Impact and Effectiveness of Widening Access to HE in Wales

Working Paper Series[1] – WAQNCW2013-1

Modelling Access to Higher Education: an evaluation of previous approaches – Caroline Wright

[1] **Working Papers are intended to present initial analysis and findings. Please do not cite them without permission of the author(s).**

Foreword

The Wales Institute of Social and Economic Research, Data and Methods (WISERD) is currently evaluating the impact and effectiveness of Widening Access to higher education (HE) in Wales. This project is funded through the Economic and Social Research Council's (ESRC) Secondary Data Analysis Initiative (ES/K004247/1); and by the Higher Education Funding Council for Wales (HEFCW). It will be completed by December 2014.

Access to higher education has become an extremely controversial area of policy, as successive UK administrations have sought to balance increasing student fees with ensuring that HE is open to individuals from as wide a range of social backgrounds as possible. Moreover, relatively distinctive approaches have been adopted in the different devolved administrations of the UK. For example, currently, the Welsh Government has undertaken to pay the increased costs to students arising from the abolition of the fees cap. However, the evidence-base for evaluating different approaches to widening access is relatively weak. Accordingly, WISERD, the HEFCW and the Welsh Government (WG) are collaborating on this innovative research study.

The research analyses how individuals who are resident in Wales progress through secondary school, into sixth forms and further education colleges for post-16 education and on to HE. It also explores what are the key factors here in determining whether individuals progress through the education system to HE or not. What are the relative impacts of the social characteristics of individuals, their previous educational attainment and their progression through the education system? What does this imply for the effects of barriers at the point of entry to HE, such as fees levels, entry processes and so forth? Answers to these questions are known for England, but not for other parts of the UK.

The analysis is based on the innovative use of three linked sources of information, the data for each of which are collected initially for administrative purposes. These are: the National Pupil Database (NPD) for Wales; the Lifelong Learning Wales Record (LLWR); and Higher Education Statistics Agency (HESA) data. By linking these together, it is possible to trace individual trajectories through the education system to entry to HE. It is also possible to compare systematically the trajectories of those who do participate in HE with those who do not. Moreover, using sophisticated statistical techniques, it is possible to determine which are

the most influential factors in shaping patterns of HE participation. Results here will be compared with those that have been produced by similar analyses in England.

A second part of the proposed study (funded by additional resources made available by the HEFCW) investigates the development of distinctive approaches to widening access to HE by successive Welsh administrations since devolution in 1999. Of key significance here is to establish the rationales that underpin the approaches adopted in Wales; and to compare these with those that have informed policy approaches in the other countries of the UK and England, in particular. In addition, the study examines the ways in which national policies have been implemented by the Welsh universities, paying special attention to the assumptions about the determinants of HE participation that are in play here. This part of the study is based on fieldwork, comprising the analysis of official and semi-official documents and interviews with politicians and senior officials responsible for widening access policies; and with the professionals inside the universities responsible for implementing these policies.

The results of the research will be fed directly into the deliberations of the WG and the HEFCW on the future development of policies on widening access to HE, which will be especially intensive over the next few years. Moreover, they will also provide the basis for working with the professionals in the universities with responsibility for implementing widening access policies, to integrate the use of analyses of administrative data more firmly into their day-to-day practices.

Introduction

Access to higher education (HE) has become an extremely controversial area of policy, as successive UK administrations have sought to balance increasing student fees with ensuring that HE is open to individuals from as wide a range of social backgrounds as possible. There have also been numerous initiatives that have sought to increase participation by disadvantaged groups, by raising levels of educational attainment or changing attitudes and aspirations in relation to HE (Gorard *et al.*, 2007). Moreover, somewhat distinctive approaches have been adopted by the administrations in the different countries of the UK. This entails a complex set of circumstances; but also opens up the possibilities for fruitful 'home international' comparative analysis (Gallacher and Raffe, 2011).

