LLAKES
CENTRE FOR LEARNING AND LIFE CHANCES IN KNOWLEDGE ECONOMIES AND SOCIETIES

IOE LONDON
Leading education and social research
Institute of Education
University of London

Estimating the effect of lifelong learning on women's earnings using a switching model

Richard Dorsett, Silvia Lui and Martin Weale

LLAKES Research Paper 30

Centre for Learning and Life Chances in Knowledge Economies and Societies

LLAKES is an ESRC-funded Research Centre - grant reference RES-594-28-0001.

To request printed copies of this paper or other LLAKES Research Papers,
please contact the LLAKES Centre - llakescentre@ioe.ac.uk.

To view this and other LLAKES Research Papers online, please visit www.llakes.org.uk.

ISSN 2042-5929

Estimating the effect of lifelong learning on women's earnings using a switching model *

Richard Dorsett, Silvia Lui and Martin Weale
National Institute of Economic and Social Research,
2, Dean Trench Street,
London SW1P 3HE

8th September 2011

Abstract

This paper examines the effect of lifelong learning on women's employment and wages in the United Kingdom. Using data from the British Household Panel Survey, a variant of the mover-stayer model is developed in which hourly wages are either taken from a stationary distribution (movers) or are closely related to the hourly wage one year earlier (stayers). The model allows for individual-specific effects through the inclusions of a fixed number of discrete mass points and also addresses the potential endogeneity of lifelong learning decisions. Once employment effects are taken into account, all forms of lifelong learning show substantial returns.

JEL Codes: C33, C35, I20, J24, J31, J63
Key Words: Lifelong Learning, Switching Regression, Sample Selection, Wage Dynamics

*We are grateful to the ESRC for supporting this research through the LLAKES research centre at the Institute of Education and the National Institute of Economic and Social Research. We also gratefully acknowledge comments from participants in seminars at the Institute of Education and the National Institute. The study used the British Household Panel Survey (University of Essex. Institute for Social and Economic Research, British Household Panel Survey: Waves 1-18, 1991-2009 [computer file]. 7th Edition. Colchester, Essex: UK Data Archive [distributor], July 2010. SN: 5151). The usual disclaimer applies.

Summary

For many years there has been concern about the educational attainment of the population in the United Kingdom. In order to meet these concerns the government has followed a joint policy of encouraging post-compulsory education and training and promoting lifelong learning, i.e. study by those who ended their education some years earlier. The underlying assumption is that there is a substantial body of the population which may not have realised their full potential during their period of school and post-school education. But, while it is generally accepted that there are economic benefits to post-compulsory education, realised as increased earnings, it is less clear what the economic benefits are to lifelong learning.

There are a number of problems in investigating this. First of all, while many social surveys ask people about their qualifications, they do not generally distinguish lifelong qualifications from those obtained as part of the more traditional process of education. Secondly, even more than with traditional education, there is the risk that, should higher earnings be associated with those with lifelong qualifications, that may reflect their underlying ability as much as the benefits obtained from their qualifications. Thirdly, while specific surveys may be designed to collect information on the effects of qualifications, it is much harder to do this specifically for lifelong learning.

In this study we use the British Household Panel Survey to examine the effects of lifelong learning on women's employment and earning. We develop an approach to distinguish the impact of lifelong learning from the effects of other individual-specific characteristics on earnings. We set our analysis in a framework for modelling hourly earnings adapted from the mover-stayer model developed by sociologists to study a range of social phenomena. In our model people's hourly wages are either reasonably closely related to those they earned in the previous year (the stayers) or they are equal to a value which reflects factors such as their age and education and on a random term. Since this random term in one year is not related to its value in the previous year those whose earnings are determined in this way are identified as movers. The model allows

individuals' characteristics (including qualifications acquired through lifelong learning) to influence whether they are a mover or stayer at a given point in time. In all these cases we can examine whether qualifications gained through lifelong learning affect women's hourly wages. However, not all women of working age work, and we need to introduce an additional equation to our model which summarises women's employment outcomes. This too, reflects the possible impact of qualifications obtained through lifelong learning.

The small number of respondents in the British Household Panel survey who have undertaken lifelong learning of any kind means that it is not possible to explore in detail the effects of different types of qualifications. We classify women's educational attainment using the framework provided by national vocational qualifications, but with academic qualifications classified to the appropriate levels on this scale. We then distinguish women who undertake lifelong learning without increasing their attainment level from those whose lifelong learning qualifications do result in upgraded attainment. A simple analysis of the data suggests that, in general terms, those who have lifelong qualifications are more likely to be employed and earn more than those who do not. Qualifications which upgrade attainment are more valuable than those which do not.

Our econometric analysis confirms this finding. We identify wage effects of 6-11% for lifelong learning without upgrading and 22-32% for upgrading of qualifications. Obviously these are only averages across what are inevitably heterogenous mixes of qualifications. However, we also find that acquisition of qualifications through lifelong learning has a substantial impact on women's employment prospects. These effects are most marked the least-qualified were the women before lifelong learning. For a twenty-five year old woman not initially qualified even to level 1, upgrading of qualifications results in an hourly earnings premium of 22%. But when the effects of employment are taken into account, the total financial return to acquisition of qualifications is estimated at 74%. This effect, though large, is nevertheless smaller than would be suggested by simple application of the probability of employment for women, initially not qualified to level 1, with and without upgrading though lifelong study. For older people and those with initial qualifications the total effects are smaller, but we find that at all qualification levels employment effects enhance the impact of lifelong learning on qualifications. Our model incorporates as a special case the situation in which wage dynamics are represented in first differences and we show that, relative to our model, such a simplification is rejected statistically.

From a policy point of view, the results point to very substantial economic benefits from lifelong study. The importance of taking account of employment as well as wage effects is very clearly demonstrated.

1 Introduction

In many countries, government policy has been to encourage lifelong learning as a means of increasing productivity and achieving progression in the labour market. In the United Kingdom at least, such policy follows from the observation that family background (or social class) is an important determinant of participation in post-compulsory education. Indeed, based on studies of a number of countries, Erikson & Goldthorpe (2002) argue that "children of more advantaged class origins take more ambitious educational options". Most of the evidence suggests that the returns to post-compulsory education are substantial (McIntosh 2006). Unless family background is also a strong determinant of the capacity of individuals to benefit from such education, this observation suggests that there is probably a large number of people who would have benefitted from post-compulsory education had they undertaken it. This leads to the conclusion that there are also likely to be substantial economic benefits to lifelong learning, that is to study undertaken after the completion of normal full-time education.

Study of this type is also fairly widespread. In the United Kingdom, about thirty per cent of both men and women with a degree-level qualification by age twenty-nine acquired it after having had a break from full-time education (Purcell, Wilton & Elias 2007). In 1994, 31 per cent of new undergraduates were aged twenty-five or over; by 2007 this proportion had risen to 43 per cent (Higher Education Statistics Agency 1995, Higher Education Statistics Agency 2008). Nor is the United Kingdom's situation unusual. Using a definition of lifelong learning broader than that adopted above, the UNESCO Institute for Lifelong Learning (2009) provides data on twenty-eight countries. In the Scandinavian countries and Finland 50-70% of adults aged 26-45 participated in organised learning and education. In the United Kingdom and New Zealand the rate was also above 50% while in Australia, Germany and the United States it was over 40%. Much lower rates, of just over 20% were observed in Chile and Hungary while in Portugal the rate was only about 15%. The pattern does vary by age. Apart from the Nordic countries, the United States has almost the highest participation rate for people aged 46-65.

Badescu & Saisana (2008) look at data drawn from European Labour Force surveys and find that in the United Kingdom in 2005 27.7 per cent of the population aged 25-64 received education or training in the four weeks preceding the survey, a proportion well ahead of the unweighted mean for the European Union (9.7 per cent) and exceeded only by Sweden with 33.4 per cent.

However, a number of studies suggest that, even if care is taken to compare like with like, lifelong learning is not as beneficial as conventional learning would have been to those who undertake it. In the United States, Light (1995) reports a range of penalties to interrupted education; these depend on the number of years of education before the interruption, the duration of the interruption and the total number of years of education. Holmlund, Liu & Skans (2008) come to similar conclusions for Sweden although they also suggest that the penalty is eroded with the passage of time. By contrast, Ferrer & Menendez (2009) suggest that, in Canada, graduates who delay their education receive a premium relative to those who do not. Looking at the United Kingdom, Egerton & Parry (2001) report substantial penalties for late learners. Jenkins, Vignoles, Wolf & Galindo-Rueda (2002) found that wage growth for people who underwent lifelong learning was generally not significantly faster over a ten-year period than for those who did not, with the implication that the former suffered a wage penalty compared to those who had obtained their qualifications without a break in their education. Purcell et al. (2007) provide case studies which illustrate the difficulty that mature graduates have had in finding "appropriate" employment. A study by Blanden, Buscha, Sturgis & Urwin (2010) finds no benefit to lifelong learning in aggregate for men although some evidence of benefit when looking at particular sub-categories. However, using an econometric approach that addresses the possible endogeneity of the decision to undertake lifelong learning de Coulon & Vignoles (2008) provide more encouraging evidence. Using British cohort data, they find strong and statistically significant positive effects of lifelong learning on wage growth for women aged 26-34.

In this paper we examine the effects of lifelong learning on women's earnings using a dynamic model of wage rates, albeit one with more structure than is found in the studies discussed above. We make contributions on both the methodological and substantive fronts, developing a modelling approach under which some people receive a wage that is a random draw from a stationary distribution, while others have a wage that is closely related to that of the previous year. Conceptually, this is a variant of the mover-stayer model (Goodman 1961) and the econometric framework we develop builds upon earlier research applying the model to income dynamics (Dutta, Sefton & Weale 2001). The first group – those whose wage is a random draw – are 'movers' in the sense that their position in the wage distribution is (conditionally) unrelated to their previous position. The second group are 'stayers' by analogous reasoning. Our approach uses its dual-regime structure to explain the observed variation in earnings. Essentially, the model allows wages to be estimated using linear regression, adjusted for employment status (which is observed) and whether the individual is a mover or a stayer (which is unobserved and is therefore identified probabilistically). In the case of movers, the regression is in levels while for the stayers the regression is in differences. We allow for the influence of unobserved heterogeneity on employment probability and movers' wages, through the inclusion of a specified number of mass points. This allows unobserved heterogeneity to be approximated in a nonparametric fashion (Heckman & Singer 1984). We include a correction term in the model to control for the potential endogeneity of lifelong learning decisions.

