{\lambda}}{1 + e^{\lambda}}; \quad \lambda = \alpha + \beta x + \varepsilon. \tag{7}
$$

Then it can be seen that

$$
\ln\left\{\frac{y}{\gamma - y}\right\} = \lambda. \tag{8}
$$

Therefore, with the inclusion of a disturbance term, the equation for the generic element of the sample becomes

$$
\ln\left\{\frac{y}{\gamma - y}\right\} = \alpha + \beta x + \varepsilon. \tag{9}
$$

For a given value of γ, one can easily calculate the value of the dependent variable on the LHS. Then the values of α and β may be found by linear least-squares regression.

4. A Model with a Binary Response

In some examinations, the only published outcome is either a *Pass* or a *Failure.* Such outcomes may be the result of imposing a threshold level or pass-mark upon a marking scheme wherein the original scores vary between the bounds of 0 and 100. The summary results of *Pass* or *Failure* would seem to provide little information upon which to base a statistical analysis. Nevertheless, if there are enough observations—which is to say that numerous examination results have been recorded in conjunction with the attributes of the candidates—then the parameters of the previous model will continue to be estimable.

To understand these matters, let us recapitulate upon elements of the model. In the first place, we have defined the inherent ability of the examination candidate to be a systematic linear function $\xi = \alpha + x_1\beta_1 + \cdots + x_k\beta_k$ of a set of measurable attributes $x_1, \ldots, x_k$. Numerous unmeasurable influences will also affect a candidate's examination performance; and, to some extent, the assessment of this performance will be subjective. Therefore a stochastic element will accompany the mapping of the inherent ability into an examination score. This element may be represented by a random variable ε with an expected value of zero and a variance which is the same for all candidates.

The systematic and stochastic effects are combined in a latent variable $\lambda = \xi + \varepsilon$ which is transformed via the logistic function of Figure 1 into an examination score of $\pi = \pi(\lambda)$. In the previous analysis, it has been assumed that the actual examination score is reported on a scale of 0 to 100. Now it will be assumed that the result is reported only as *Pass* or *Failure.* These outcomes may be denoted by a variable $y \in \{0, 1\}$. Let $\rho \in (0, 100)$ denote the pass mark. Then the situation can be represented by the following scheme:

$$
(10) \qquad \begin{aligned} y = 0 \quad &\text{if} \quad \pi(\lambda) < \rho, \qquad \textit{Failure} \\ y = 1 \quad &\text{if} \quad \rho \leq \pi(\lambda). \qquad \textit{Pass} \end{aligned}
$$

Here the pass-mark ρ represents a threshold level relative to the examination score $\pi = \pi(\lambda)$. We may also define a corresponding threshold value κ relative to the latent variable λ with the effect that

$$
(11) \qquad \lambda > \kappa \qquad \text{if and only if} \qquad \pi > \rho.
$$

From the point of view of statistical modelling, it is redundant to specify the transformation from λ to π, and the analysis may be conducted more