In general terms, however, there is a dearth of studies which have attempted to carry out such comparative analysis. One of the aims of WISERD's research project on widening access to HE is to contribute to filling this gap. More specifically, the recent availability of linked administrative data-sets, which brings together information from the National Pupil Database (NPD), the Lifelong Learning Wales Record (LLWR) and Higher Education Statistics Agency (HESA), permits the comparison of patterns of entry to HE in Wales with those in England. In particular, Chowdry *et al.* (2013) have analysed entry into HE in England in a pioneering study. Accordingly, one of the objectives of the WISERD project is to replicate Chowdry et al.'s analysis, thereby allowing systematic comparisons to be made between Wales and England.

Equally, however, the WISERD project seeks to develop alternative approaches to the modelling of access to HE. With this in mind, then, this paper presents a critical evaluation of the methodology adopted by Chowdry *et al.* (2013) in order to build upon their approach; and to formulate alternative methodological approaches which may provide new insights into the patterns of access to HE in Wales and – by implication, at least – England too.

An overview of Chowdry *et al.* (2013)

- 'Widening participation in higher education: analysis using linked administrative data' was published in the *Journal of the Royal Statistical Society A*, 176 (2) in 2013.

- The paper uses linked, individual level data from the NPD, the National Information System for Vocational Qualifications and HESA, to track two cohorts of young people from age 11 through to age 20.

- The authors use data for both participants and non-participants in HE to assess the effect of socio-economic status (SES) on (1) the propensity to participate in HE. This provides a more robust approach than many previous analyses, which have relied on individual-level administrative data from HE institutions alone. They also model (2) the propensity to attend a 'high-status' university, using data for HE participants only. Both modelling strands consider the supplementary effect of SES, once attainment has been taken into account.

- They find that pupils from lower SES backgrounds are much less likely to participate in HE than pupils from higher SES backgrounds; and are also much less likely to attend a high-status universities.

- However, these differences do not emerge at the point of entry to HE. Rather, lower SES pupils tend not to achieve the same levels of educational attainment as higher SES pupils throughout education; and consequently do not have the same options available to them. Previous educational attainment is the most important factor in shaping patterns of entry to HE.

- That said, socio-economic differences in participation remain on entry to HE, even allowing for prior attainment. However, the effect of SES is much smaller than that of previous educational attainment.

A Problematic Methodology?

In methodological terms, Chowdry *et al.*'s (2013) analysis appears to provide a robust approach; and clearly allows a highly authoritative analysis to be developed that takes us beyond the bulk of previous studies within the area. However, a number of important methodological issues also arise from their work. The following section identifies and discusses some of these issues. These are:

- Stratified analysis
- Use of fixed effects models
- Use of linear rather than logistic regression

- No consideration of other random effects
- Use of contextual analysis
- Measuring socio-economic background

Stratified Analysis

Stratified Analysis, in this instance, is based on the assumption that males and females are so different in terms of the outcome variable (propensity to participate in HE and propensity to attend a high-status university), that they should be stratified into two different groups and modelled separately. There are advantages to this style of modelling. For example, in a model with several explanatory (*x)* variables, a dummy variable for gender would simply tell us the difference between males and females, controlling for all the other *x* variables. However, if the coefficient on education varies across genders, this would not be apparent unless an interaction term were included. One interaction in a model is fine, but including an interaction between gender and every other *x* variable would quickly become very complicated.

That said, by not fitting an overall single model, it is not possible readily to test for the differences between male and females. Further, by doing separate analyses for both males and females, statistical power is also being lost, as typically data are reduced by 50 per cent for each model. This may not be a problem in the Chowdry *et al*. paper, as they have a very large number of observations (*n)*, but it may result in a substantial loss of statistical power for other research projects, including the WISERD one, owing to the much smaller population.

A potential solution to this problem is to undertake *both* forms of analyses (stratified and non-stratified) and assess which is more appropriate, given the nature of the Wales dataset and the intermediate results that are produced.