Both cross-sectional wage equations and wage equations in first differences are restricted forms of our more general model. Tests reported later in the paper reject the restrictions that are implied by these popular specifications. This result compounds the findings from Dutta et al. (2001) who showed that the mover-stayer structure offered a better means of understanding income inequality in the UK than did other popular specifications. The specification whereby women either receive a wage closely related to their previous wage or receive one drawn from a stationary distribution can, depending on the estimated parameters, accommodate the idea that high earners face higher earnings variance than do low earners. Furthermore, age-dependent transition probabilities may result in a higher degree of autoregression for one age- group than for another. These are both features of wage dynamics which Browning, Ejrnaes & Alvarez (2010) suggest are important. Finally, in a methodological advance, we extend the basic two-regime switching regression, where the regimes are endogenous but unobserved, by jointly modelling selection into employment. This last feature of our approach is essential for a study of women's earnings and distinguishes our work from most other studies of the earnings mobility; these restrict their analysis to the sub-sample of individuals with useable earnings data and do not address possible selection bias (for example, Blanden et al. (2010), Meghir & Pistaferri (2004), Ulrick (2008) and Browning et al. (2010)). We also address the potential endogeneity of lifelong learning decisions.

Substantively, the results further our understanding of the effectiveness of lifelong learning. In particular, by examining this within a mover-stayer model, we are able to identify the routes by which lifelong learning might affect wages. It becomes possible to assess not only whether lifelong learning affects wages directly but also whether it has a role in assigning individuals to be movers or stayers and thereby have their wages subject to differing sets of influences. Other analyses of lifelong learning have used regression techniques that do not permit such detailed insights. We base our analysis on the British Household Panel Survey, a nationally representative longitudinal survey dataset; we use the data spanning the period from 1991-2007.

The paper has the following structure. The next section describes our data and the pattern of lifelong learning shown by them. In section 3 we set out our econometric analysis. Section 4 presents the parameter estimates and in Section 5 we present simulation results to show the effect of lifelong learning. Section 6 discusses the relationship between our findings and other related work and Section 7 concludes.

2 Earnings, Employment and Lifelong Learning in the British Household Panel Survey

The British Household Panel Survey (BHPS) started in 1991 and is an annual survey of each adult member of a nationally representative sample of more than 5,000 households (around 10,000 individuals). Among other things, it provides information on employment status, pay, hours worked and educational attainment on a continuing basis. It is a longitudinal survey with the same individuals interviewed in each successive wave. If an individual leaves the original household, that individual together with all the adult members in their new household will also be interviewed. Children become eligible for interview when they reach the age of 16. The sample thus remains representative of the British population as it changes through the 1990s and 2000s.

We focus on data collected from the original sample households over seventeen waves from 1991 to 2007. Members of these households are repeatedly surveyed regardless of changes to household membership. Only women are considered (a companion paper looked at the effect of lifelong learning on men's wages (Dorsett, Lui & Weale 2010)). We limit ourselves to women aged 25 to 55 in order to concentrate on working lives beyond completion of the conventional period of education but to avoid the years leading up to the state pension age for women, which was 60 during our sample period. Thus, for those younger than 25 in 1991 or older than 55 in 2007, we consider only the data they provide while in this age range. We drop observations where individuals report themselves as self-employed because of the difficulties in defining their hourly wages. We also ignore those who provide proxy responses or whose data are incomplete while they are in this age range. Our sample is confined to those who respond in successive waves – where there is a break in response, that individual only features in our estimation sample up to the wave in which that break occurred. Finally, we trim the data to remove the observations whose reported hourly wages fall into the top and bottom 1% of the distribution.

In our analysis we define lifelong learning as the acquisition of any qualifications after the age of 25. This age threshold was chosen in order to allow for a period to elapse following the completion of full-time education for most people. We focus on qualification acquisition rather than participation in training since this is more fully recorded in the data but also since this has merit in its own right. Our analysis differs from the approach adopted in some related work (Taniguchi 2005, Hällsten 2010) in that we look at average effects from people upgrading their qualifications from one level to a higher one, or acquiring qualifications without upgrading their qualification levels rather than the premium associated with the acquisition of a particular qualification. The advantage of this approach is that, in our data set, there are more cases of upgrading and acquisition in general than there are of upgrading to any particular qualification (such as a degree) and thus our prospects for identifying effects of lifelong learning are enhanced. The drawback is, of course, that any effect identified is only an average. But this objection also applies, at least to some extent, if one focuses on a particular qualification since the benefit resulting from its acquisition almost certainly depends on the previous educational attainment of the individual in question. Ideally one would analyse each possible transition in educational status separately. However, there are too few observations for such an approach to be practical.

We look at the effects of lifelong learning undertaken in each of the last five years and also if it has been undertaken since our respondent entered the sample, i.e. since 1991 or after reaching the age of twenty-five, whichever comes later. In our econometric work we look only at wage dynamics from 1996 onwards; this means that we have a full record of lifelong learning in the last five years for everyone in our sample. We also know whether they have undertaken it since 1991 or, if later, since they reached the age of twenty-five. The BHPS does not, however, tell us about people who undertook lifelong learning before the first wave of the survey in 1991.

2.1 The Pattern of Lifelong Learning

The BHPS provides very detailed information on qualifications. These were classified to match the national scale which ranges from 0 (for those with no or only minimal qualifications) to 5 for those with post-graduate degrees. The system was originally designed to represent national vocational qualifications (NVQs) but academic qualifications have also been calibrated against it, allowing most qualifications to be represented on an equal basis. We note that, using this or indeed any categorical classification of qualifications means that the acquisition of a qualification is not necessarily associated with an increase in qualification level. In common with other work (e.g. Blanden et al. (2010)) we merge categories 4 and 5. Our classification of qualifications[1] is shown in table 1.

Table 2 provides a summary picture of the extent of lifelong learning. The main panel of the table compares individuals' highest current qualifications when first observed to their highest qualification five years later. This captures the prevalence of lifelong learning that results in qualification upgrading. The row below the transition table shows the probability of doing some kind of lifelong learning to rise with the level of initial qualification. Roughly one-tenth of those with no qualifications initially undertook lifelong learning during the years observed. For those with intermediate-level qualifications the incidence was about one-fifth, while for those with the highest initial qualifications it was closer to one-third. The gradient with regard to upgrading was less clear. Overall, five per cent upgraded.

In our subsequent analysis we focus our attention on two variables, first whether someone has acquired a qualification and secondly, if they did, whether it led to an upgrade of their qualification level.

[1] This classification differs slightly from the National Qualifications Framework which classes GSCEs at grades D to G as level 1 and grades A* to C as level 2.

Level 1
Youth training certificate
Trade appenticeship
Clerical and commercial qualifications
City and Guilds Certification Part I
SCOTVEC National Certiciate Modules
NVQ/SVQ level 1
GCSEs
SCEs grade D-E or 4-5
O grades A-C or 1-3
Standard grades 4-7
CSEs
O-levels (pre-1975), OLs (post-1975)
SLCs
Level 2
City and Guilds Certification Part II
SCOTVEC Higher National Units
NVQ/SVQ level 2
CPVE
1 A level
Standard grades 1-3
GNVQ
AS level
School Certificate or Matriculation
1 Higher School Certificate
Level 3
City and Guilds Certification Part III
SCOTVEC National Certificate or Diploma
ONC, OND, BEC/TEC/BTEC General Certificate
NVQ/SVQ level 3
2 or more A levels
2 or more Higher School Certificates
Higher grades
Certificate of 6th year studies
Level 4
HNC, HND, BEC/TEC/BTEC/SCOTVEC Higher Certificate or Higher Diploma
NVQ/SVQ level 4
Nursing qualifications (e.g. SEN, SRN, SCM, RGN)
Teaching qualification
University diploma or Foundation degree
University or CNAA First Degree (e.g. BA, B.Ed, BSc)
University or CNAA Higher Degree (e.g. MSc, PhD)

Table 1: The Classification of Qualifications

		Initial qualification level					
		0	1	2	3	4	All
Qualification	0	92.50	0.00	0.00	0.00	0.00	26.15
level	1	4.13	94.66	0.00	0.00	0.00	41.59
five	2	1.50	1.12	95.29	0.00	0.00	5.20
years	3	1.13	1.99	4.71	94.66	0.00	7.96
later	4	0.75	2.24	0.00	5.34	100.00	19.10
Upgrading		7.50	5.34	4.71	5.34		4.99
Lifelong learning		9.57	17.27	18.82	23.66	30.21	17.88
N		533	805	85	131	331	1,885

Table 2: Transition Probabilities (per cent) over a Five-year Window and the Incidence of Lifelong Learning

2.2 Employment, Wages and Lifelong Learning

The BHPS did not introduce an explicit question on hourly pay until wave 8. However, in all waves it asks employees to give information on the number of hours they work in a normal week and the number of hours they worked as overtime. The survey also collects usual monthly earnings before tax and other deductions in employees' current main job[2]. For all waves, we derive each employee's gross hourly wage as follows:

$$hourly\ wage = \frac{monthly\ earnings}{\frac{52}{12} \times (weekly\ regular\ hours + 1.5 \times weekly\ overtime\ hours)} \tag{1}$$

We use the calendar year average of the Retail Price Index excluding mortgage interest payments (RPIX) to deflate nominal wages to 2007 prices. We refer to this deflated variable as the hourly wage.