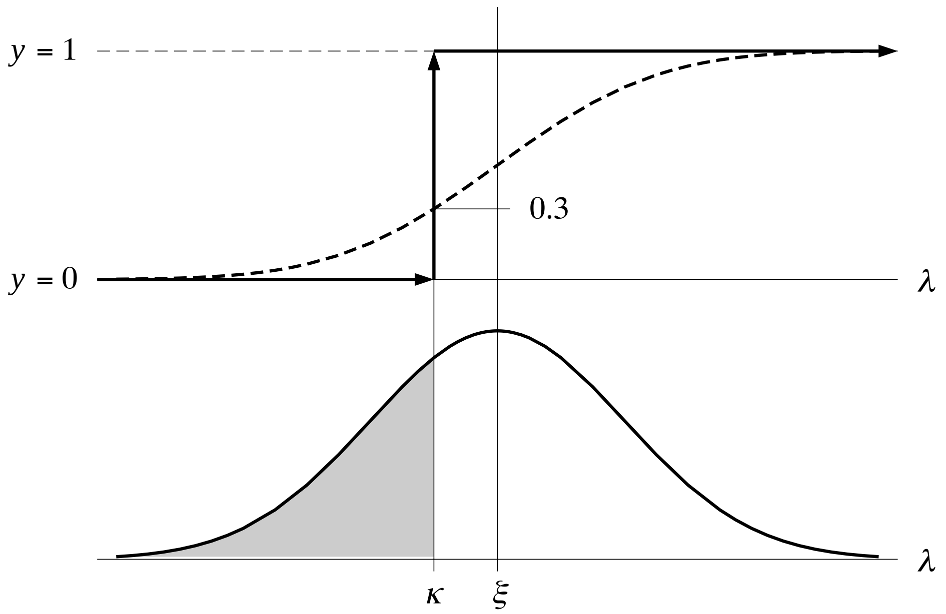

Figure 2. If $\lambda = \xi + \varepsilon$ exceeds the threshold value κ, then the step function, indicated by the arrows in the upper diagram, will deliver $y = 1$. When $\varepsilon \sim N(0,1)$ and $\xi - \kappa = 0.5$, the probability that λ will fall short of the threshold is 0.3, which is the area of the shaded region in the lower figure.

simply in terms of λ and κ. Indeed, unless one can attribute a value to the pass-mark ρ, it is impossible to construct a definitive transformation. In effect, it is impossible to determine whether a given set of summary results have been generated by applying a stringent pass-mark to an easy exam or by applying a generous pass-mark to a difficult exam.

The problem of the unknown pass-mark is reflected in the unidentifiability of the intercept parameter α within the linear function $\xi = \alpha + x_1\beta_1 + \cdots + x_k\beta_k$ which represents the candidate's inherent ability. If the pass-mark is unknown, then an arbitrary value may be attributed to α; and it is reasonable to set $\alpha = 0$. An alternative procedure is to allow the statistical data to determine a value for α and to set $\kappa = 0$, which puts the threshold value for $\lambda = \xi + \varepsilon$ at zero. The effect of either procedure is to subtract one degree of freedom from the parametrisation of the model.

The basis of our model is now a simple probability distribution which determines the probabilities P_0 and P_1 of the outcomes $y = 0$ and $y = 1$ which correspond, respectively, to a *Failure* and a *Pass:*

$$\begin{aligned} P_0 &= P(\lambda < \kappa) & \qquad P_1 &= P(\lambda \geq \kappa) \\ &= P(\varepsilon < \kappa - \xi), & &= P(\varepsilon \geq \kappa - \xi). \end{aligned} \tag{12}$$

Here the function $x'\beta = \xi$ represents the mean of the probability distribution which determines the values of P_0 and P_1. Its exact location relative to the threshold level κ is a characteristic of the individual student which is determined by their particular attributes. Notice that, if ε has a probability distributon which is symmetric about the expected value of zero, then, when $\xi = \kappa$, there are equal chances of *Pass* and *Failure.* Figure 2 illustrates the case of a candidate for whom $\xi > \kappa$, and for whom the probability of failure is $P_0 = 0.3$.

There is a variety of choices for the distribution function of ε. The most obvious choice is the cumulative normal distribution, which has been used in constructing Figure 2. A more tractable choice is the logistic function of the sort which we have used in describing the mapping from λ to π and which is depicted in Figure 1, where it already has the requisite upper bound of 1.