Use of fixed effects modelling to assess school effects

Chowdry *et al*. (2013) use fixed effects models to assess school effects. This is problematic for several reasons:

- It is not possible to interpret the 'school effect' by adding in a separate intercept, or dummy, for each school. A fixed effects approach simply 'models out' the effect of school. This may be an appropriate strategy if you are simply assessing the effectiveness of a national policy, such as widening participation overall. However, if

you are interested in exploring what it is about schools that makes a difference to young people's educational trajectories, then a random effects model is more appropriate.

- There is no single parameter to assess between-group differences: it provides no metric for school effects. This means that research using these methods, are unable to compare the size of the school effects, with the other effects in the model i.e. you don't get any understanding of relative effects.

- It is not possible to include group-level predictors e.g. school type, because all degrees of freedom have been consumed at the group-level (this has to be done in a separate analysis).

- There are potentially unstable estimates, unless all cluster sizes are *large*. Clusters based on a small *n* will be highly volatile and therefore unreliable. It is also problematic to attribute equivalent meaning to schools with vastly different populations.

The benefits of using random effects models instead include

- They produce an interpretable parameter relating to the higher level variance (absent from the fixed effects model).

- They partition out the variance at each level of the model, allowing us to consider more than one higher level, for example, local authorities and/or neighbourhoods.

- They allow us to add in contextual-level variables - for example, school type - to try and explain the school-level variance. This cannot be done in a fixed effects model; by adding in separate intercepts for each school, all the degrees of freedom in the model are used up and there will be no more variance remaining in the model to explain.

- They provide more stable estimates of between-school variance by 'borrowing strength' from an overall distribution of variance, rather than treating each mean as equally valid regardless of its size.

- They allow the use of more complex (cross-classified) modelling structures, if required.

- They allow the use of logistic regression, rather than linear regression, which is a more appropriate method for binary outcomes.

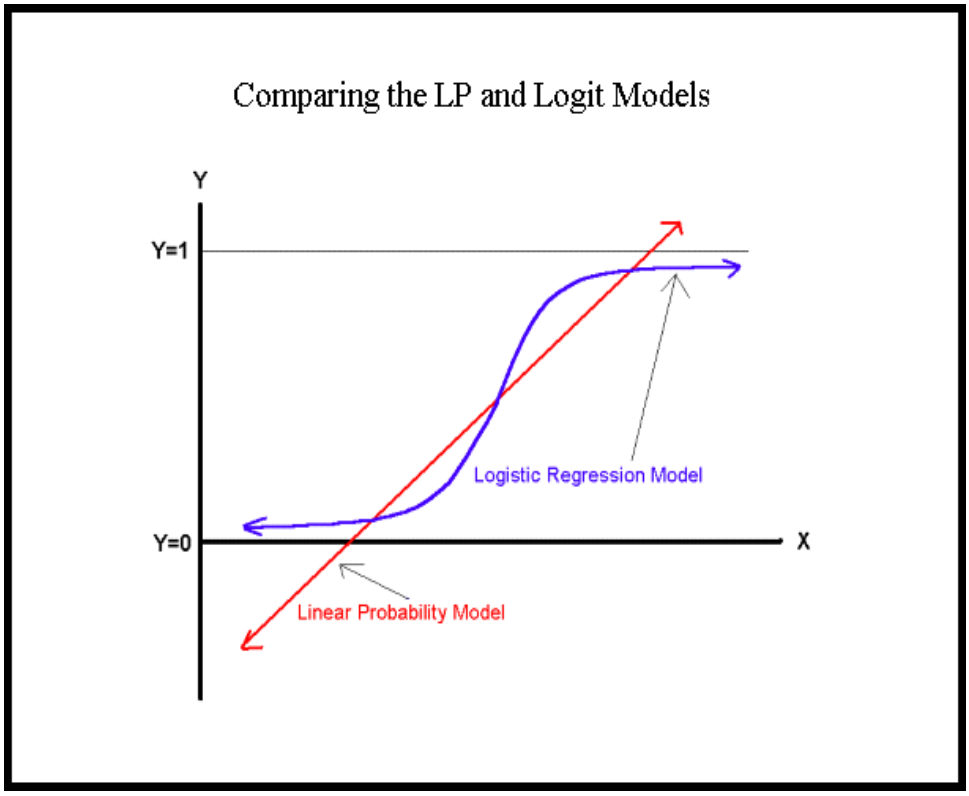
It is easy to identify high and low achieving schools/local authorities/neighbourhoods for more in depth qualitative analysis. A potential solution to this problem is to undertake both forms of analyses (school effects modelled as fixed effects *and* as random effects) and see if the random effects models produce substantively different results.