Table 3 provides a summary of average hourly wages and non-employment rates for the women in our sample, differentiating between those with no lifelong learning, those who undertake lifelong learning without upgrading their highest level of qualification and those who do upgrade their highest level of qualification as a result of lifelong learning. This shows that wages mostly increase with qualification level. Employment, on the other hand, is lowest among those with no initial qualifications but among those with some qualifications, the relationship is less clear. It is not the case that higher levels

[2] This is a derived variable wPAYGU.

of qualification are monotonically associated with higher probabilities of employment. This differs from what Dorsett et al. (2010) had found for men, with more qualified men more likely to be employed than less qualified men. The fact that such a pattern does not exist for women may reflect the fact that women more commonly have periods out of the labour market while they bring up children. Alternatively, it may indicate a stronger income effect for women than men. More directly of interest is the apparent effect of lifelong learning on wages and employment. Lifelong learning with no qualification upgrade is associated with higher wages for those with no qualifications, but the effect for those with some qualifications is smaller or, for those with level 3, non-existent. Where qualifications are upgraded as a result of lifelong learning, the apparent premium is larger. This is particularly the case for those initially with level 2 qualifications (interestingly, the same was found for men). Turning to the effect of lifelong learning on the probability of being employed, the impression from Table 3 is that acquiring a qualification goes a long way towards removing the variation across the different qualification groups in the chances of working. Among those who acquired a new qualification, the probability of employment is lowest (79 per cent) among those with no qualifications initially and highest (86 per cent) among the most qualified initially. The difference between these two groups (7 percentage points) is considerably smaller than the difference among those who do not undertake any lifelong learning (27 percentage points). The effect of upgrading is less equalising. Among those with no qualifications initially who upgrade, the employment rate is 71 per cent, while for those with higher initial qualification levels, the employment rate ranges between 83 and 87 per cent.

With this background we can now proceed to our econometric analysis.

3 Econometric Analysis

In this section, we discuss in more detail the mover-stayer model, describe the econometric approach and present estimation results. We begin by considering the variables

Initial education level	No lifelong learning	With qualification but not upgrading	With upgrading	Total
Number				
0	3140	265	434	3,839
1	5092	1525	811	7,428
2	522	219	48	789
3	940	418	135	1,493
4	1640	1337		2,977
All	11334	3764	1428	16,526
Wages (2007 Price)				
0	£6.24	£6.82	£7.16	
1	£7.81	£8.20	£8.96	
2	£7.55	£7.93	£12.59	
3	£9.84	£9.26	£11.01	
4	£12.79	£13.80		
Non-employment rates				
0	54.27%	20.75%	29.49%	
1	31.52%	15.41%	13.32%	
2	40.61%	15.98%	16.67%	
3	30.32%	16.27%	14.07%	
4	28.05%	14.36%		

Table 3: Summary Data: Intial Qualifications, Earnings, Employment and Lifelong Learning, 1996-2008 Average. Pooled Data

to include in the analysis.

3.1 Variables used in the Analysis

The main variables of interest are those that relate to lifelong learning. We are concerned with both the short- and long-term effects of lifelong learning and wish to distinguish people who upgrade their level of qualification from those who gain qualifications but at a level equal to or below those of their existing highest qualifications.

With this aim in mind, we set up a range of dummy variables to reflect lifelong learning history. $Acquired_{t-i}$ takes a value of 1 if someone acquired a qualification between the interview year $t-i-1$ and the interview year $t-i$ $(i = 0, 1)$ whether they upgraded their educational status or not, while $Upgraded_{t-i}$ takes a value 1 if they acquired a qualification which upgraded their educational status; otherwise these variables take the value of 0. We also use the subscript $t-2+$ to indicate acquisition of a qualification two or more years ago. *Ever Acquired* and *Ever Upgraded* summarise these, taking a value of 1 for people whom the data set shows to have acquired qualifications at some time in the past.

We include additional variables in the analysis to control for other sources of variation within our sample. These include: qualification level when first observed (*Orig Qual 1-Orig Qual 4*); a dummy variable indicating whether the highest qualification at that time was academic (*Highest Qual Academic*); age; whether from an ethnic minority group or not; marital status (single or partnered); the presence of children aged 0-1, 2-5, 6-10 or 11-16 (all represented by 0/1 dummy variables); region (using dummies to indicate the region within Britain people live in); whether the individual was employed when first observed; whether a new job was started within the last year; log GDP or its change as an indicator of the state of the economy; and the time between interviews. Apart from those relating to lifelong learning, the regressors included in the model are either exogenous (age, ethnic group, wave of survey) or relate to an earlier time period in order to reduce concerns about endogeneity. Some of these variables were excluded from particular equations in order to assist with identification of the model, as described later.

3.2 A Mover-Stayer Framework

This section sets out the framework within which we explore the effect of lifelong learning on earnings. We assume that, at any given time, workers can be characterised as either movers or stayers. Movers are so called because they move about the wage distribution; they receive a wage rate possibly very different from what they had previously earned. Stayers, by contrast, stay at much the same point in the wage distribution as they had been in the previous year; thus their wage rates are closely explained by the previous year's wage rates.

There are a number of possible reasons why people might be movers. Perhaps the most obvious is that they lose their jobs and have to take whatever the labour market offers, with or without a period of unemployment in between. But they may also be people who have been in stagnant jobs with little prospect for progression who have the good fortune to come across more favourable labour market opportunities. Or people who have done reasonably well but still find that a better opportunity has come along. Being a mover need not even be associated with a change of employer. It is perfectly possible that people will move from one post to another offering sharply better pay within the same employer. It is rather less likely that someone's wage rate will fall sharply while they remain with the same employer, if for no other reason that such a change would be likely to appear as constructive dismissal. Nevertheless, one might expect to see some connection between being a mover and a change of job.

While there may be a number of ways in which movers and stayers could be defined, the approach we adopt is that movers are assumed to receive a wage rate set by a standard Mincerian wage equation in the levels of wages. For these movers the wage rate of the previous period has no bearing on the current wage rate except, of course, insofar as both are affected by the same individual characteristics, such as the level of education. For stayers by contrast, the idea that the wage rate is closely related to that of the previous period points naturally to their wages being determined by an equation with the structure of equation (3) below, in the first difference of log earnings.

Since mover-stayer status is not observed, we determine it statistically. To do this, we assume it is driven by a latent variable in a probit model, in much the same way that it is commonly assumed that employment is driven by a latent variable. The estimated model allows us to determine the probability that particular observations are those of stayers rather than movers or *vice versa* just as a probit model can be used to identify the probability that someone will be employed. Strictly, it is obviously impossible for someone who was previously recorded as not employed to be a stayer; her wage rate cannot be closely related to that of the previous period because there was no wage rate in the previous period. The model is specified so that such women can still be classified as stayers. Rather than allowing current wages to be related to wages in the previous year, they are allowed to be related to wages when last employed.

Econometrically, our model can be seen as a switching regression in which the two distinct states cannot be identified except through estimation of the model and is of the type first discussed by Quandt (1958). Over and above this, however, we have to extend the model to take account of selection into employment. The fact that our model includes an equation in first differences might suggest that it encompasses a model specified in this way. In fact, for our model to reduce to the model in first differences we require i) there are no selection effects from employment present, ii) all earnings of people who were not employed in the previous period can be explained by the movers' equation (which our model in any case requires) and iii) all earnings of people employed in the previous period can be explained by the stayers' equation.

We acknowledge that people who study for lifelong qualifications may have an earnings capacity different from those who do not do so. We include an adjustment term to correct for this potential endogeneity. This term is the generalised residual resulting from the estimation of an ordered probit equation with a dependent variable corresponding to one of three possible cases: no lifelong learning, lifelong learning without an upgrade, and lifelong learning with an upgrade. This is discussed in more detail below.

Lastly, we noted above that a virtue of the first difference model was that it removed individual fixed effects associated with the level of earnings. The movers' equation is in the level of log wages and so does not control for such effects. We address this by including a number of discrete mass points in the movers' equation in order to approximate unobserved heterogeneity, as we explain in the next section. The issue also arises in the employment equation. The sort of people who undertake lifelong learning at some time in their lives may be more or less likely to be employed than those who do not. We also include mass points in the employment equation; we are, however, unable to explore whether they play a role in the switching equation.

We now set out the components of the mover-stayer model.

3.3 Movers

For movers, wages are given by a stationary Mincerian equation

$$y_{it} = X_{it}\beta_1 + m_{it} + u_{1it} \tag{2}$$

where y_{it} represents log hourly wages deflated by the retail price index and X_{it} is a vector of variables which influence the wage rate. Such variables include age, qualifications, lifelong learning, region of residence and log real GDP *per capita*. Thus, for a mover, the wage rate is not directly related to previous wages except insofar as the variables which influence the wage of a mover have also influenced their wage on the previous occasion when they were a mover. An individual-specific effect, $m_{it} \in \{m^1, m^2, ..., m^K\}$, is included to capture the influence of unobserved characteristics on wages. As discussed below, this provides a discrete approximation of the distribution of unobserved heterogeneity.

3.4 Stayers

The hourly earnings of stayers are assumed to be related to those of the previous period. We specify the stayers' wage equation as

$$\Delta y_{it} = X_{it}\beta_2 + u_{2it} \tag{3}$$

It should be noted that there is no loss of generality in specifying the vector of driving variables X_{it} to be the same in both equations; provided it is general enough, differences in specification can be accommodated by restrictions on the elements of β_1 and β_2.Since this equation is estimated in differences, the individual-specific effect on the level of wages is swept out. If someone was not employed in the previous period, then the change is measured relative to the last observed wage rate.

3.5 Switching

A respondent is a mover if the indicator variable $I_{it} = 1$ and a stayer if $I_{it} = 2$. This indicator is driven by the latent variable, I_{it}^*. The probability, P_{it} that observation y_{it} is drawn from (3) rather than (2) is driven by the latent variable

$$I_{it}^* = Z_{it}\gamma + \varepsilon_{it} \tag{4}$$

with $I_{it} = 1$ if $I_{it}^* \leq 0$ and $I_{it} = 2$ if $I_{it}^* > 0$.