It remains to illustrate how the form of a cumulative probability density function can be inferred from a set of observations on $y \in \{0, 1\}$ coupled with the corresponding ability ratings ξ. When plotted on a graph, the observations (y, ξ) constitute a set of points scattered along the axes $y = 0$ and $y = 1$ which are represented in Figure 2. Imagine that intervals of equal length are demarcated along these axes over a range bounded by the minimum and the maximum of the observed values of ξ. In each interval, the number of points falling on the axes $y = 0$ and $y = 1$ are counted and the proportion p of those lying on the upper axis is calculated. The value of p coupled with the value of ξ which corresponds to the mid-point of the interval are the coordinates of a point which can be marked on the graph. When all such points are joined, a curve is described which should have roughly the appearance of a cumulative probability density function.

The procedure described above has only a tenuous connection with the mathematical procedures which are used in practice in fitting a cumulative probability distribution function to the data and in estimating the parameters of the model. Nevertheless, it does indicate how a model of examination performance can be constructed on the basis of data which might appear, at first sight, to be inadequate for the purpose.

5. A Model of Ordered Qualitative Responses

Often the percentage marks which generate a student's assessment rating are unavailable, and we have to make do with an ordinal scale of ratings which can be made to correspond to a sequence of consecutive integers. An example is provided by the Honours classification according to which universities award their degrees. This comprises 6 categories altogether, including a Pass degree and a Failure. The GNVQ assessment regime depends upon four categories; but, for expository purposes, it will simplify matters if these are reduced to three. This can be done by assuming that there are no failures. In fact, in the preliminary stage of the GNVQ assessment, there are no failures; and, therefore, the analysis which follows is wholly appropriate to that situation.

The transformation from the student's potential to their assessment rating now comprises two threshold values κ_1, κ_2 which are applied to the latent variable $\lambda = \xi + \varepsilon$ wherein $\xi = x_1\beta_1 + \cdots + x_k\beta_k$. The mechanism, which represents an elaboration of the one displayed previously under (12), is a follows:

$$\begin{aligned} &y = 0 \quad \text{if} \quad \lambda < \kappa_1, && \textit{Pass} \\ &y = 1 \quad \text{if} \quad \kappa_1 \leq \lambda < \kappa_2, && \textit{Merit} \\ &y = 2 \quad \text{if} \quad \kappa_2 \leq \lambda. && \textit{Distinction} \end{aligned} \tag{13}$$

The three events are mutually exclusive; and their respective probabilities are determined by the location of the threshold values and by the probability distribution of ε. The probability that $y = 0$, ie. of a *Pass*, is

$$\begin{aligned} P_0 &= P(\lambda < \kappa_1) \\ &= P(\varepsilon < \kappa_1 - \xi). \end{aligned} \tag{14}$$

The probability that $y = 1$, ie. of a *Merit*, is

$$\begin{aligned} P_1 &= P(\kappa_1 \leq \lambda < \kappa_2) \\ &= P(\lambda < \kappa_2) - P(\lambda < \kappa_1) \\ &= P(\varepsilon < \kappa_2 - \xi) - P(\varepsilon < \kappa_1 - \xi). \end{aligned} \tag{15}$$

The probability that $y = 2$, ie. of a *Distinction*, is

$$P_2 = P(\kappa_2 - \xi \leq \varepsilon). \tag{16}$$

There is a variety of choices for the distribution function of ε. The most obvious choice is the cumulative normal distribution, but the most tractable is the logistic function. In applying either of these functions, one can exploit their property of symmetry which implies, for example, that $P(\kappa_2 - \xi \leq \varepsilon) = P(\varepsilon \leq \kappa_2 - \xi)$ since $E(\varepsilon) = 0$.

If the assumption is made that the assessment ratings of the individual students are independently distributed random variables, then the probability of obtaining a particular sample of ratings is just the product of the individual probabilities. The likelihood function is the probability density function of the sample seen as a function of the parameters of the transformation which maps from the measured indices of student potential—which are the variables $x_1, \ldots, x_k$—to the individual assessment ratings. The parameters which are to be estimated consist, therefore, of the threshold values κ_1, κ_2, and the regression coefficients $\beta_1, \ldots, \beta_k$ comprised by $\xi = x_1\beta_1 + \cdots + x_k\beta_k$.

Estimates of these parameters may be obtained by finding the values which maximise the likelihood function. This is a matter of nonlinear optimisation which, in this case, is relatively straightforward. There are several commercially available computer programs which are capable of calculating the parameters of an ordered-response model; and, for this project, the *Stata.* [10] was chosen.

6. Monitoring the GNVQ Assessment Process at the Local Level

Once the parameters of the ordered-response model have been estimated on the basis a wide sample of data, a simple system for monitoring the GNVQ assessment process in individual locations could be established. It is a straightforward matter to compare the proportions of assessments falling into the various categories in a given location with the proportions which prevail nationally. However, the intention is to compare the actual local proportions with those which are predicted on the basis of a statistical model which takes account of the potentials of the students.

Such predictions depend upon the availability of the indices $x_1, x_2, \ldots, x_k$ for each student. Using the estimated regression coefficients, the academic potential $\xi = x_1\beta_1 + x_2\beta_2 + \cdots + x_k\beta_k$ can be calculated for each student. From the graph of Figure 4, the probabilities P_0, P_1 and P_2 that they will obtain respectively, a *Pass*, a *Merit* or a *Distinction* can be derived. The average over all the students of these probabilities provides the local predicted proportions for the assessment categories. A simple graphical presentation can be provided which uses histograms (bar charts) to compare the local proportions in the assessment categories to the national proportions and to the proportions predicted for the location.

Summary indices may be devised which show the extent of the divergences of the predicted and the actual assessment proportions as well as the national and the local assessment proportions. The indices would be based upon weighted sums of squares of the differences of the proportions in each category. The weighting functions would be developed in the light of experience so as to be sensitive to those anomalies which cause the greatest concern to the national administrators of the GNVQ.

7. The Index of GCSE Performance

The study by Armitage and Nutall [4] revealed that the most effective variables for predicting BTEC grades were previous GCSE grades and age. It was not revealed how the information on these grades was used, but it is presumed that they were combined into a single index by use of a points system. A points system has been used, for example, in a study conducted at the

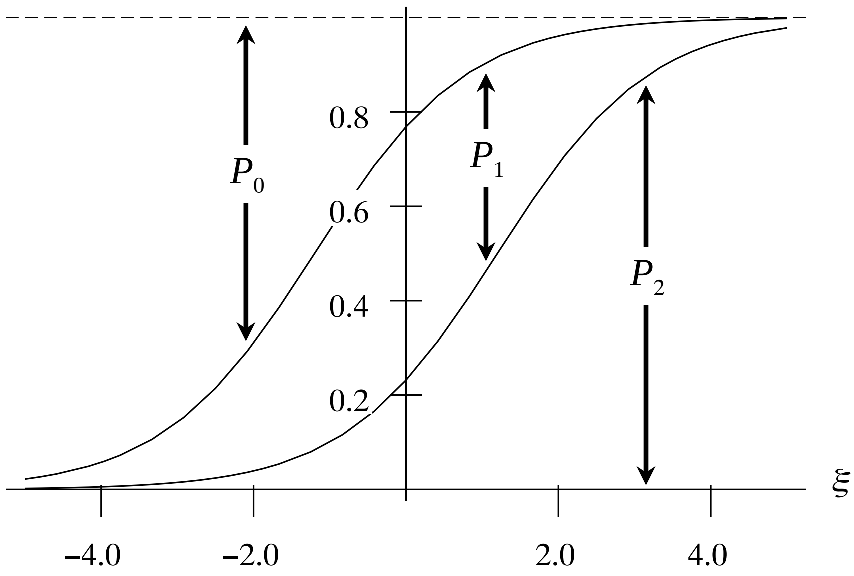

Figure 3. The probabilities P_0, P_1 and P_2 can be determined by two logistic functions of the argument $\xi = x'\beta$ in such a way that they are guaranteed to sum to unity.

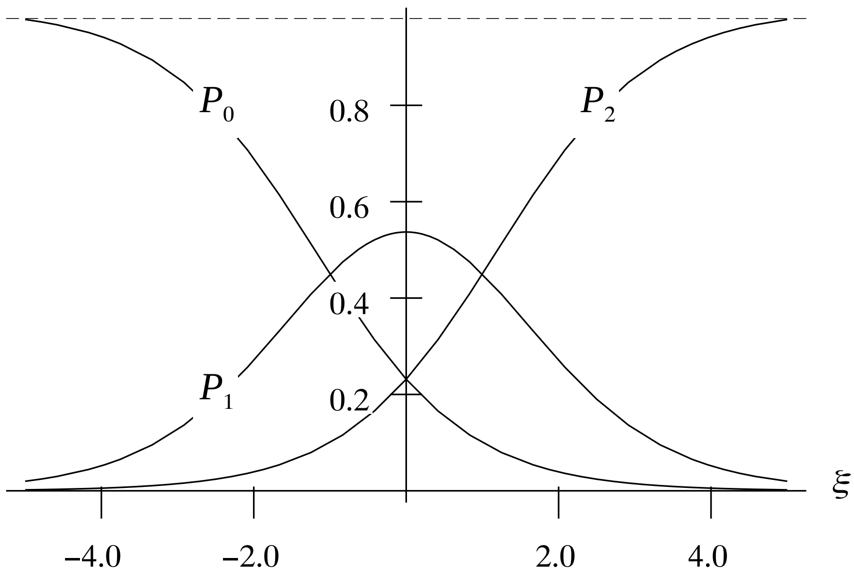

Figure 4. Figure 3 may be reconstructed so as to show more directly how the probabilities $P_0 = P_0(\xi)$, $P_1 = P_1(\xi)$ and $P_2 = P_2(\xi)$, which represent respectively the chances of gaining a *Pass* a *Merit* and a *Distinction*, vary as a function of $\xi = x'\beta$ which represents the student's academic potential.

University of Newcastle into the value-added by BTEC courses. Here the grades were recoded into numerical equivalents as follows:

Distinction	5
Merit	4
Pass	3
Refer	2
Fail	1

There is a straightforward way of constructing an index. Let v_i be the value attributed to a result in the ith class. For example, in the scheme above, $v_5 = 5$ is the value attributed to a *Distinction.* Then, if there are altogether k classes, the index of examination performance might be calculated as

$$I = \sum_{i=1}^{k} v_i n_i, \tag{17}$$

where n_i is the number of the exam results of the candidate in question which fall into the ith class.

The objection is that such an index is too simple. In the first place, it entails an arbitrary judgment regarding the relative merits of the various grades. Thus, for example, by adopting such a scheme one is asserting *a priori* that, in terms of its favourable influence upon the outcome of a subsequent exam, a *Distinction* is almost twice as effective as a *Pass* grade and exactly three times as effective as a *Refer* grade. These relative values should be determined empirically rather than chosen *a priori.*

Another problem with the simple index is that it is liable to confound quantity with quality. Thus, for example, if the points system which is tabulated above were adopted, then it would be implied that five fails are equivalent to one distinction in the extent to which they promote a favourable outcome in a subsequent exam. This problem can be overcome only by using separate indices of quantity and quality.

The present study uses a quadratic weighting scheme in constructing an index of GCSE performance. This should allow the index to be determined flexibly by the data. To describe the scheme, let $v_1, v_2, \ldots, v_k$ represent a set of values attributed to k examination classes, including the failure. These values are assumed to follow an arithmetic or linear progression. Let $n_1, n_2, \ldots, n_k$ be the numbers of results in each class which have been obtained by an individual candidate, and let $n = n_1 + n_2 + \cdots + n_k$ be the total number of examinations sat by the candidate. Then a vector of three elements t, ℓ, q is constructed as follows:

$$(18) \qquad \begin{bmatrix} t \\ \ell \\ q \end{bmatrix} = \frac{1}{n}\begin{bmatrix} 1 & 1 & \cdots & 1 \\ v_1 & v_2 & \cdots & v_k \\ v_1^2 & v_2^2 & \cdots & v_k^2 \end{bmatrix}\begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_k \end{bmatrix}.$$

It will be observed that the value of t is unity, identically; and this reflects the fact that the proportions n_i/n of the results in each grade must sum to one. The proposed index takes the form of

$$I = \alpha n + \beta t + \gamma\ell + \delta q,$$

where α, β, γ and δ are parameters which are determined by statistical estimation.

These parameters can be given simple interpretations. The parameter α measures the quantity effect; and it is expected to have a positive value if having sat a greater number of examinations at the previous level is conducive to better examination result at the present level.

The parameter β is a constant term which is absorbed by the threshold levels of the binary model and the ordinal model. The parameters γ and δ are both associated with quality effects. The former is the parameter of the linear effect and the latter is the parameter of the quadratic effect. Together they provide the necessary flexibility in the mapping from the point values v_i to the index I.

There is a natural presumption that the γ coefficient will be positive, since one would expect good results in the previous examinations to presage good results at the present level. However, if the quadratic parameter δ is positive, then it is quite possible for γ to take a negative value. There is no presumption about the sign of the δ coefficient. In the language of economists, a positive δ would imply increasing marginal returns to excellence and a negative coefficient would imply diminishing marginal returns.

If δ were to take a large negative value, then one might find, at some stage in the progression, that previous excellence is unfavourable to present achievement. In other words, the diminishing marginal returns could become negative returns.

Bibliography

[1] Agresti, A., (1984), Analysis of Ordered Categorical Data, John Wiley: New York.

[2] Armitage, P., (1980), Procedures used in the Armitage System of Statistical Monitoring, Discussion Paper of the Student Assessment Research

Group, Institute of Educational Technology, The Open University, Milton Keynes.

[3] Armitage, P., (1981), Statistical Monitoring in the Open University: An Evaluation of Armitage (1980), Discussion Paper of the Student Assessments Research Group, Institute of Educational Technology, The Open University, Milton Keynes.

[4] Armitage, P., and D.L. Nutall, (1983), The Moderating Instrument Research Project: A Feasibility Study, School of Education, The Open University, Milton Keynes.

[5] Becker, W.E. and P.E. Kennedy, (1992), A graphical exposition of ordered probit, Econometric Theory, 8, 127–31.

[6] Cramer, J.S., (1991), The Logit Model: An Introduction for Economists, Edward Arnold: London.

[7] Gourieroux, C., (1984), Econometrie Des Variables Qualitatives, Editions Economica: Paris.

[8] Greene, W.H., (1990), Econometric Analysis, Macmillan: New York.

[9] Maddala, G.S., (1983), Limited-Dependent and Qualitative Variables in Econometrics, Econometric Society Monograph No. 3, Cambridge University Press, Cambridge U.K.

[10] *Stata:* Statistics, Graphics and Data Management, (1993), Stata Corporation.

[11] McCullagh, P., (1980), Regression models for ordinal data, Journal of the Royal Statistical Society, Series B, 42, 109–42.

[12] McKelvey, R. and W. Zavoina, (1975), A statistical model for the analysis of ordinal level dependent variables, Journal of Mathematical Sociology, 4, 103–20.

[13] Pratt, J.W., (1981), Concavity of the log-likelihood, Journal of the American Statistical Association, 76, 137–59.

[14] Rahiala, M. and and T. Terăsvirta, (1988), Formation of firms' production decisions in Finnish manufacturing industries, Journal of Applied Econometrics, 3, 125–37.

[15] Terza, J., (1985), Ordinal probit: a generalisation, Communications in Statistics, 14, 1–12.