Use of linear rather than logistic regression

Chowdry *et al*. (2013) do not to use logistic regression because this type of model will not converge, given the large number of fixed effects (>4000) that need estimating. Rather, they use linear regression analysis instead. However, as will be argued here, the logistic regression is preferable to linear regression for several reasons.

If we first consider the different relationships described by the two models, by referring to Figure 1, the *y* axis refers to propensity to attend a high-status university and the *x* axis refers to attainment level. It is unlikely that attainment would have a constant and even effect on a student's likelihood of attending a high status university, as depicted by the linear relationship. It is much more likely that there is a threshold of sorts, whereby achieving a certain set of A-Level grades (AAB) increases the propensity to attend a high-status university substantially. This type of relationship is better reflected by the logistic regression (blue) line, which is far steeper than the linear regression (red) line. Essentially, the logistic model is more sensitive than the linear model when further away from the extremes - in the middle - and conversely, less sensitive than the linear, at the extremes (Jiménez and Salas-Velasco, 2000). Again, this is intuitive, where the 'S' shaped logistic distribution is very flat, this indicates that there really is not much variation amongst the very low-achieving students. This is the case until you hit the (AAB) threshold on the *x* axis and your propensity to attend a high status university leaps up. Where we see the curve flatten out, this indicates that once you have exceeded a certain level of attainment, propensity to attend a high-status university does not alter very much according to small differences in grades.

Figure 1: The different results from a Linear Probability Model (red line) and a Logistic Regression Model (blue line)



Secondly, linear models can produce unobtainable results (those falling above 1 and below 0). Indeed, Chowdry *et al*. (2013) report that 17 per cent of the results for males and 7 per cent of the results for females were <0 or >1 for the first strand of modelling (propensity to participate in HE); and 10 per cent of the results for males and 3 per cent of the results for females were <0 or >1 for the second strand of modelling (propensity to attend a high status university).

By modelling the school effects using random effects models, it will be possible to use logistic regression, so these problems will be avoided.

No consideration of other random effects

The classifications of schools and local authorities were chosen for inclusion in this research owing to the influence that these contexts are known to have on young people's educational outcomes (Cheng, 1995; Sutton Trust, 2007, 2008; BIS, 2009; Chowdry et al., 2010 and 2013). It is reasonable to assume that schools have a direct influence on the experience that young people have in education, further, that the ethos of the school, the teachers that teach there, the size of classes and the type of pupil that attends the school, will all have a bearing on their student's likelihood of participating in higher education. Although the role of local authorities has diminished in recent years, they still have many statutory duties, which impact

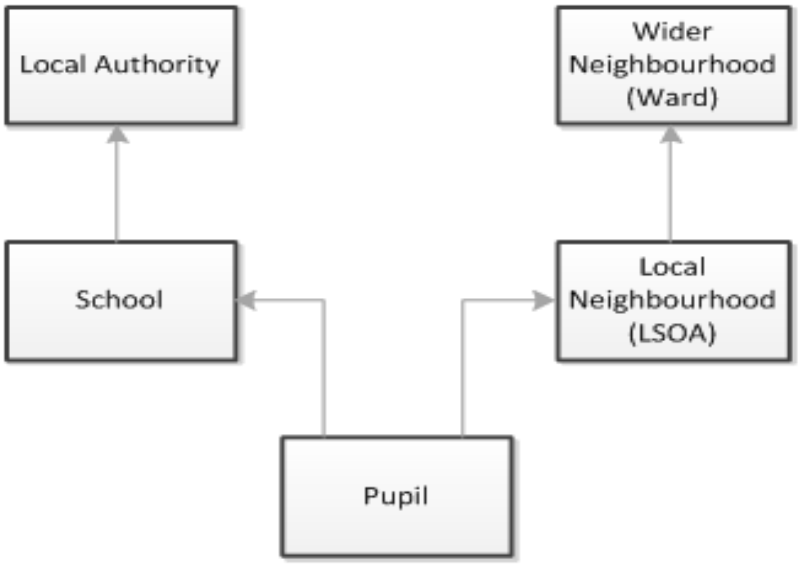
both directly and indirectly on how schools function. For example, they employ teachers in maintained schools; determine school budgets; maintain schools in their areas; and ensure that the curriculum provided by maintained schools comprises the National Curriculum (DfE, 2013). Once again, it is assumed that these types of functions will impact on the culture of the school, and subsequently on the experience that the student has at school, which potentially affects their likelihood of participating in higher education.