3.6 Selection into Employment

We address the issue of selection into employment in the following way. Someone is employed if the indicator $J_{it} = 1$ and not employed if $J_{it} = 0$. This indicator is driven by the variable

$$J_{it}^* = W_{it}\delta + e_{it} + \eta_{it} \tag{5}$$

with $J_{it} = 1$ if $J_{it}^* > 0$ and $J_{it} = 0$ if $J_{it}^* \leq 0$. W_{it} is a vector of variables which drives the employment choice. An individual-specific effect, $e_{it} \in \left\{e^1, e^2, ..., e^K\right\}$, is included to capture the influence of unobserved characteristics on the probability of being employed.

3.7 Lifelong Learning

Our analysis needs to take account of the consequences of potential endogeneity of lifelong learning decisions. We distinguish lifelong learning which results in upgrading qualifications from lifelong learning which results in no such upgrade. Someone undertakes lifelong learning with upgrading if $K_{it} = 2$, lifelong learning without upgrading if $K_{it} = 1$ and does not do so if $K_{it} = 0$. This process is driven by the latent variable

$$K_{it}^{*} = V_{it}\zeta + \nu_{it} \tag{6}$$

with $K_{it} = 2$ if $K_{it}^{*} > \bar{K}_{2t} \geq 0$, $K_{it} = 1$ if $\bar{K}_{2t} > K_{it}^{*} \geq 0$ and $K_{it} = 0$ if $K_{it}^{*} < 0$

Our approach to dealing with potential endogeneity of lifelong learning decisions is discussed below.

3.8 Estimation Strategy

The model has the following likelihood function:

$$\begin{aligned} L_{it} = & \prod_{I_{it}\in 1,2,J_{it}=1} \Big\{ F\left(\eta_{it} > -W_{it}\delta - e_{it}, \varepsilon_{it} > -Z_{it}\gamma\right) f\left(u_{1it} \mid \eta_{it} > -W_{it}\delta - e_{it}, \varepsilon_{it} > -Z_{it}\gamma\right) \\ & + F\left(\eta_{it} > -W_{it}\delta - e_{it}, \varepsilon_{it} \leq -Z_{it}\gamma\right) f\left(u_{2it} \mid \eta_{it} > -W_{it}\delta - e_{it}, \varepsilon_{it} \leq -Z_{it}\gamma\right) \Big\} \\ \times & \prod_{J_{it}=0} F\left(\eta_{it} \leq -W_{it}\delta - e_{it}\right) \end{aligned} \tag{7}$$

We allow the error terms to be freely correlated across equations and assume a multivariate normal distribution: $(u_{1it}, u_{2it}, \varepsilon_{it}, \eta_{it}) \sim N(0, \Sigma)$ where

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{1\varepsilon} & \sigma_{1\eta} \\ & \sigma_2^2 & \sigma_{2\varepsilon} & \sigma_{2\eta} \\ & & 1 & \sigma_{\varepsilon\eta} \\ & & & 1 \end{bmatrix} \tag{8}$$

Note that σ_{12} is not estimable (Maddala 1983, p. 224) since individuals cannot be simultaneously in two states.

Consider the case of $I_{it} = 1$. The truncated normal density is

$$f\left(u_{1it},\varepsilon_{it},\eta_{it}\mid\eta_{it}>-W_{it}\delta-e_{it},\varepsilon_{it}\leq-Z_{it}\gamma\right) = \frac{f\left(u_{1it},\varepsilon_{it},\eta_{it}\right)}{\Phi\left(W_{it}\delta+e_{it},Z_{it}\gamma,\rho_{\varepsilon\eta}\right)} = \frac{f\left(u_{1it}\right)f\left(\varepsilon_{it},\eta_{it}\mid u_{1it}\right)}{\Phi\left(W_{it}\delta+e_{it},Z_{it}\gamma,\rho_{\varepsilon\eta}\right)} \tag{9}$$

where $\Phi()$ represents the cumulative standard normal distribution. Integrate over $\varepsilon_{it},\eta_{it}$ to get the marginal truncated density for u_{1it}

$$f\left(u_{1it}\mid\eta_{it}>-W_{it}\delta-e_{it},\varepsilon_{it}\leq-Z_{it}\gamma\right)=\frac{f\left(u_{1it}\right)\int_{-W_{it}\delta}^{\infty}\int_{0}^{-Z_{it}\gamma}f\left(\varepsilon_{it},\eta_{it}\mid u_{1it}\right)\mathrm{d}\varepsilon_{it}\mathrm{d}\eta_{it}}{\Phi\left(W_{it}\delta+e_{it},-Z_{it}\gamma,\rho_{\varepsilon\eta}\right)} \tag{10}$$

noting that

$$f\left(\varepsilon_{it},\eta_{it}\mid u_{1it}\right)\sim N\left(\left(\begin{array}{c}\frac{\rho_{1\varepsilon}}{\sigma_1}\left(y_{it}-X_{it}\beta_1-m_{it}\right)\\ \frac{\rho_{1\eta}}{\sigma_1}\left(y_{it}-X_{it}\beta_1-m_{it}\right)\end{array}\right),\left(\begin{array}{cc}1-\rho_{\varepsilon1}^2 & \sigma_{\varepsilon\eta}-\rho_{1\varepsilon}\rho_{1\eta}\\ & 1-\rho_{\eta1}^2\end{array}\right)\right) \tag{11}$$

where $\rho_{1\varepsilon}=\frac{\sigma_{1\varepsilon}}{\sigma_1}$ and $\rho_{1\eta}=\frac{\sigma_{1\eta}}{\sigma_1}$. Since $\rho_{\varepsilon\eta}=\sigma_{\varepsilon\eta}$we can write

$$\begin{aligned}&f\left(u_{1it}\mid\eta_{it}>-W_{it}\delta-e_{it},\varepsilon_{it}\leq-Z_{it}\gamma\right) \\ = &\frac{\Phi\left(-\frac{Z_{it}\gamma+\frac{\rho_{1\varepsilon}}{\sigma_1}\left(y_{it}-X_{it}\beta_1-m_{it}\right)}{\sqrt{1-\rho_{1\varepsilon}^2}},\frac{W_{it}\delta+e_{it}+\frac{\rho_{1\eta}}{\sigma_1}\left(y_{it}-X_{it}\beta_1-m_{it}\right)}{\sqrt{1-\rho_{1\eta}^2}},-\frac{\rho_{\varepsilon\eta}-\rho_{1\varepsilon}\rho_{1\eta}}{\sqrt{1-\rho_{1\varepsilon}^2}\sqrt{1-\rho_{1\eta}^2}}\right)\phi\left(\frac{y_{it}-X_{it}\beta_1-m_{it}}{\sigma_1}\right)/\sigma_1}{\Phi\left(W_{it}\delta+e_{it},-Z_{it}\gamma,-\rho_{\varepsilon\eta}\right)}\end{aligned} \tag{12}$$

Doing the same kind of thing for the case of $I_{it}=2$ results in

$$\begin{aligned}&f\left(u_{2it}\mid\eta_{it}>-W_{it}\delta-e_{it},\varepsilon_{it}>-Z_{it}\gamma-s_{it}\right) \\ = &\frac{\Phi\left(\frac{Z_{it}\gamma+\frac{\rho_{2\varepsilon}}{\sigma_2}\left(\Delta y_{it}-X_{it}\beta_2\right)}{\sqrt{1-\rho_{2\varepsilon}^2}},\frac{W_{it}\delta+e_{it}+\frac{\rho_{2\eta}}{\sigma_2}\left(\Delta y_{it}-X_{it}\beta_2\right)}{\sqrt{1-\rho_{2\eta}^2}},\frac{\rho_{\varepsilon\eta}-\rho_{2\varepsilon}\rho_{2\eta}}{\sqrt{1-\rho_{2\varepsilon}^2}\sqrt{1-\rho_{2\eta}^2}}\right)\phi\left(\frac{\Delta y_{it}-X_{it}\beta_2}{\sigma_2}\right)/\sigma_2}{\Phi\left(W_{it}\delta+e_{it},Z_{it}\gamma,\rho_{\varepsilon\eta}\right)}\end{aligned} \tag{13}$$

Substituting back into the likelihood function, the denominator terms cancel out giving:

$$
\begin{aligned}
L_{it} &= \prod_{I_{it} \in 1,2, J_{it}=1} \left\{ \Phi(-\frac{Z_{it}\gamma + \frac{\rho_{1\varepsilon}}{\sigma_1}(y_{it} - X_{it}\beta_1 - m_{it})}{\sqrt{1-\rho_{1\varepsilon}^2}}, \frac{W_{it}\delta + e_{it} + \frac{\rho_{1\eta}}{\sigma_1}(y_{it} - X_{it}\beta_1 - m_{it})}{\sqrt{1-\rho_{1\eta}^2}}, \right. \\
& - \frac{\rho_{\varepsilon\eta} - \rho_{1\varepsilon}\rho_{1\eta}}{\sqrt{1-\rho_{1\varepsilon}^2}\sqrt{1-\rho_{1\eta}^2}}) \qquad (14) \\
& \phi\left(\frac{y_{it} - X_{it}\beta_1 - m_{it}}{\sigma_1}\right) / \sigma_1 \\
&+ \Phi\left(\frac{Z_{it}\gamma + \frac{\rho_{2\varepsilon}}{\sigma_2}(\Delta y_{it} - X_{it}\beta_2)}{\sqrt{1-\rho_{2\varepsilon}^2}}, \frac{W_{it}\delta + e_{it} + \frac{\rho_{2\eta}}{\sigma_2}(\Delta y_{it} - X_{it}\beta_2)}{\sqrt{1-\rho_{2\eta}^2}}, \frac{\rho_{\varepsilon\eta} - \rho_{2\varepsilon}\rho_{2\eta}}{\sqrt{1-\rho_{2\varepsilon}^2}\sqrt{1-\rho_{2\eta}^2}}\right) \\
& \left. \phi\left(\frac{\Delta y_{it} - X_{it}\beta_2}{\sigma_2}\right) / \sigma_2 \right\} \\
&\times \prod_{J_{it}=0} \Phi\left(-W_{it}\delta - e_{it}\right) \qquad (15)
\end{aligned}
$$

This likelihood contribution includes the individual-specific effects m_{it}. These can be integrated out. To do this, the distribution of unobserved heterogeneity is approximated by a finite number, K, of mass points in the mover and employment equations. The specification characterises women as represented by one of K such mass points each of which occurs with probability p_k. This is tantamount to allowing the population to be made up of K types of women. Each type of woman differs but every woman of a given type is identical with regard to those unobserved characteristics thought to influence wages. The number of types of women, K, is unknown but is chosen on the basis of likelihood ratio tests. Writing the likelihood contribution for woman i at time t that would obtain were she of type k as L_{it}^k, her unconditional contribution to the likelihood at this time is:

$$L_{it} = \sum_{k=1}^{K} p_k L_{it}^k$$

This approach to controlling for individual-specific effects is flexible in the sense that it avoids the need to assume a particular distribution of unobserved heterogeneity. It is also flexible in the sense that, while it assumes the distribution of unobserved heterogeneity can be approximated by a finite number of mass points, it does not restrict each woman to be characterised by the same mass point at all times. This property becomes increasingly desirable the longer the period covered by the data.

The model is estimated using maximum likelihood on a pooled dataset. As well as including mass points in the employment and movers' equations in order to capture the effect of unobserved heterogeneity, the effect of correlation across waves for individual respondents was addressed by allowing for clustering in the computation of standard errors. Strictly, we maximise a log pseudolikelihood.

3.9 Identification

Strictly, the assumed error structure achieves model identification without the need for exclusion restrictions. In practice, though, exclusion restrictions were imposed. This was done in order to provide a more intuitive basis for identification and to reduce reliance on arbitrary assumptions about errors. In common with numerous other studies, variables appearing in the employment equation only are family background variables – whether partnered at time $t-1$ and the ages of children present in the household at that time. Variable in the switching equation but not in the wages equations include the *Wave Gap* which indicates the interval between interviews and *Recent Job* indicating whether the current job has started since the previous interview. Here, the rationale is that, people are more likely to be movers if the gap between interviews is long than if it is short and that those with a recent job are more likely to have experienced a wages shock that would be likely to classify them as movers.

As mentioned previously, the decision to undertake lifelong learning is potentially endogenous so the estimation needs take account of possible correlations between ν_{it} in equation (6) and the errors in the four equations (2 to 5) of the main model. Ideally, this would be dealt with by jointly estimating all five equations. However, to avoid the computational burden involved with high-order normal integrals we use instead a two-step approach. This follows in the spirit of Kim (2004) who considers the case of a Markov switching model with an endogenous continuous regressor in the outcome equations (the equivalent of our Δy_{it} and y_{it} equations). The full covariance matrix can be written

$$Cov\left(\nu_{it},\eta_{it},\varepsilon_{it},u_{2it},u_{1it}\right)=\left[\begin{array}{ccccc} 1 & \sigma_{\eta\nu} & \sigma_{\varepsilon\nu} & \sigma_{2\nu} & \sigma_{1\nu} \\ \sigma_{\eta\nu} & 1 & \sigma_{\varepsilon\eta} & \sigma_{2\eta} & \sigma_{1\eta} \\ \sigma_{\varepsilon\nu} & \sigma_{\varepsilon\eta} & 1 & \sigma_{2\varepsilon} & \sigma_{1\varepsilon} \\ \sigma_{2\nu} & \sigma_{2\eta} & \sigma_{2\varepsilon} & \sigma_2^2 & \sigma_{12} \\ \sigma_{1\nu} & \sigma_{1\eta} & \sigma_{1\varepsilon} & \sigma_{12} & \sigma_1^2 \end{array}\right]$$

Applying a Cholesky decomposition, we can recast the error terms in such a way that the correlation structure is maintained:

$$\left[\begin{array}{c} \nu_{it} \\ \eta_{it} \\ \varepsilon_{it} \\ u_{2it} \\ u_{1it} \end{array}\right]=\left[\begin{array}{ccccc} b_{11} & 0 & 0 & 0 & 0 \\ b_{21} & b_{22} & 0 & 0 & 0 \\ b_{31} & b_{32} & b_{33} & 0 & 0 \\ b_{41} & b_{42} & b_{43} & b_{44} & 0 \\ b_{51} & b_{52} & b_{53} & b_{54} & b_{55} \end{array}\right]\left[\begin{array}{c} \omega_{1it} \\ \omega_{2it} \\ \omega_{3it} \\ \omega_{4it} \\ \omega_{5it} \end{array}\right]$$

where $(\omega_{1it,}\omega_{2it,}\omega_{3it,}\omega_{4it,}\omega_{5it})$ are independent standard normal variables.

This allows our model to be written in reverse order as:

$$\begin{aligned}
K_{it}^* &= V_{it}\varphi + b_{11}\omega_{1it} \\
J_{it}^* &= \widetilde{W_{it}\delta} + e_{it} + \Delta_{41}(L)K_{it,1} + \Delta_{42}(L)K_{it,2} + b_{21}\omega_{1it} + b_{22}\omega_{2it} \\
I_{it}^* &= \widetilde{Z_{it}\gamma} + \Delta_{31}(L)K_{it,1} + \Delta_{32}(L)K_{it,2} + b_{31}\omega_{1it} + b_{32}\omega_{2it} + b_{33}\omega_{3it} \\
\Delta y_{it} &= \widetilde{X_{it}\beta}_2 + \Delta_{21}(L)K_{it,1} + \Delta_{22}(L)K_{it,2} + b_{41}\omega_{1it} + b_{42}\omega_{2it} + b_{43}\omega_{3it} + b_{44}\omega_{4it} \\
y_{it} &= \widetilde{X_{it}\beta}_1 + m_{it} + \Delta_{11}(L)K_{it,1} + \Delta_{12}(L)K_{it,1} + b_{51}\omega_{1it} + b_{52}\omega_{2it} + b_{53}\omega_{3it} + b_{54}\omega_{4it} + b_{55}\omega_{5it}.
\end{aligned}$$

Here $\Delta_{ij}(L)$ are lag operators and $K_{it,j}$ are dummy variables which take the value 1 if $K_{it}=j$ and 0 otherwise $(j=1,2)$ The tildes indicate the removal of K_{it} from the respective regressor set. Endogeneity of $K_{it,j}$ stems from their correlation with ω_{1it}. We can substitute this out to give

$$
\begin{aligned}
J_{it}^{*} &= \widetilde{W_{it}\delta} + e_{it} + \Delta_{41}(L)K_{it,1} + \Delta_{42}(L)K_{it,2} + \frac{b_{21}}{b_{11}}\left(K_{it}^{*} - V_{it}\varphi\right) + b_{22}\omega_{2it} \\
I_{it}^{*} &= \widetilde{Z\gamma} + \Delta_{31}(L)K_{it,1} + \Delta_{32}(L)K_{it,2} + \frac{b_{31}}{b_{11}}\left(K_{it}^{*} - V_{it}\varphi\right) + b_{32}\omega_{2it} + b_{33}\omega_{3it} \\
\Delta y_{it} &= \widetilde{X_{it}\beta_2} + \Delta_{21}(L)K_{it,1} + \Delta_{22}(L)K_{it,2} + \frac{b_{41}}{b_{11}}\left(K_{it}^{*} - V_{it}\varphi\right) + b_{42}\omega_{2it} + b_{43}\omega_{3it} + b_{44}\omega_{4it} \\
y_{it} &= \widetilde{X_{it}\beta_1} + m_{it} + \Delta_{11}(L)K_{it,1} + \Delta_{12}(L)K_{it,1} + \frac{b_{51}}{b_{11}}\left(K_{it}^{*} - V_{it}\varphi\right) + b_{52}\omega_{2it} + b_{53}\omega_{3it} + b_{54}\omega_{4it} \\
&\quad + b_{55}\omega_{5it}
\end{aligned}
$$

Kim's approach addresses the case of a continuous endogenous regressor and involves a two-step procedure. In the first step, a reduced form equation for the endogenous regressor is estimated. In the second step, the primary equation is estimated, with the residual from the first-step equation included as an additional regressor. Our case is slightly different in that the potentially endogenous regressor – the acquisition of a lifelong learning qualification – is categorical rather than continuous. Following Vella & Verbeek (1999) and Orme (2001), we replace the $(K_{it}^{*} - V_{it}\varphi)$ with the generalised residual from the K_{it}^{*} regression, $\bar{\nu}_{it}$. Since $\bar{\nu}_{it}$ is correlated with ω_{1it} but not with ω_{kit} for $k > 1$, inclusion of this term as a regressor in each of the other equations controls for the endogeneity of K_{it}. Since the ω_{kit} terms are independent standard normal, our model becomes:

$$
\begin{aligned}
J_{it}^{*} &= W_{it}\delta + e_{it} + \Delta_{41}(L)K_{it,1} + \Delta_{42}(L)K_{it,2} + \frac{b_{21}}{b_{11}}\bar{\nu}_{it} + \zeta_{4it} \\
I_{it}^{*} &= Z_{it}\gamma + \Delta_{31}(L)K_{it,1} + \Delta_{32}(L)K_{it,2} + \frac{b_{31}}{b_{11}}\bar{\nu}_{it} + \zeta_{3it} \\
\Delta y_{it} &= X_{it}\beta_2 + \Delta_{21}(L)K_{it,1} + \Delta_{22}(L)K_{it,2} + \frac{b_{41}}{b_{11}}\bar{\nu}_{it} + \zeta_{2it} \\
y_{it} &= X_{it}\beta_1 + m_{it} + \Delta_{11}(L)K_{it,1} + \Delta_{12}(L)K_{it,1} + \frac{b_{51}}{b_{11}}\bar{\nu}_{it} + \zeta_{1it}
\end{aligned}
$$

Now

$$Cov\left(\zeta_{4it}, \zeta_{3it}, \zeta_{2it}, \zeta_{1it}\right) = \mathbf{CC}', \text{ where } \mathbf{C} = \begin{bmatrix} b_{22} & 0 & 0 & 0 \\ b_{32} & b_{33} & 0 & 0 \\ b_{42} & b_{43} & b_{44} & 0 \\ b_{52} & b_{53} & b_{54} & b_{55} \end{bmatrix}$$

As with the linear case, the coefficients on the generalised residual terms provide a statistical test of endogeneity.