In addition to the educational contexts that may influence young people's outcomes, much of the literature places an importance on where children live, this is often referred to as 'neighbourhood effects' (Van Ham et al., 2012). For example, HEFCE (2005) argues that there are 'broad and deep divisions in the chances of going to HE according to where you live' (p.10). Indeed, area-based approaches remain the corner-stone of widening participation strategies, based on Communities First areas in Wales and POLAR areas in England (Taylor *et al*., 2013), For these reasons, it is important to assess the influence of contexts other than school and include an analysis of neighbourhoods. Thus, the lower layer super output area (LSOA), that the pupil lives in, which may be understood as their local neighbourhood, and the ward that the LSOA belongs to, which may be understood as their wider neighbourhood are both included as classifications in the research. Each of these contexts represents a social and/or educational context, which may impact upon outcomes for young people and allows the analysis of contextual-level variability in educational participation.

However, the potential impact that the neighbourhood a student lives in is not addressed by Chowdry *et al.* (2013). Here, an alternative option is to consider both the effects of educational context and social context, i.e. schools *and* neighbourhoods. Unlike most applications of the multilevel approach there is not a simple hierarchy where each lower level is nested in a higher unit. The left side of the diagram shown in Figure 2 gives the hierarchical structure for schools while the right-side gives the hierarchical structure for neighbourhoods. Consequently, there are two sets of hierarchies which are potentially crossed given that not all the pupils from the same LSOA go to the same school, nor do all the pupils at one school live in the same LSOA. This complex structure will be analysed using non-hierarchical, cross-classified multilevel models (Goldstein, 1994). The cross-classified multilevel structure shows the macro contexts. Individuals' are simultaneously nested in both schools, which are nested in Local Authorities, and LSOAs, which are nested in Wards so that individuals can be understood as occupying more than one set of contexts - as they do in real life. The LSOA's are taken here as defining the local neighbourhood in which pupils

live, while Wards are taken as the wider neighbourhood. Each of these classifications may hold significance over the educational outcomes being measured.

Figure 2: Classification diagram for cross classified models



In order to critically evaluate current policy, exploratory modelling of the possible neighbourhood effect(s) should be undertaken.

Use of Contextual Analysis

Contextual analysis describes a form of analysis which considers individual-level fixed effects, such as gender, ethnicity and social class, and additionally includes contextual-level variables, for example, school type or neighbourhood deprivation. Chowdry *et al.* (2013) include school type in their fixed effects models, in an attempt to capture a 'school effect'.

This is problematic for at least three reasons:

- Individual-level fixed effects may cause problems in this example because regression analysis assumes that the error terms are independent of one another (that is, there should be no pattern to the residuals). However, this is very unlikely, as research has shown that pupils from the same school are more similar to one another than pupils from different schools, which will result in clustering. Interpreting the residuals as if they are independent, when they are not, will lead to type 1 errors, that is, finding significance where there is none.

- It may lead to incorrect standard errors for group-level predictors caused by omitted variable bias (for example, school type may be statistically significant, but this may be a symptom of its correlation with other variables).

- It assumes that all group-level variance can be explained by group-level predictors, which is not accurate. There may well be variance at the school level (that is, schools may vary according to how many pupils they get into HE). However, this type of variance may not be explicable by the school-level variables that are added to the model. Using this method could lead us to assume that there is no school-level variance, because the school-level variable is not significant, when in fact there is. Adopting a random effects approach, would allow us to test both of these things.

Measuring socio-economic background

Chowdry *et al.* (2013) opt to combine individual and neighbourhood level data to create an index of socio-economic background using principal components analysis. They include Free School Meal eligibility; the index of multiple deprivation; a classification of neighbourhood type; and three very local area-based measures from the 2001 census (specifically the proportion of individuals in the area: who work in higher or lower managerial or professional occupations; whose highest educational qualification is level 3 or above; and those who own their own home, either outright or through a mortgage.)

This approach runs the risk of falling foul of the ecological fallacy: the fallacious attribution of aggregate results on to individuals. The classifications based on the 2001 Census are also significantly out of date, whereas the WISERD project will have the benefit of being able to use the 2011 census data.

However, it is much more difficult to develop alternative approaches which avoid these difficulties. Nevertheless, the WISERD project will attempt to address this issue; and, in doing so, will draw upon the previous work in the sociology of education, which has provided important insights into the relationships between social class and entry to higher education. This is a literature that Chowdry *et al.* (2013) effectively ignore.

Summary

The aim of this paper was to critically evaluate the methods adopted by Chowdry *et al*. in their 2013 paper: *Widening participation in higher education: analysis using linked*

administrative data; and to make suggestions as to how these methods could be developed for the WISERD research project: *Impact and Effectiveness of Widening Access to HE in Wales*. Critiques were made in relation to each of the following areas:

- Stratified analysis
- Use of fixed effects models
- Use of linear rather than logistic regression
- No consideration of other random effects
- Use of contextual analysis
- Measuring socio-economic background

Proposals made as to how these methods could be improved upon are as follows:

- Conduct stratified and non-stratified analysis of gender, compare the two approaches and adopt the better of the two.
- Explore school effects using both fixed and random effects models, in order to explore whether substantively different results are generated using the different methods.
- Use logistic rather than linear regression.
- Additionally explore random effects relating to local authority and neighbourhood. Again, this is exploratory, and will only be adopted is significant results are produced.
- Consider contextual level effects by including contextual variables into the random effects model and measure if variance has been reduced.
- Consider new strategies for measuring socio-economic background.

These latter issues will be implemented as the WISERD project develops its own approaches to the modelling of entry to HE.

References

BIS 2009 Applications, Offers and Admissions to Research Led Universities

Cheng, Y. (1995) England and Wales Youth Cohort Study. Staying on in full-time education after 16: do schools make a difference Sheffield: Research Strategy Branch, Department for Education and Employment

Chowdry, H., Crawford, C., Dearden, L., Goodman, A. and Vignoles, A. (2013) 'Widening Participation in Higher Education: analysis using linked administrative data, *Journal of the Royal Statistical Society A*, 176:2, 431-57.

Gallacher, J. and Raffe, D. (2011) 'Higher Education Policy in Post-Devolution UK: more convergence than divergence?', *Journal of Education Policy*, 27:4, 467-490.

Goldstein, H. (1994) 'Multilevel cross-classified models' in Sociological Methods and Research, 22, 364-375.

Gorard, S., Adnett, N., May, H., Slack, K., Smith, E. and Thomas, L. (2007) *Overcoming the Barriers to Higher Education*, Stoke-on-Trent: Trentham.

HEFCE (2005) Young participation in Higher Education

Jimenez, J. D., Salas-Velasco, M., 2000. Modeling Educational Choices. A Binomial Logit Model Applied to the Demand for Higher Education. Higher Education, 40: 293-311.

The Sutton Trust. 2007. University admissions by individual schools.

Taylor, C., Rees, G., Sloan, L. and Davies, R. (2013) 'Creating an Inclusive Higher Education System? Progression and outcomes of students from low participation neighbourhoods at a Welsh university', *Contemporary Wales*, 26, 138-61.