This two-step approach raises additional identification issues. Statistical identification is achieved through the non-linearity of the generalised residual terms. However, we also impose an exclusion restriction in order provide an additional basis for identification that can be argued on more economic grounds. Specifically, we include year dummies in the lifelong learning (first-step) regression but not in the main (second-step) model. The justification for this is that these dummies can capture changes over time in the probability of undertaking lifelong learning, perhaps reflecting shifts in the policy environment. Furthermore, if we believe that it is predominantly the strength of the economy that influences employment and earnings, year dummies should not affect the dependent variables in the other equations of our model since they all control for the effect of GDP. Hence, the exclusion restriction may be justified. It should be noted that a strength of the approach taken here is that the year dummies are unarguably exogenous; something that is not guaranteed with all variables used as instruments.

4 Estimation Results

We begin by showing, in table 4, the results of the ordered probit regression (equation 6), which determines whether women undertake no lifelong, undertake lifelong learning without upgrading their qualifications or undertake lifelong learning and, in doing so, upgrade their qualifications. The results suggest that initial qualification level is a significant determinant of undertaking lifelong learning. So too is the possession of qualifications which do not fit into the grading scheme. The probability of undertaking lifelong learning increases non-linearly with age, peaking at 36 years. Being employed at the start of the survey increases the chance of undertaking lifelong learning. As noted, year dummies serve to strengthen the identification of the model. Individually, none of these dummies is significant. Taken together, the p-value of their joint significance is 0.1266. This falls somewhat short of conventional thresholds for statistical significance and the results for the main model need to be considered with this in mind. However, the available data offer no credible alternative instruments.

Our main results are presented in table 5. The only statistically significant effect we see in the two earnings equations is that women who upgrade their qualifications earn as movers, on average 0.13 log units more than those who acquire qualifications without upgrading. Qualifications of the latter type on average raise movers' pay by 0.036 log units but this is not statistically significant at a 10 per cent level. Stayers who acquire qualifications but do not upgrade gain 0.054 log units in the year of acquisition and those who upgrade gain a further 0.034 log units, but neither of these effects is statistically significant. Those who upgrade enjoy as stayers a long run effect of 0.007 log units growth to their wages (0.008 log units for upgrading less 0.001 units for acquisition); this cumulates to a substantial amount for whose who remain as stayers for any length of time. However acquisition of qualification has a substantial long-run effect, significant at the 1 per cent level, on employment prospects. Those who upgrade their qualifications experience employment disruption, significant at a 10 per cent level, in the short term but in the long term enjoy further support for their employment prospects, but this long-term addition for women who upgrade is not statistically significant. Acquisition and upgrading of qualifications has little impact on the switching equation.

The time taken for lifelong learning (with or without an upgrade) to have a positive effect on employment chances may reflect the sort of problems that some people have in finding suitable jobs after gaining qualifications, as discussed by Purcell et al. (2007).

With regard to the other coefficients, we see that wage rates of movers rise with their level of qualification. Qualifications at level 4 attract wages markedly higher than any other level (a premium of 0.48 log units). Academic qualifications appear prized relative to vocational qualifications in terms of their associated wage rates. The qualification

	Coeff.	s.e	
Orig Qual 1	0.222	0.056	***
Orig Qual 2	0.246	0.078	***
Orig Qual 3	0.313	0.065	***
Orig Qual 4	0.349	0.054	***
Orig Qual other	0.165	0.040	***
High Qual Academic	-0.073	0.040	*
Age	0.062	0.016	***
Age^2 lagged	-0.093	0.019	***
Not White	-0.081	0.085	
London	-0.003	0.065	
South-West	0.130	0.060	**
East Anglia	-0.226	0.089	**
East Midlands	0.069	0.058	
West Midlands	0.083	0.059	
North-West	-0.045	0.058	
Yorks Humb.	0.033	0.057	
North	0.114	0.062	*
Wales	0.192	0.067	***
Scotland	0.040	0.059	
Δ ln GDP	0.088	2.961	
Wave Gap	0.019	0.016	
Children age 0-1	-0.520	0.077	***
Children age 2-5	-0.109	0.051	**
Children age 6-10	0.044	0.048	
Children age 11-16	0.027	0.046	
Partnered	-0.134	0.034	***
1996	-0.062	0.077	
1997	-0.128	0.080	
1998	-0.049	0.083	
1999	-0.032	0.083	
2000	0.004	0.086	
2001	-0.006	0.078	
2002	-0.051	0.081	
2003	-0.117	0.084	
2004	0.017	0.082	
2005	0.105	0.081	
2006	0.038	0.084	
Employed at Start	0.092	0.035	
Cut 1	2.668	0.363	***
Cut 2	3.540	0.363	***

N	16526
LR chi2(38)	394.96
Prob > chi2	0
Pseudo R2	0.0391
Log likelihood	-4849.66

Note: * Significant at 10% ** Significant at 5% *** Significant at 1%

Table 4: The Decision to undertake Lifelong Learning: Probit Model Parameters

	Mover			Stayer			Switching			Employment		
N=16,526	Coeff	s.e.		Coeff	s.e.		Coeff	s.e.		Coeff	s.e.	
Acquired_t				0.054	0.040					0.115	0.769	
Acquired_{t-1}				0.001	0.008					0.752	0.128	***
Acquired_{t-2+}				-0.001	0.004					0.578	0.124	***
Ever acquired	0.036	0.030					0.006	0.077				
Upgraded_t				0.034	0.024					-0.686	0.400	*
Upgraded_{t-1}				0.013	0.018					-0.126	0.183	
Upgraded_{t-2+}				0.008	0.006					0.110	0.158	
Ever upgraded	0.131	0.042	***				0.086	0.111				
Orig Qual 1	0.025	0.040		-0.004	0.006		0.045	0.117		0.320	0.123	***
Orig Qual 2	0.108	0.046	**	-0.005	0.008		0.056	0.166		0.225	0.172	
Orig Qual 3	0.163	0.055	***	0.007	0.007		0.483	0.157	***	0.164	0.153	
Orig Qual 4	0.483	0.050	***	0.006	0.006		0.589	0.128	***	0.305	0.148	**
High Qual Academic	0.080	0.033	**	0.016	0.004	***	0.280	0.082	***	0.021	0.092	
Age	0.019	0.010	*	-0.006	0.002	***	0.016	0.030		0.131	0.034	***
$\text{Age}^2/100$	-0.025	0.012	**	0.007	0.002	***	0.001	0.035		-0.211	0.044	***
Not White	-0.040	0.072		0.008	0.007		0.005	0.169		-0.370	0.183	**
London	0.100	0.047	**	0.001	0.005		-0.041	0.120		0.034	0.144	
South-West	-0.093	0.041	**	-0.002	0.006		-0.189	0.118		-0.057	0.139	
East Anglia	-0.138	0.062	**	0.005	0.008		-0.295	0.180		-0.010	0.187	
East Midlands	-0.144	0.039	***	-0.005	0.006		-0.370	0.128	***	-0.276	0.143	*
West Midlands	-0.084	0.042	**	0.004	0.006		-0.111	0.130		-0.123	0.130	
North-West	-0.073	0.048		0.007	0.005		0.094	0.126		-0.301	0.142	**
Yorks Humb.	-0.089	0.043	**	0.006	0.006		-0.217	0.117	*	-0.323	0.132	**
North	-0.165	0.048	***	0.005	0.005		0.041	0.138		-0.189	0.146	
Wales	-0.120	0.055	**	0.003	0.007		-0.051	0.148		-0.296	0.184	
Scotland	-0.053	0.056		0.000	0.005		-0.168	0.125		-0.237	0.147	
Employed at Start	0.177	0.044	***	-0.023	0.007	***	0.232	0.120	*	1.672	0.222	***
Δ ln GDP				0.246	0.243		1.027	3.105		-2.524	2.047	
Ln GDP	0.761	0.095	***									
Newly Employed	-0.137	0.056	**				-2.025	0.319	***			
Wave Gap							-0.045	0.019	**	0.024	0.012	**
Recent Job							-0.723	0.056	***			
Partner Lagged										0.152	0.081	*
children aged 0-1										-1.416	0.211	
children aged 2-5										-0.933	0.152	
children aged 6-10										-0.350	0.117	
children aged 11-16										0.050	0.098	
Gen. Residual	0.005	0.016		-0.032	0.021		-0.083	0.045	*	0.296	0.402	
Constant	-2.042	0.489	***	0.139	0.035	***	0.523	0.666		-2.085	0.621	***
mass point 2	-0.704	0.036	***							-0.150	0.268	
P(mass point 2)	0.092	0.017	***									
mass point 3	0.964	0.160	***							2.128	0.529	***
P(mass point 3)	0.010	0.005	**									
mass point 4	0.137	0.081	*							-1.996	0.940	**
P(mass point 4)	0.090	0.142										

Note: * Significant at 10% ** Significant at 5% *** Significant at 1%

Table 5: Model Parameters

	Coeff	s.e.	z	Notes:
$\ln\sigma_1^2$	-1.245	0.040	-30.07	movers
$\ln\sigma_2^2$	-2.047	0.027	-75.92	stayers
$\text{atanh}\rho_{1s}$	0.085	0.177	0.52	movers, switching
$\text{atanh}\rho_{2s}$	0.437	0.102	3.71	stayers, switching
$\text{atanh}\rho_{1p}$	0.162	0.212	0.99	movers, employment
$\text{atanh}\rho_{2p}$	-0.305	0.055	-3.72	stayers, employment
$\text{atanh}\rho_{ps}$	-0.619	0.176	-3.54	switching, employment

Table 6: Equation Standard Errors and the Correlations between them

premia shown here are broadly consistent with other estimates of the returns to education (McIntosh 2006). Being newly employed leads to a pay penalty of 0.14 log units as a mover and the switching equation shows that someone who is newly employed (i.e. has not been working for at least a year) is much more likely to be a mover than someone who is not newly employed. Both of these effects are highly significant. Women are also more likely to be movers if the gap between the adjacent waves of the survey is large and also if they have changed their jobs recently. Women with partners are more likely to be employed and those with children less likely to be employed; the magnitude of this last effect decreases with the age of the children. There is little evidence that selection effects associated with the decision to undertake lifelong learning influence wages directly. However, the significance of the generalised residual term in the switching equation shows the need to control for the endogeneity of the learning decision with regard to mover-stayer status.

On the basis of likelihood ratio tests, we select four mass points; the coefficients associated with these are all significant in either the movers' equation or the employment equation or both. The probabilities of women being associated with mass points 1, 2, 3 and 4 are 80.8%, 9.2%, 1% and 9.0% respectively. Movers' wages are, for those associated with mass points 2 and 3, sharply different from those associated with mass point 1. Employment prospects are sharply different for those associated with mass points three and four.

The parameters of table 6 imply that the variance of the movers' equation is 0.08

while that of the stayers' equation is 0.02. Conditioning on previous wages, the variance of wages for stayers is smaller than that for movers, in line with the ideas underlying the model specification. The statistically significant correlations between the errors of the different equations further indicate the value of the model structure as compared with, say, considering employment and wages independently.

Finally we note the significance of the coefficients on qualification levels, *Newly Employed, Wave Gap* and *Recent Job* in the switching equation. These point to the stability of women's earnings depending on their qualifications and labour market experience.

5 Simulating the Effects of Lifelong Learning

Table 6 shows the effects of lifelong learning in each of the equations of our model. However, it is difficult to discern from this the overall effect of such learning for two reasons. First, the direct effect of learning on wages depends on whether someone is a mover or a stayer. This is determined probabilistically by the parameters of the switching equation. Second, a feature of lifelong learning as shown by this model is that it increases the probability of employment. This effect is much more marked for women who upgrade their educational level than it is for those who gain a qualification without upgrading. Furthermore, as we see from table 5, the coefficient on *Newly Employed* is negative. This means that reducing the risk of a woman not being employed increases her earnings should she be a mover when re-employed (i.e. not re-employed on earnings close to those in her previous job). Subsequently, as a stayer her earnings are also increased because they are set by cumulating changes from her earnings when last a mover. Thus an assessment of the full effects of lifelong learning needs to take account not only of the impact on the probability of employment but also of the feedback that this has on the expected level of earnings; in addition it needs to take account of the probabilistic nature of the switching model. This can be done only by simulation.

In the simulations presented here we consider a white women with a partner who has two children, one at age twenty-eight and one at age thirty. We explore the effects of lifelong learning at age twenty-five and at age forty. Instead of considering someone in any particular region we use the weighted average of the regional dummies with the weight for each region being the share of women within that region averaged over the years 1996 to 2007. We similarly assume the average population figures for those whose highest qualification is academic. lnGDP is assumed to grow from its mean value for the period 1996-2007 at the real rate of 0.02 units per annum. We simulate the earnings and employment history of someone between the ages of twenty-three and sixty, measuring the change in the present discounted value of current and future labour income from the time of qualification acquisition to age sixty, using a discount rate of five per cent per annum.

Our simulations are stochastic at two levels. We draw model parameters from a multivariate normal distribution with mean given by the parameters shown in tables 5 and covariance matrix given by the associated covariance matrix of these parameters. We carry out five hundred of these draws for each of the cases considered. Each parameter set has its own covariance matrix for the disturbances associated with the four individual equations of table 5. We then consider a panel of ten thousand women ageing from twenty-three to sixty. From the relevant covariance matrix we draw ten thousand sets of disturbances for each equation and for each year between ages twenty-three and sixty. We use the movers' equation with the age parameters set to age twenty-three to give an initial distribution for earnings and the employment equation together with its set of disturbances to establish which of our ten thousand women are employed.

For each subsequent year, the switching equation with its set of disturbances is used to determine which women are movers rather than stayers. The relevant disturbances determine earnings in each case and the employment equation again determines which of the panel are employed. The exercise is repeated five times using the same set of disturbances, so as to provide earnings and employment paths as functions of the five different qualification levels. For each of these five qualification levels four further sets of simulations are carried out; the first two assume that lifelong learning takes place at age twenty-five without or with upgrading. The third and fourth simulations assume that the lifelong learning takes place at age forty. We then calculate the percentage increase in the present discounted value of earnings with each type of lifelong learning at ages twenty-five and forty relative to that at the same ages without lifelong learning.

In table 7 we show these effects of lifelong learning calculated as the means of the five hundred simulations. The standard deviations of the simulations measured in percentage points and computed across the five hundred parameter sets are also shown. Finally we show the proportions of the five hundred simulations which indicate that lifelong learning of each type has a negative effect on present discounted earnings. This provides a set of P-values for the hypothesis that lifelong learning has a negative effect on discounted future earnings. The results are shown both taking account of the probability of employment, in the first three columns and showing the wage terms on the assumption that the women are employed in each year. Thus the differences between the two reflects the direct effect of lifelong learning on the probability of employment.

Table 7 shows that acquisition of qualifications has effects on wages statistically significant at a 5% level in all cases. The overall returns which take into account the effects of qualifications on employment are appreciably above the pure wage effects. This reflects the fact that, compared to men, women often have a more marginal attachment to the labour market such that relatively small changes can have a substantial influence on employment probability. The large impact of acquiring qualifications without upgrading on women initially at level 0 can be put in its context by noting in table 3 that the non-employment rate of women with level 0 qualifications and no lifelong learning is 54 per cent. With lifelong learning but no upgrading this falls to 21 per cent. Thus the employment rate rises from 45 per cent to 79 per cent. If there were no change to wage rates but the whole of this increased employment were due to the acquisition of qualifications, the return shown in table 7 would be 75 per cent. In fact the employment effect, approximated as the total effect less the wage effect, is 41 per cent. Thus our articulated model with its accommodation of individual-specific effects results in returns

	Full model			Wages only		
	mean	s.d	P<0	mean	s.d	P<0
			Not upgrading			
			Lifelong learning age 25			
Orig Qual 0	47%	0.12pp	0%	6%	0.03pp	2.4%
Orig Qual 1	28%	0.07pp	0%	8%	0.04pp	0.9%
Orig Qual 2	27%	0.07pp	0%	7%	0.03pp	1.0%
Orig Qual 3	30%	0.09pp	0%	10%	0.05pp	0.6%
Orig Qual 4	28%	0.09pp	0%	11%	0.05pp	0.5%
			Lifelong learning age 40			
Orig Qual 0	44%	0.12pp	0.0%	7%	0.03pp	1.4%
Orig Qual 1	23%	0.07pp	0.1%	8%	0.03pp	0.4%
Orig Qual 2	22%	0.08pp	0.1%	7%	0.03pp	0.7%
Orig Qual 3	24%	0.08pp	0.1%	10%	0.04pp	0.6%
Orig Qual 4	21%	0.08pp	0.1%	10%	0.05pp	0.9%
			Upgrading			
			Lifelong learning age 25			
Orig Qual 0	74%	0.17pp	0%	22%	0.06pp	0.0%
Orig Qual 1	52%	0.12pp	0%	26%	0.07pp	0.0%
Orig Qual 2	50%	0.11pp	0%	25%	0.06pp	0.0%
Orig Qual 3	58%	0.15pp	0%	32%	0.10pp	0.0%
			Lifelong learning age 40			
Orig Qual 0	68%	0.16pp	0%	22%	0.05pp	0.0%
Orig Qual 1	41%	0.10pp	0%	24%	0.06pp	0.0%
Orig Qual 2	40%	0.10pp	0%	23%	0.05pp	0.0%
Orig Qual 3	41%	0.12pp	0%	25%	0.07pp	0.0%

Table 7: Returns to Lifelong Learning

which, while high, are much lower than a crude analysis might indicate.

6 Discussion

An obvious question is how the switching regression adopted here compares with other approaches used to explore earnings dynamics. While retaining the pooled structure used here, the model allows us to compare our results directly with models which rely on treating everyone as either movers or as stayers. If everyone were regarded as a mover – as they implicitly are in cross-sectional analyses that have no dynamic component – we would find a large negative constant in the switching equation, ensuring that all observations were classified as coming from movers, and all the other terms in that equation would be statistically insignificant. We can confidently reject the hypothesis that all terms in the switching equation other than the constant are not significant (p-value of 0.000). This test also allows us to reject the idea that everyone is a stayer – the implicit assumption in the first-difference model. If the mover equation explained the earnings of people who had not previously been employed, and the stayers' equation explained everyone else's earnings, then the first-difference model would be valid. Such a situation would be generated by a large negative coefficient on *Newly Employed* in the switching equation, by statistical insignificance of all other variables and by a positive constant which is large enough to ensure that the probability of being a mover is negligible unless someone is newly employed. All other variables would be statistically insignificant. This hypothesis is also rejected by the test result presented above. Thus, our model rejects as incomplete descriptions of the data two popular alternative models used to explore earnings. The statistical significance of the correlations between the disturbances of the four equations in our system suggests that it is not appropriate to ignore employment selection effects. Overall, the approach used in this paper offers an important generalisation of more conventional models.

Such methodological differences need to be borne in mind in comparing our results with other studies. However, as noted in the introduction, econometric analyses of lifelong learning are few in number. Blanden et al. (2010) used a model in the log level of wages and a structure very similar to our movers' equation (2). However, they addressed the problem of individual fixed effects by introducing individual-specific constants which were invariant over the sample period. They did not correct for the possible endogeneity of either the employment decision or the decision to undertake lifelong learning. In a regression where they did not distinguish between acquisition and upgrading, but looked simply at whether lifelong learning was undertaken, they found an eventual impact of 0.221 log units, comparable to our estimate of 0.201 log units for women who upgrade their qualifications. Instead of looking at the effects of aggregated upgrading Blanden et al. (2010) also look at the earnings of people classified by their qualification level after lifelong learning, independently of the qualification level before lifelong learning. They find very high returns with 30 per cent or higher for level 3 (whether academic or vocational) and 27.8 per cent for level 4 academic qualifications, but note, not surprisingly, that there are problems arising from small sample sizes and the former estimate, in particular, is poorly determined. On the other hand de Coulon & Vignoles (2008) estimate a model in first differences, implicitly assuming everyone to be a stayer. Their results suggest that lifelong learning undertaken between the ages of 26 and 34 significantly increases British women's wage growth by roughly 20 per cent. They also show a significant effect on employment but do not give an interpretation of this in terms of its impact on employment rates.

The results presented in this paper are in line with these findings. Our results suggest that lifelong learning that results in an upgrade to the previously-held highest level of qualification increases women's wages by about 20 per cent. Where there is no such upgrade, there is little effect. Analyses that do not distinguish between these two types of lifelong learning cannot detect this important difference. To allow a rough comparison, we note from Table 2 that about one quarter of lifelong learning in our sample results in a qualification upgrade. Hence, an appropriate estimate from our model of the average effect of lifelong learning – that is, an estimate that takes no account of whether qualifications are upgraded – would be in the region of a 5 per cent increase.

7 Conclusions

In this paper we have investigated the effect of lifelong learning on women's earnings using data from the British Household Panel Survey. We have developed a modelling framework intended to represent an intuitive feature of the labour market. People are likely to continue in their jobs for some time and while this happens their hourly earnings are not likely to change very much from one year to the next. We have denoted such people as stayers. But there will also be occasions when they experience substantial movement to their earnings. We have represented this as moving to an earning level which need not be closely related to previous earnings; we denote such people as movers. Movers might, of course, have to move as a result of job-loss. But they may also be people who have not been able to realise their potential in their previous job and who find substantially better opportunities. We find this to be a realistic representation of the way in which hourly earnings change over time. We show it is superior to a framework in which wages are modelled simply by analysis of growth rates from one period to the next.

However, for the model to be fully satisfactory, it is necessary to take account of possible interdependence between employment decisions and opportunities and earnings movements. We take this into account. It is also necessary to take account of the possibility that people may differ in their individual characteristics – both observed and unobserved – and that this may affect their earnings and employment. This is done using a statistical approach which identifies a number of 'clusters' and allocates people to one or other of these on the basis of statistical fit.

The mover-stayer framework could be used to explore earnings movements in the absence of lifelong learning. However, given the structure, it is straightforward to introduce variables to represent the effects of lifelong learning and these allow us to establish its influence on women's earnings and employment. We take due account of the fact that the decision to undertake lifelong learning may be endogenous in a way which could influence our estimates of the mover-stayer model.

Our results offer a positive view of the effects of lifelong learning on women's hourly earnings. We find that qualifications which result in women's educational status being upgraded have a clear effect on their earnings. Lifelong learning appears to provide a one-off boost to wages growth for those in stable employment. It also influences the probability of being in work and thereby indirectly increases earnings for movers. These results are robust to controlling so as to distinguish the effects of lifelong learning from the characteristics of people who, at some point, undertake lifelong learning.

As noted in the Introduction, lifelong learning is widespread in a number of countries and it is common for government policy to encourage it. In the UK, explicit targets for skills development were set out in an official review of skills needs (Leitch Report 2006). Presented as a means of increasing productivity, growth and social justice, the recommendations are for skills upgrading at all levels and for continued progression for those in the highest skills group. The results of our analysis speak to the importance of acknowledging the distinction between simply acquiring a new qualification and acquiring a qualification that results in a demonstrable and visible skills upgrade with the latter being considerably more valuable, especially after employment effects are taken into account.

References

Badescu, M. & Saisana, M. (2008), Participation in Lifelong Learning in Europe. What can be Measured and Compared? Joint Research Centre, Ispra.

Blanden, J., Buscha, F., Sturgis, P. & Urwin, P. (2010), Measuring the Returns to Lifelong Learning. Centre for the Economics of Education Discussion Paper No 110.

Browning, M., Ejrnaes, M. & Alvarez, J. (2010), 'Modelling Income Processes with Lots of Heterogeneity', *Review of Economic Studies* **77**, 1353–1381.

de Coulon, A. & Vignoles, A. (2008), An Analysis of the Benefit of NVQ2 Qualifications acquired at Age 24-26. Centre for the Economics of Education Discussion Paper No 106.

Dorsett, R., Lui, S. & Weale, M. (2010), Modelling Earnings and the Effect of Lifelong Learning. NIESR Discussion Paper No 352. http://www.niesr.ac.uk/pdf/300310_103442.pdf.

Dutta, J., Sefton, J. & Weale, M. (2001), 'Income Distribution and Income Dynamics in the United Kingdom', *Journal of Applied Econometrics* **16**, 599–616.

Egerton, M. & Parry, G. (2001), 'Lifelong Debt: Rates of Return to Lifelong Learning', *Higher Education Quarterly* **55**, 4–27.

Erikson, R. & Goldthorpe, J. (2002), 'Intergenertional Inequality: a Sociological Perspective', *Journal of Economic Perspectives* **16**, 31–44.

Ferrer, A. & Menendez, A. (2009), The Returns to Flexible Post-secondary Education: the Effect of Delaying School. http://ideas.repec.org/p/ubc/clssrn/clsrn_admin-2009-26.html.

Goodman, L. (1961), 'Statistical Methods for the Mover-stayer Model', *Journal of the American Statistical Association* **56**, 841–868.

Hällsten, M. (2010), 'Late Entry in Swedish Tertiary Education: Can the Opportunity fo Lifelong Learning Promote Equality over the Life Course', *British Journal of Industrial Relations* . Forthcoming.

Heckman, J. & Singer, J. (1984), 'A Method for minimising the Impact of Distributional Assumptions in Econometric Models for Duration Data', *Econometrica* **52**, 271–320.

Higher Education Statistics Agency (1995), *Students in Higher Education Institutions, 1994-5.*

Higher Education Statistics Agency (2008), *Students in Higher Education Institutions, 2007-8.*

Holmlund, B., Liu, Q. & Skans, O. (2008), 'Mind the Gap? Estimating the Effects of Postponing Higher Education', *Oxford Economic Papers* **60**, 683–710.

Jenkins, A., Vignoles, A., Wolf, A. & Galindo-Rueda, F. (2002), 'The Determinants and Effects of Lifelong Learning', *Applied Economics* **35**, 1711–1721.

Kim, C. (2004), 'Markov-switching models with endogenous explanatory variables', *Journal of Econometrics* **122**, 127–136.

Leitch Report (2006), Prosperity for All in the Global Economy: World Class Skills. http://www.hm-treasury.gov.uk/d/leitch_finalreport051206.pdf.

Light, A. (1995), 'The Effect of Interrupted Schooling on Wages', *Journal of Human Resources* **30**, 472–502.

Maddala, G. (1983), *Limited-Dependent and Qualitative Variables*, Cambridge University Press, New York.

McIntosh, S. (2006), 'Further analysis of the returns to academic and vocational qualifications', *Oxford Bulletin of Economics and Statistics* **68**, 225–251.

Meghir, C. & Pistaferri, L. (2004), 'Income Variance Dynamics and Heterogeneity', *Econometrica* **72**, 1–32.

Orme, C. (2001), Two-step Inference in Dynamic Non-linear Panel Data Models. http://personalpages.manchester.ac.uk/staff/chris.orme.

Purcell, K., Wilton, N. & Elias, P. (2007), 'Hard Lessons for Lifelong Learners? Age and Experience in the Graduate Labour Market', *Higher Education Quarterly* **61**, 57–82.

Quandt, R. (1958), 'The Estimation of Parameters of a Linear Regression obeying Two Different Regimes', *Journal of the American Statistical Association* **53**, 873–880.

Taniguchi, H. (2005), 'The Influence of Age at Degree Completion on College Wage Premiums', *Research in Higher Education* **46**, 861–881.

Ulrick, S. (2008), 'Using Semi-parametric Methods in an Analysis of Earnings Mobility', *Econometrics Journal* **11**, 478–498.

UNESCO Institute for Lifelong Learning (2009), *Global Report on Adult Learning and Education, 2009.* http://www.unesco.org/uil/en/UILPDF/nesico/GRALE/gralen_en.pdf.

Vella, F. & Verbeek, M. (1999), 'Two-step estimation of panel data models with censored endogenous variables and selection bias', *Journal of Econometrics* **90**, 239–263.

A Appendix: Model testing results

The estimated model includes four mass points in the movers' equation. This was selected on the basis of likelihood ratio tests. Table 8 shows the log likelihood obtaining as the number of mass points K is increased from 1 (no unobserved heterogeneity) to 5. The final column shows the likelihood ratio test statistic associated with $K-1$ rather than K mass points. The 95% critical value of 7.81 (with 3 degrees of freedom) is not achieved beyond $K=4$ and we therefore choose 4 rather than 5 mass points.

K	Log-likelihood	$\chi^2(3)$
1	-4316.412	
2	-4311.004	48.5
3	-4276.439	65.1
4	-4274.482	18.1
5	-4272.128	6.5

Table 8: Choosing the number of mass points in the movers' and employment equations

For more information, please contact
llakescentre@ioe.ac.uk
LLAKES Centre
Institute of Education
20 Bedford Way
WC1H 0AL
London
UK

National Institute of Economic and Social Research

University of BRISTOL

UNIVERSITY OF Southampton School of Education

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL