National Institute of Economic and Social Research
Discussion Paper No 352

# Economic Benefits of Lifelong Learning*

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

29th March 2010

**Abstract**

This paper examines the effect of lifelong learning on men's employment and wages. 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). Mover-stayer status is not observed and we therefore model wages using an endogenous switching regression, extended to take account of non-random selection into employment. The model is estimated by maximum likelihood, using generalised residuals to correct for possible endogeneity of lifelong learning decisions. The results show modest effects significant at a 10% level for men who undertake life-long learning without upgrading their educational status and more powerful and significant effects for those who do upgrade their status. For the latter, the influence of lifelong learning on employment prospects is an important influence on the overall return.

**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.

# 1 Introduction

In a number of advanced economies it has become increasingly common for people to undertake lifelong learning, that is a period of study after the completion of formal education. For example, Holmlund et al. (2008) report that in 2002 just over forty per cent of Swedish university entrants had completed secondary school more than five years earlier, while only about one third progressed to university within one year of completing secondary school. Similarly, 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 et al. 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). Using a much broader definition of lifelong learning the UNESCO Institute for Lifelong Learning (2009) indicates that in the United Kingdom over fifty per cent of adults aged twenty-six to forty-five report recent participation in some form of adult learning or education, with a participation rate of forty-one per cent for people aged forty-six to fifty-five and twenty-one per cent for people aged fifty-six to sixty five. On this basis participation rates in the United States are slightly lower for those forty-five and under and slightly higher for those aged over forty-five.

In many countries, government policy has been to encourage lifelong learning as a means of increasing productivity and achieving progression in the labour market. However, a number of studies suggest that 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 et al. (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 et al. (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 et al. (2008) finds no benefit to lifelong learning in aggregate for men although some evidence of benefit when looking at some sub-categories.

In this paper we add to the empirical literature by presenting new evidence on the effect of lifelong learning on employment outcomes in Britain. We base our analysis on a nationally representative longitudinal survey dataset spanning the period from 1991-2008 and consider the effects of lifelong learning in an extension of the classic mover-stayer model (Goodman 1961). Dutta et al. (2001) showed that such a structure offered a better means of understanding income inequality in the UK than did other popular specifications. The mover-stayer model is sufficiently flexible to allow for the possibility that people receive a wage that is either a random draw from a stationary distribution or that is closely related to their wage rate of the previous year. This dual-regime framework defines our two groups. 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. Intuitively, the stayers in this model are characterised by employment stability while the movers are likely to have experienced some disruption to their employment. However, it is not necessarily the case that being a stayer is preferable to being a mover since it is possible that some stayers are in jobs that offer little in the way of progression and that, for them, becoming a mover might allow them to improve their prospects.

The paper makes contributions on both the substantive and the methodological fronts. 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. Specifically, 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. Methodologically, 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 distinguishes our work from most other studies of the earnings mobility, that 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. (2008) , Meghir & Pistaferri (2004) and Ulrick (2008)). However, our approach does nest the more traditional analysis in which earnings are simply differenced to remove individual-specific fixed effects. We also address the potential endogeneity of lifelong learning.

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, beginning with a traditional model specified in first differences so as to remove individual fixed effects, and then moving on to the specification of our more general mover-stayer model. Section 4 presents the implied returns to lifelong learning and in section 5 we discuss the relationship between our findings and other related work. Section 6 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 and 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 all seventeen waves from 1991 to 2007. Members of these households are repeatedly surveyed regardless of changes to household membership. In common with most analysis of wages (see, for example, Ramos (2003), Dickens (2000), Cappellari & Jenkins (2008), Ulrick (2008), Meghir & Pistaferri (2004) and Lillard & Willis (1978) ), we consider only men. We limit ourselves to men aged 25 to 60 in order to concentrate on working lives beyond completion of the conventional period of education. Thus, for those younger then twenty-five in 1991 or older than sixty 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. Table 1 summarises the populations we study.

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

Table 1: Number of men in each survey wave

| wave | Total in each wave used for analysis |
|---|---|
| 6 | 1466 |
| 7 | 1431 |
| 8 | 1355 |
| 9 | 1273 |
| 10 | 1217 |
| 11 | 1114 |
| 12 | 1016 |
| 13 | 923 |
| 14 | 859 |
| 15 | 799 |
| 16 | 745 |
| 17 | 698 |
| total | 12896 |

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. We look at the effects of lifelong learning 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. (2008)) we merge categories 4 and 5. Our classification of qualifications is shown in table

2.

Table 3 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 upgrading to be roughly 5 per cent, with little variation by qualification level (note that those with level 4 qualifications cannot upgrade, by definition). A very different impression is formed when considering the incidence of lifelong learning, regardless of whether this resulted in a qualification upgrade (last row of Table 3 ). Here there is a clear gradient. Among those with no qualifications, about 10 per cent will undertake some learning. This is substantially higher for those with a level 1 qualification (19 per cent) and higher still for those with level 2 qualifications. For those with level 3 or level 4 qualifications, the participation rate is three times that for those with no qualifications.

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.

## 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[1]. 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 4 provides a summary of average hourly wages and non-employment rates for the men 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

[1] This is a derived variable wPAYGU.

| 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 2: The Classification of Qualifications

| | | Initial qualification level | | | | | |
|---|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 4 | All |
| Qualification | 0 | 94.10% | 0% | 0% | 0% | 0.00 | 21.02% |
| Level | 1 | 2.82% | 93.44% | 0% | 0% | 0% | 31.62% |
| Five | 2 | 1.28% | 1.90% | 94.56% | 0% | 0% | 8.88% |
| Years | 3 | 1.54% | 1.55% | 1.36% | 92.47% | 0% | 15.75% |
| Later | 4 | 0.26% | 3.11% | 4.08% | 7.53% | 100% | 22.74% |
| Upgrading | | 5.90% | 6.56% | 5.44% | 7.53% | 0% | 5.15% |
| Lifelong learning | | 10.26% | 18.83% | 27.89% | 30.47% | 31.62% | 22.11% |
| N | | 390 | 579 | 147 | 279 | 351 | 1,746 |

Table 3: Transition Probabilities over a Five-year Window and the Incidence of Lifelong Learning

probabilities also increase with qualification level when considering those who do not undertake lifelong learning but no such smooth pattern is evident among those with experience of lifelong learning. 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, particularly for those with qualifications at level 2 or lower. Above that level, the increases associated with lifelong learning appear more marginal, at least relative to the average wage that prevails where no lifelong learning is undertaken. Where qualifications are upgraded as a result of lifelong learning, the apparent premium is larger still. This is particularly the case for those initially with level 2 qualifications. However, it remains true that it is among those with qualifications at or below level 2 that the strongest effects are seen. It is also among this group that lifelong learning without qualification upgrade appears to have the highest impact on the probability of being employed; this is particularly striking for those with no qualifications. Lifelong learning that involves an upgrade to qualifications also improves employment probabilities, except among those initially with a level 3 qualification. Interestingly, though, these increases are not as great as those associated with lifelong learning that does not involve a qualification upgrade.

| Initial Education Level | No lifelong Learning | With Qualification but not Upgrading | With Upgrading | Total |
|---|---|---|---|---|
| 0 | 2034 | 277 | 273 | 2584 |
| 1 | 3024 | 812 | 434 | 4270 |
| 2 | 756 | 356 | 114 | 1226 |
| 3 | 1329 | 714 | 163 | 2206 |
| 4 | 1555 | 1047 | 0 | 2602 |
| Total | 8698 | 3206 | 984 | 12888 |
| | Average Hourly Earnings (2005 prices) | | | |
| 0 | £7.98 | £9.40 | £9.99 | |
| 1 | £9.84 | £10.50 | £10.65 | |
| 2 | £10.01 | £10.69 | £13.12 | |
| 3 | £12.28 | £12.35 | £12.51 | |
| 4 | £15.76 | £16.09 | | |
| | Non-employment Rates | | | |
| 0 | 38.3% | 11.2% | 16.5% | |
| 1 | 18.3% | 8.4% | 13.4% | |
| 2 | 13.5% | 10.1% | 9.6% | |
| 3 | 9.3% | 10.4% | 17.2% | |
| 4 | 9.0% | 7.2% | | |

Table 4: Summary Data: Intial Qualifications, Earnings, Employment and Lifelong Learning, 1996-2008 Average. Pooled Data

While intriguing, such descriptive statistics can only offer a partial insight into the effect of lifelong learning. To proceed, we need to use econometric methods.

# 3 Econometric analysis

In this section, we discuss in more detail the mover-stayer model, describe the econometric approach and present estimation results. As a preliminary step, and partly to motivate our approach, we first present the results of estimating a model in first differences since this is a more standard and familiar approach. With both the model in first differences and the mover-stayer model, the question of the appropriate variables to include in the model arises and we turn to this issue first of all.

## 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. We also want to have some sense of whether the effects of lifelong learning are age-dependent as Blanden et al. (2008) suggests. However, our ability to explore the effects of large number of dummy variables is limited by the relatively small number of people who upgrade their educational status.

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 $t-i$ $(5 \geq i \geq 0)$ 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. These dummy variables allow us to identify the impact effect of lifelong learning in each of the last five years.

We also set up dummies which indicate whether people have acquired qualifications since the first wave of the BHPS in 1991 and also if they have acquired qualifications since 1991 with upgrading. We explore the effects of these in three age ranges, 25-34, 35-49 and 50-60; these are indicated by the dummy variables *Ever Acquired 25-34*$_{t-i}$, *Ever Upgraded 25-34*$_{t-i}$ and so on. *Ever Acquired 25-34*$_{t-i}$ , for example, takes a value of 1 if someone aged 25-34 acquired a qualification between 1991 and year $t-i$, whether they upgraded their educational status or not

We do not distinguish multiple qualifications from single qualifications in the *Ever Acquired* and *Ever Upgraded* terms. However, if someone acquires a qualification without upgrading this will be reflected in the relevant *Ever Acquired* term. If they subsequently upgrade, then the relevant *Ever Upgraded* term will also take a value of one. However if someone acquires a qualification in one year and another one two years later, then more than one $Acquired_{t-i}$ and, if relevant $Upgraded_{t-i}$ dummies may take values of one.

We include additional variables in the analysis to control for other sources of variation within our sample. These include: qualification level when first observed; a dummy variable indicating whether the highest qualification at that time was academic or not; age; whether from an ethnic minority group or not; marital status (single or partnered), the presence of children (represented by a 0/1 dummy variable); 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; GDP or its change as an indicator of the state of the economy; and the time between interviews. Some of these variables were included as instrumental variables to assist with identification of the model, as described later.

## 3.2 A Model in First Differences

Analysis of the effects of training and qualification gain is usually carried out by exploring whether the pay of those who gain qualifications rises faster than those who do not gain such qualifications. This approach is of necessity employed when looking at cohort studies such as the British Cohort Survey; since they collect data only at long intervals it is not practical to explore whether the underlying dynamic processes are more complicated than this and, if so, what the implications of that might be. An attraction of working with first differences is that this removes any individual-specific effects which might explain the levels of people's earnings; such effects might bias the results of a study of the levels of earnings if they are correlated with the propensity to gain qualifications.

The equation we estimate in log hourly pay is, with $X_{it}$ explanatory variables and $\beta$ a vector of coefficients,

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

However this is estimated only for those men who were employed in adjacent years. There is a risk that this generates selection bias; we explore this using Heckman's standard procedure. We set up an employment indicator $H_{it}$ which takes a value 1 if someone is employed and providing hourly wage data in periods $t$ and $t-1$ but a value 0 otherwise. This is driven by the latent variable $H_{it}^*$ with $H_{it} = 1$ if $H_{it}^* > 0$ and $H_{it} = 0$ otherwise. $H_{it}^*$

$$H_{it}^* = W_{it}\lambda + \eta_{it} \tag{3}$$

where $W_{it}$ is a vector of explanatory variables and $\lambda$ a coefficient matrix. The distribution of the residual terms has the structure

$$\begin{pmatrix} u_{it} \\ \eta_{it} \end{pmatrix} \tilde{}N\left(0, \begin{bmatrix} \sigma^2 & \rho \\ \rho & 1 \end{bmatrix}\right)$$

In fact we find that there is little evidence of selection effects. We therefore present the results of our OLS regression for men in table 5 first in unrestricted form and secondly after restricting the terms in acquisition of lifelong learning without upgrading, and also all the dynamic terms in upgrading except for the most significant to zero, in an attempt to improve the resolution of the model. The results are not very encouraging. The only statistically significant coefficient in the unrestricted results suggests that men aged 50-60 with lifelong learning qualifications fare less well in terms of pay growth than do those without. The set of restrictions applied to the unrestricted form are nevertheless easily accepted *(F(12,1695)=0.74, p=0.71)*. But even after imposing the restrictions, the only remotely significant inference which can be drawn is that educational ugprading leads to a rise in earnings after a lag of one year. However, this conclusion has a p-value of 0.08 which is hardly very convincing. Thus these findings are broadly in line with those presented by Blanden et al. (2008) that, in the British Household Panel Survey we cannot identify clear effects of lifelong learning on hourly earnings. The p-values for the absence of selection effects in the analysis for men were 0.84 with the unrestricted model and 0.63 with the restricted model. We note, however, that the absence of selection effects is entirely dependent on the inclusion of the term *Employed at start* as a control variable. Without this term, the relevant p-values fall to zero to four decimal places.

Finally we note the great volatility of hourly earnings. The standard error of the restricted regression for men is 0.245 indicating a substantial degree of unexplained movement from one year to the next.

### 3.3 A Mover-stayer Framework

One possible explanation of the above results is that the structure of the models does not fully address the heterogeneity in people's experiences and that therefore it is inherently mis-specified. The underlying assumption that a first-difference process fully describes the evolution of people's wage rates may simply be incorrect. Separately, the structure above does not allow us to explore the effects of lifelong learning on employment. While we included learning terms in the selection equation, this identifies only people who are employed in adjacent surveys and a focus on the role of learning terms in the selection equation cannot therefore be used as a good indication of their importance as determinants of employment; we therefore cannot explore whether lifelong learning might lead to an important financial return by enhancing employment prospects. We now develop a richer structure which allows us to study employment and earnings effects jointly and also allows us to explore a possible role played by individual effects associated with the people who undertake lifelong learning.

We explore this issue using the mover-stayer framework as applied to the wage distribution. The idea is that movers can be distinguished from stayers. The former receive a wage rate possibly very different from what they had previously earned – they move about the wage distribution. The latter, by contrast, stay at much the same point in the wage distribution as they had been in the previous year; thus their wages are closely explained by the previous year's wage rate.

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

| S.E.=0.245 N=12888 | Coeff. | Std. Err | t-stat | Coeff. | Std. Err | t-stat |
|---|---|---|---|---|---|---|
| | Unrestricted | | | Restricted | | |
| $\text{Acquired}_t$ | 0.012 | 0.015 | 0.780 | | | |
| $\text{Acquired}_{t-1}$ | 0.010 | 0.017 | 0.590 | | | |
| $\text{Acquired}_{t-2}$ | 0.006 | 0.013 | 0.420 | | | |
| $\text{Acquired}_{t-3}$ | -0.008 | 0.011 | -0.730 | | | |
| $\text{Acquired}_{t-4}$ | 0.012 | 0.014 | 0.830 | | | |
| $\text{Upgraded}_t$ | 0.024 | 0.045 | 0.530 | | | |
| $\text{Upgraded}_{t-1}$ | 0.038 | 0.027 | 1.430 | 0.046 | 0.026 | 1.750 |
| $\text{Upgraded}_{t-2}$ | -0.010 | 0.023 | -0.420 | | | |
| $\text{Upgraded}_{t-3}$ | -0.017 | 0.027 | -0.650 | | | |
| $\text{Upgraded}_{t-4}$ | -0.021 | 0.036 | -0.580 | | | |
| Ever Acquired 25-34 | -0.015 | 0.010 | -1.420 | | | |
| Ever Acquired 35-49 | -0.002 | 0.005 | -0.380 | | | |
| Ever Acquired 50-60 | -0.017 | 0.008 | -2.050 | | | |
| Ever Upgraded 25-34 | 0.027 | 0.023 | 1.170 | 0.015 | 0.023 | 0.630 |
| Ever Upgraded 35-49 | 0.003 | 0.013 | 0.250 | 0.002 | 0.012 | 0.130 |
| Ever Upgraded 50-60 | 0.007 | 0.015 | 0.450 | -0.009 | 0.012 | -0.720 |
| Orig Qual 1 | -0.002 | 0.007 | -0.320 | -0.002 | 0.007 | -0.330 |
| Orig Qual 2 | -0.003 | 0.008 | -0.370 | -0.003 | 0.008 | -0.450 |
| Orig Qual 3 | -0.003 | 0.007 | -0.490 | -0.004 | 0.006 | -0.640 |
| Orig Qual 4 | 0.012 | 0.008 | 1.560 | 0.010 | 0.007 | 1.400 |
| Orig Qual other | 0.005 | 0.006 | 0.930 | 0.005 | 0.006 | 0.930 |
| High Qual Academic | 0.002 | 0.005 | 0.430 | 0.002 | 0.005 | 0.440 |
| Age lagged | -0.021 | 0.003 | -7.550 | -0.020 | 0.002 | -8.040 |
| $\text{Age}^2$ lagged | 0.022 | 0.003 | 6.910 | 0.021 | 0.003 | 7.320 |
| Not White | 0.025 | 0.018 | 1.330 | 0.025 | 0.018 | 1.340 |
| London | 0.009 | 0.010 | 0.890 | 0.009 | 0.010 | 0.930 |
| South-West | 0.003 | 0.008 | 0.360 | 0.003 | 0.008 | 0.390 |
| East Anglia | 0.009 | 0.011 | 0.880 | 0.009 | 0.010 | 0.860 |
| East Midlands | 0.005 | 0.008 | 0.640 | 0.005 | 0.008 | 0.720 |
| West Midlands | 0.004 | 0.007 | 0.560 | 0.004 | 0.007 | 0.510 |
| North-West | 0.001 | 0.007 | 0.130 | 0.001 | 0.007 | 0.140 |
| Yorks | 0.002 | 0.008 | 0.250 | 0.002 | 0.008 | 0.240 |
| North | -0.006 | 0.008 | -0.720 | -0.006 | 0.008 | -0.690 |
| Wales | 0.004 | 0.010 | 0.420 | 0.003 | 0.010 | 0.310 |
| Scotland | -0.001 | 0.008 | -0.180 | -0.001 | 0.008 | -0.170 |
| Employed at start | -0.088 | 0.023 | -3.880 | -0.086 | 0.022 | -3.840 |
| $\Delta$ ln GDP | 0.310 | 0.332 | 0.930 | 0.337 | 0.331 | 1.020 |
| Constant | 0.567 | 0.064 | 8.830 | 0.550 | 0.059 | 9.250 |

Table 5: OLS Results for Men: First Differences

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 of the form of equation (2), in the first difference of log earnings.

There is no observed characteristic which makes possible a precise distinction between movers and stayers. Rather we assume that the process is driven by a latent variable; it is thus determined statistically, in much the same way as 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.

However, it is obviously impossible for someone who was previously recorded as not employed to be a stayer. His 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 it has this property.

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 equation 2 as a component might suggest that it encompasses the model in first differences. This is not completely correct, because, as mentioned above, in the model in differences the selection equation requires someone to be employed in two successive periods while in the full model we simply require someone to be employed in the current period. The model, does of course, encompass the first differences model of section 3.2 if 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 and iii) all earnings of people employed in the previous period can be explained by the stayers' equation. We test the restrictions this implies in Appendix A.

We noted above that a virtue of the first difference model was that it removed individual fixed effects associated with the level of earnings and the risk that, if these are correlated with the acquisition of qualifications, they may lead to biases. This issue re-emerges since our movers' equation is set out in the level of log wages. People who study for lifelong qualifications may have higher earnings capacity than those who do not study. We address the problem by means of an endogeneity adjustment computed on the basis of an ordered probit equation; this gives us generalised residuals which can be used to test for endogeneity effects.

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

### 3.4 Movers

For movers, wages are given by a stationary Mincerian equation

$$y_{it} = X_{it}\beta_1 + u_{1it} \tag{4}$$

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 current and lagged real GDP *per capita*. Thus, for a mover, the wage rate is not directly related to previous wages except in so far as the variables which influence the wage of a mover have also influenced their wage on the previous occasion when they were a mover. We explain how this feature is imposed in our model in section 3.10 and how we ensure that all previously non-employed men are treated as movers.

### 3.5 Stayers

The hourly earnings of stayers are assumed to be related to those of the previous period. We use equation 2 of section 3.2 and specify the stayers' wage equation as

$$\Delta y_{it} = X_{it}\beta_2 + u_{2it} \tag{5}$$

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 $\beta_1$ and $\beta_2$.

## 3.6 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 (5) rather than (4) is driven by the latent variable

$$I_{it}^* = Z_{it}\gamma + \varepsilon_{it} \tag{6}$$

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

## 3.7 Selection into Employment

The nature of the model complicates the selection process. Someone cannot be observed to be a mover unless employed in both the current and the previous period. On the other hand, the wage for a mover can be observed conditional only on working in the current period. The specification of the model needs to reflect this; it would not be very satisfactory to estimate wages only for those employed in both current and previous periods, but have nothing to say about those employed only in the current period.

We address this issue 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

$$J_{it}^* = W_{it}\delta + \eta_{it} \tag{7}$$

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. This replaces equation (3) of section 3.2 which looked at people being employed in two successive periods.

## 3.8 Lifelong Learning

Our analysis needs to take account of the consequences of the 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

$$K_{it}^* = V_{it}\zeta + \nu_{it} \tag{8}$$

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 $K_{it}^* < 0$

## 3.9 Estimation Strategy

The model has the following likelihood function:

$$
\begin{aligned}
L &= \prod_{I_{it}\in 1,2,J_{it}=1}\Big\{F\left(\eta_{it}>-W_{it}\delta,\varepsilon_{it}>-Z_{it}\gamma\right)f\left(u_{1it}\mid\eta_{it}>-W_{it}\delta,\varepsilon_{it}>-Z_{it}\gamma\right)\\
&\quad +F\left(\eta_{it}>-W_{it}\delta,\varepsilon_{it}\le -Z_{it}\gamma\right)f\left(u_{2it}\mid\eta_{it}>-W_{it}\delta,\varepsilon_{it}\le -Z_{it}\gamma\right)\Big\}\\
&\times \prod_{J_{it}=0}F\left(\eta_{it}\le -W_{it}\delta\right)
\end{aligned}
\tag{9}
$$

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

$$
\Sigma=\begin{bmatrix}\sigma_1^2 & \sigma_{12} & \sigma_{1\varepsilon} & \sigma_{1\eta}\\ & \sigma_2^2 & \sigma_{2\varepsilon} & \sigma_{2\eta}\\ & & 1 & \sigma_{\varepsilon\eta}\\ & & & 1\end{bmatrix}
\tag{10}
$$

Note that $\sigma_{12}$ 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

$$
\begin{aligned}
f\left(u_{1it},\varepsilon_{it},\eta_{it}\mid\eta_{it}>-W_{it}\delta,\varepsilon_{it}\le -Z_{it}\gamma\right) &= \frac{f\left(u_{1it},\varepsilon_{it},\eta_{it}\right)}{\Phi\left(W_{it}\delta,Z_{it}\gamma,\rho_{\varepsilon\eta}\right)}\\
&= \frac{f\left(u_{1it}\right)f\left(\varepsilon_{it},\eta_{it}\mid u_{1it}\right)}{\Phi\left(W_{it}\delta,Z_{it}\gamma,\rho_{\varepsilon\eta}\right)}
\end{aligned}
\tag{11}
$$

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

$$
f\left(u_{1it}\mid\eta_{it}>-W_{it}\delta,\varepsilon_{it}\le -Z_{it}\gamma\right)=\frac{f\left(u_{1it}\right)\int_{-W_{it}\delta}^{\infty}\int_{0}^{-Z_{it}\gamma}f\left(\varepsilon_{it},\eta_{it}\mid u_{1it}\right)\mathrm{d}\varepsilon_{it}\mathrm{d}\eta_{it}}{\Phi\left(W_{it}\delta,-Z_{it}\gamma,\rho_{\varepsilon\eta}\right)}
\tag{12}
$$

noting that

$$
f\left(\varepsilon_{it},\eta_{it}\mid u_{1it}\right)\sim N\left(\begin{pmatrix}\frac{\rho_{1\varepsilon}}{\sigma_1}\left(y_{it}-X_{it}\beta_1\right)\\ \frac{\rho_{1\eta}}{\sigma_1}\left(y_{it}-X_{it}\beta_1\right)\end{pmatrix},\begin{pmatrix}1-\rho_{\varepsilon 1}^2 & \sigma_{\varepsilon\eta}-\rho_{1\varepsilon}\rho_{1\eta}\\ & 1-\rho_{\eta 1}^2\end{pmatrix}\right)
\tag{13}
$$

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

$$
\begin{aligned}
&f\left(u_{1it}\mid\eta_{it}>-W_{it}\delta,\varepsilon_{it}\le -Z_{it}\gamma\right)\\
&=\frac{\Phi\left(-\frac{Z_{it}\gamma+\frac{\rho_{1\varepsilon}}{\sigma_1}\left(y_{it}-X_{it}\beta_1\right)}{\sqrt{1-\rho_{1\varepsilon}^2}},\frac{W_{it}\delta+\frac{\rho_{1\eta}}{\sigma_1}\left(y_{it}-X_{it}\beta_1\right)}{\sqrt{1-\rho_{1\eta}^2}},-\frac{\rho_{\varepsilon\eta}-\rho_{1\varepsilon}\rho_{1\eta}}{\sqrt{1-\rho_{1\varepsilon}^2}\sqrt{1-\rho_{1\eta}^2}}\right)\phi\left(\frac{y_{it}-X_{it}\beta_1}{\sigma_1}\right)/\sigma_1}{\Phi\left(W_{it}\delta,-Z_{it}\gamma,-\rho_{\varepsilon\eta}\right)}
\end{aligned}
\tag{14}
$$

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

$$\begin{aligned} & f\left(u_{2it} \mid \eta_{it} > -W_{it}\delta, \varepsilon_{it} > -Z_{it}\gamma\right) \qquad (15) \\ = & \frac{\Phi\left(\frac{Z_{it}\gamma+\frac{\rho_{2\varepsilon}}{\sigma_2}(\Delta y_{it}-X_{it}\beta_2)}{\sqrt{1-\rho_{2\varepsilon}^2}}, \frac{W_{it}\delta+\frac{\rho_{2\eta}}{\sigma_2}(\Delta y_{it}-X_{it}\beta_2)}{\sqrt{1-\rho_{2\eta}^2}}, \frac{\rho_{\varepsilon\eta}-\rho_{2\varepsilon}\rho_{2\eta}}{\sqrt{1-\rho_{2\varepsilon}^2}\sqrt{1-\rho_{2\eta}^2}}\right)\phi\left(\frac{\Delta y_{it}-X_{it}\beta_2}{\sigma_2}\right)/\sigma_2}{\Phi\left(W_{it}\delta, Z_{it}\gamma, \rho_{\varepsilon\eta}\right)} \end{aligned}$$

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

$$\begin{aligned} L = & \prod_{I\in 1,2}\left\{\Phi\left(-\frac{Z_{it}\gamma+\frac{\rho_{1\varepsilon}}{\sigma_1}(y_{it}-X_{it}\beta_1)}{\sqrt{1-\rho_{1\varepsilon}^2}}, \frac{W_{it}\delta+\frac{\rho_{1\eta}}{\sigma_1}(y_{it}-X_{it}\beta_1)}{\sqrt{1-\rho_{1\eta}^2}}, -\frac{\rho_{\varepsilon\eta}-\rho_{1\varepsilon}\rho_{1\eta}}{\sqrt{1-\rho_{1\varepsilon}^2}\sqrt{1-\rho_{1\eta}^2}}\right)\right. \\ & \phi\left(\frac{y_{it}-X_{it}\beta_1}{\sigma_1}\right)/\sigma_1 \\ & + \Phi\left(\frac{Z_{it}\gamma+\frac{\rho_{2\varepsilon}}{\sigma_2}(\Delta y_{it}-X_{it}\beta_2)}{\sqrt{1-\rho_{2\varepsilon}^2}}, \frac{W_{it}\delta+\frac{\rho_{2\eta}}{\sigma_2}(\Delta y_{it}-X_{it}\beta_2)}{\sqrt{1-\rho_{2\eta}^2}}, \frac{\rho_{\varepsilon\eta}-\rho_{2\varepsilon}\rho_{2\eta}}{\sqrt{1-\rho_{2\varepsilon}^2}\sqrt{1-\rho_{2\eta}^2}}\right) \\ & \left.\phi\left(\frac{\Delta y_{it}-X_{it}\beta_2}{\sigma_2}\right)/\sigma_2\right\} \\ & \times \prod_{I=3}\Phi\left(-Z_{it}\gamma\right) \qquad (16) \end{aligned}$$

The effect of lifelong learning can be examined by including an appropriate variable among the regressors in each of the equations in the model. However, the possibility of lifelong learning decisions being endogenous needs to be taken into account. Thus the estimation needs take account of possible correlations between $\nu_{it}$ in equation (8) and the errors in the four equations (4 to 7) of the main model. Ideally, this would be dealt with by jointly estimating all five equations. However, to avoid the computational burden involved with higher-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 $\Delta y_{it}$ and $y_{it}$ equations). The full covariance matrix can be written

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

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

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

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:

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

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 $\omega_{1it}$. We can substitute this out to give

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

Kim's approach addresses the case of a continuous endogenous regressor and involves including a residual term from the regression of the endogenous variable on instrumental variables uncorrelated with the error terms in the outcome equations in order to overcome the endogeneity-induced bias. The significance of the estimated coefficient attached to the residual term provides a test of endogeneity. Our case is slightly different in that the potentially endogenous regressor – the acquisition of a lifelong learning qualification – is binary 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 $\omega_{1it}$ but not with $\omega_{kit}$ 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 $\omega_{kit}$ terms are independent standard normal, our model becomes:

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

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.

The model is estimated using maximum likelihood. The nature of the model, in particular the fact that the two regimes cannot be observed, means that including individual effects is problematic. Consequently, it is estimated on a pooled dataset. 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.

## 3.10 Control Variables

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 variables operate as instrumental variables in order to aid identification. Variables appearing in the employment equation only are family background variables – whether partnered at time $t-1$ and whether children were present in the household at that time. The intuition behind this exclusion restriction is that individuals within a couple are able to specialise into paid and non-paid labour, depending on their preferences and comparative earnings potential, and the advantage of such specialisation becomes greater when there are dependent children. Consequently, household composition may influence labour supply decisions but should not affect wages. The switching equation alone includes the variable *Wave Gap* which indicates the interval between interviews and a variable *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. Finally, the equation used to estimate the generalised residuals includes dummies for individual years (in order to capture exogenous policy shifts).

Other variables included in the model merit some mention. We control for educational status in 1991 or, if later, at the age of twenty-five, using six dummies to indicate this. These are *Orig Qual 1-Orig Qual 4* indicating the level of educational attainment reported in the 1991 BHPS or when the respondent was aged twenty-five. Further dummies *Orig Qual other* and *Highest Qual Academic* indicate whether someone at that time held a qualification which cannot be placed in the standard scale and whether their highest qualification level was academic or vocational, respectively.

The effect of rising overall prosperity is controlled for by including the growth rate of GDP in equations 2, 3 and 4 – the wage equation for stayers, the switching equation and the employment equation. The logic behind this is that the rise people receive if their real wage is linked to that of the previous year may depend on overall economic performance, as may the probabilities of them being a stayer and of being employed; we use GDP growth to represent overall economic performance. By contrast, we expect the wage rate of stayers to depend on the ability of the economy to pay, and this is indicated by the log of the level of GDP rather than by its rate of change.

If someone was not employed in period $t-1$ they cannot be a stayer because, to measure the growth in earnings in year $t$ relative to $t-1$, they need to be employed in both years. We impose this on our model by allowing a dummy *Newly Employed* which takes a value 1 if someone works in year $t$ but not in $t-1$ to enter equation 3 with a large negative coefficient; we select 10 although the results are not in any way sensitive to this choice. This creates a probability indistinguishable from 1 that such an individual is a mover.

In the first differences model we made the assumption that differences between individual earnings associated with people's capacity to acquire qualifications rather than the actual acquisition of qualifications were removed by differencing. But this argument cannot be sustained for our movers' equation which is in the log level rather than the log difference of hourly wages. However, the panel nature of the British Household Panel Survey allows us to make a proper distinction between the effects of lifelong learning and the characteristics of people who undergo lifelong learning. We introduce two additional dummy control variables. The first, *Sometime Acquired*, takes a value 1 for people who gain qualifications at some time in the dataset and 0 otherwise. The second, *Sometime Upgraded*, takes a value 1 for people who upgrade at some time in the dataset and 0 otherwise. These capture the impact of the characteristics of people who acquire qualifications whether or not they have actually done so, and thus allow us to distinguish the actual acquisition of qualifications from the characteristics of the people who acquire them.

## 3.11 Parameters of the Mover-Stayer Model

Our main results are presented in table 6. This specification of the model is the result of testing down from the full parameterisation in order to obtain a sharper focus on the effects of lifelong learning. The results for the general model before imposing any restrictions are provided in Appendix A. The parameters of the ordered probit equations which generate the generalised residuals are shown separately in Appendix B. As a summary comment, our preferred (restricted) model provides results which, while clearer, are essentially similar to those provided by the unrestricted model.

We find that movers augment their earnings by 0.06 log units, significant at a ten per cent level, on acquiring a qualification, independently of whether they upgrade or not. This effect is not time dependent. Stayers, by contrast, gain an increment to hourly earnings of just over 0.06 log units between one and two years after upgrading their educational status and this is significant at a five per cent level; there is no significant effect on stayers who do not upgrade.

If this were the limit of the effects of qualification acquisition, the conclusion would be that people who were likely to be movers would tend to gain from qualification acquisition while those likely to be stayers would gain from upgrading. Should the stayers eventually be movers, they would gain from the premium paid to movers but of course might be worse off than if they had remained on the progression path available to stayers.

However, when we look at the employment equation in table 6 we can see that upgrading also affects employment prospects. People who upgrade enjoy a positive impact on their chance of employment, as represented by the *Ever Upgraded* terms. However for the first three years this is offset by negative effects, shown by the coefficients on the terms $Upgraded_t ... Upgraded_{t-2}$. Since these impact effects add to the permanent effect, we can see that the net impact on employment is decreasingly negative in the time since the qualification was acquired with the net impact being small in year *t-2* and becoming, of course, positive thereafter. This may reflect the sort of problems that some people have in finding suitable jobs after gaining qualifications, as discussed by Purcell et al. (2007). Nevertheless the overall effect of upgrading is enhanced by the improved long-term employment prospects shown in the employment equation. This has a further effect on the earnings of movers, because we see in the movers' equation that being *Newly Employed* faces a highly statistically significant wage penalty of 0.35 log units. Thus someone whose employment prospects are enhanced gains not only because they are more likely to be employed but also because their earnings are likely to be higher when they are employed.

With regard to the other coefficients, we see that wage rates of movers depend on their qualifications, with level 1 offering a premium of 0.13 log units, level 2 a premium of 0.24 log units rising to 0.48 log units for level 4. The penalty associated with being newly employed is 0.32 log units. Among stayers, those originally qualified to level 4 have a rate of growth of earnings 0.015 log units per annum larger than those with no qualifications, and the effect is statistically significant. Other qualification levels do affect earnings growth, apparently negatively in the case of level 2, but the effects are not statistically significant. Turning to the switching equation, we can see that the chance of someone being a stayer is increasing in their original level of education and that the effect is significantly enhanced if their highest qualification is academic. The positive coefficient on *Employed at Start* indicates that people initially employed are more likely to be stayers. The significant negative term on *Wave Gap* indicates that someone is more likely to be a mover the longer the gap between interviews while that on *Recent Job* shows that someone who has recently changed job is more likely to be a mover. Thus, these two terms are related to the latent variable in the way that would be expected. Looking at the employment equation, we see that educational status increases the chance of employment, although there is little to choose between the effects of levels 2, 3 and 4. Being employed when first observed has a very strong influence on subsequent employment prospects.

Finally, exponentiating the log standard error terms at the bottom of table 7 shows the standard error of mover equation to be 0.38, while that of the stayer equation is 0.14. These compare with the standard error of the free-standing regression in first differences of 0.245. Thus the model does indeed separate out those who move around the earnings distribution from those who stay close to their previous situation; the greater statistical significance of the effects of qualification upgrading in the stayer equation as compared with the OLS regression may be attributed to the extra clarity offered by making this distinction. We also note that the statistically significant correlations between the residuals of the different equations indicate the importance of this structure as compared with, say, considering employment and earnings effects independently.

| N=12888 | Equation 1m Mover | | | Equation 2m Stayer | | | Equation 3m Switching | | | Equation 4m Employment | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | coeff. | s.e. | | coeff. | s.e. | | coeff. | s.e. | | coeff. | s.e. | |
| Upgraded$_t$ | | | | | | | | | | -0.730 | 0.218 | *** |
| Upgraded$_{t-1}$ | | | | 0.065 | 0.023 | *** | | | | -0.625 | 0.186 | *** |
| Upgraded$_{t-2}$ | | | | | | | | | | -0.442 | 0.147 | *** |
| Ever Acquired | 0.060 | 0.034 | * | | | | | | | | | |
| Ever Upgraded | | | | | | | | | | 0.407 | 0.187 | ** |
| Orig Qual 1 | 0.128 | 0.045 | *** | 0.002 | 0.005 | | 0.172 | 0.124 | | 0.119 | 0.123 | |
| Orig Qual 2 | 0.241 | 0.050 | *** | -0.006 | 0.006 | | 0.406 | 0.132 | *** | 0.290 | 0.153 | * |
| Orig Qual 3 | 0.277 | 0.057 | *** | 0.006 | 0.005 | | 0.574 | 0.126 | *** | 0.378 | 0.124 | *** |
| Orig Qual 4 | 0.483 | 0.068 | *** | 0.015 | 0.005 | *** | 0.983 | 0.141 | *** | 0.326 | 0.133 | ** |
| Orig Qual other | 0.034 | 0.033 | | 0.000 | 0.003 | | 0.024 | 0.087 | | -0.108 | 0.093 | |
| High Qual Academic | 0.019 | 0.040 | | 0.005 | 0.003 | | 0.260 | 0.096 | *** | 0.053 | 0.095 | |
| Age | 0.049 | 0.014 | *** | -0.005 | 0.001 | *** | 0.128 | 0.031 | *** | 0.125 | 0.026 | *** |
| Age$^2$/ 100 | -0.046 | 0.017 | *** | 0.004 | 0.002 | ** | -0.138 | 0.039 | *** | -0.190 | 0.031 | *** |
| Not White | -0.088 | 0.070 | | -0.005 | 0.009 | | -0.239 | 0.187 | | -0.106 | 0.179 | |
| London | 0.058 | 0.053 | | 0.010 | 0.006 | * | -0.090 | 0.164 | | -0.307 | 0.138 | ** |
| South-West | -0.137 | 0.057 | ** | -0.003 | 0.005 | | -0.075 | 0.132 | | 0.073 | 0.135 | |
| East Anglia | -0.119 | 0.071 | * | 0.010 | 0.007 | | 0.028 | 0.163 | | 0.068 | 0.181 | |
| East Midlands | -0.128 | 0.051 | ** | 0.005 | 0.005 | | 0.066 | 0.138 | | -0.264 | 0.149 | * |
| West Midlands | -0.114 | 0.046 | ** | 0.000 | 0.005 | | 0.003 | 0.132 | | -0.066 | 0.132 | |
| North-West | -0.112 | 0.053 | ** | 0.000 | 0.005 | | -0.257 | 0.130 | ** | -0.288 | 0.136 | ** |
| Yorks \Humb. | -0.235 | 0.047 | *** | 0.004 | 0.005 | | -0.121 | 0.128 | | -0.079 | 0.144 | |
| North | -0.140 | 0.055 | ** | -0.003 | 0.006 | | -0.281 | 0.145 | * | -0.469 | 0.136 | *** |
| Wales | -0.177 | 0.060 | *** | 0.004 | 0.006 | | -0.464 | 0.161 | *** | -0.501 | 0.146 | *** |
| Scotland | -0.149 | 0.056 | *** | 0.004 | 0.004 | | -0.293 | 0.146 | ** | -0.216 | 0.156 | |
| Employed at Start | -0.054 | 0.077 | | -0.010 | 0.012 | | 0.785 | 0.229 | *** | 1.825 | 0.090 | *** |
| Ln GDP | 0.544 | 0.111 | *** | | | | | | | | | |
| $\Delta$ ln GDP | | | | 0.309 | 0.207 | | -1.601 | 3.548 | | 0.052 | 2.072 | |
| Newly Employed | -0.354 | 0.116 | *** | | | | -10.000 | - | | | | |
| Wave Gap | | | | | | | -0.058 | 0.023 | ** | 0.019 | 0.012 | |
| Recent Job | | | | | | | -0.719 | 0.090 | *** | | | |
| Children Lagged | | | | | | | | | | -0.070 | 0.074 | |
| Partner Lagged | | | | | | | | | | 0.300 | 0.078 | *** |
| Gen. Residual | -0.022 | 0.018 | | 0.000 | 0.004 | | -0.108 | 0.048 | ** | 0.046 | 0.044 | |
| Sometime Acquired | -0.045 | 0.037 | | 0.000 | 0.003 | | 0.073 | 0.083 | | 0.255 | 0.093 | *** |
| Sometime Upgraded | 0.071 | 0.040 | * | 0.002 | 0.005 | | -0.069 | 0.122 | | -0.209 | 0.155 | |
| Constant | -1.234 | 0.585 | ** | 0.127 | 0.036 | *** | -2.049 | 0.739 | *** | -2.538 | 0.550 | *** |

* Significant at 10% ** Signficant at 5% *** Significant at 1%

Table 6: Model Parameters

| | Coeff. | Std. Err | z-stat |
|---|---|---|---|
| ln $\sigma_1$ | -0.963 | 0.043 | -22.19 |
| ln $\sigma_2$ | -1.991 | 0.021 | -93.85 |
| arctanh $\rho_{1\eta}$ | -0.608 | 0.174 | -3.48 |
| arctanh $\rho_{2\eta}$ | -0.122 | 0.102 | -1.20 |
| arctanh $\rho_{1\varepsilon}$ | 0.344 | 0.30 | 1.15 |
| arctanh $\rho_{2\varepsilon}$ | 0.174 | 0.054 | 3.20 |
| arctanh $\rho_{\eta\varepsilon}$ | 0.523 | 0.192 | 2.73 |

Table 7: Standard Errors and Correlations in the Restricted Model

## 3.12 Marginal Probabilities

Table 8 presents the estimated marginal effects of lifelong learning on the probability of being employed, the probability of being a stayer for those who are observed to be employed and, lastly, the probability of being both employed and a stayer. Where a variable features in either the employment equation or the switching equation but not both, its marginal effect will be zero in the equation from which it is absent. However, the non-zero marginal effect in the other equation will influence the joint marginal effect (of being employed and a stayer). The format of table 8 is that for each variable, the probability of being employed, the conditional probability of being a stayer and the (unconditional) probability of being employed and a stayer are calculated setting the variable in question to zero (the 'no' column) or to one (the 'yes'column). When considering age, the reference category is age 30.

By definition, the marginal effects on the probability of being a stayer and the probability of employment have the same sign as their corresponding coefficients in the model. However, presenting the results as marginal effects allows a better understanding of the magnitude of the effects and how the influences combine across the employment and switching equations.

The results point to a negative effect of upgrading on employment that reduces over time such that in the long-term (specifically, after three years) upgrades exert a positive impact of 5.5 percentage points. Since lifelong learning does not enter into the switching equation, the effects on the joint probability of being employed and a stayer are determined by the employment equation. Overall, the long-term effect of lifelong learning is 4.5 percentage points. The negative short-term impacts of lifelong learning may reflect the time taken to find acceptable employment following a period of training.

Some of the other marginal effects presented are also interesting. Higher qualifications, particularly academic ones, are associated with increased employment and a higher probability of being a stayer. Also, men aged 50 are less likely than those aged 30 to be in work but more likely to be stayers if they are.

| | Pr(emp) | | | Pr(stayer \| emp) | | | Pr(stayer, emp) | | |
|---|---|---|---|---|---|---|---|---|---|
| | no | yes | mfx | no | yes | mfx | no | yes | mfx |
| Upgraded$_t$ | 0.832 | 0.775 | -0.057 | | | | 0.687 | 0.640 | -0.047 |
| Upgraded$_{t-1}$ | 0.832 | 0.795 | -0.037 | | | | 0.687 | 0.657 | -0.031 |
| Upgraded$_{t-2}$ | 0.832 | 0.826 | -0.006 | | | | 0.687 | 0.683 | -0.005 |
| Ever Upgraded | 0.832 | 0.886 | 0.055 | | | | 0.687 | 0.732 | 0.045 |
| Orig Qual 1 | 0.804 | 0.825 | 0.021 | 0.727 | 0.776 | 0.049 | 0.585 | 0.640 | 0.055 |
| Orig Qual 2 | 0.804 | 0.852 | 0.048 | 0.727 | 0.833 | 0.106 | 0.585 | 0.710 | 0.125 |
| Orig Qual 3 | 0.804 | 0.865 | 0.060 | 0.727 | 0.868 | 0.141 | 0.585 | 0.750 | 0.166 |
| Orig Qual 4 | 0.804 | 0.857 | 0.053 | 0.727 | 0.930 | 0.203 | 0.585 | 0.797 | 0.213 |
| Orig Qual other | 0.837 | 0.820 | -0.017 | 0.825 | 0.830 | 0.005 | 0.691 | 0.681 | -0.010 |
| High Qual Academic | 0.831 | 0.839 | 0.008 | 0.795 | 0.853 | 0.058 | 0.661 | 0.716 | 0.055 |
| Age 40 | 0.894 | 0.884 | -0.010 | 0.781 | 0.853 | 0.072 | 0.698 | 0.755 | 0.056 |
| Age 50 | 0.894 | 0.814 | -0.080 | 0.781 | 0.861 | 0.080 | 0.698 | 0.701 | 0.002 |
| Not White | 0.835 | 0.818 | -0.017 | 0.828 | 0.770 | -0.058 | 0.691 | 0.630 | -0.061 |
| London | 0.860 | 0.813 | -0.047 | 0.851 | 0.832 | -0.019 | 0.732 | 0.677 | -0.056 |
| South-West | 0.860 | 0.870 | 0.010 | 0.851 | 0.836 | -0.016 | 0.732 | 0.727 | -0.005 |
| East Anglia | 0.860 | 0.869 | 0.009 | 0.851 | 0.857 | 0.005 | 0.732 | 0.744 | 0.012 |
| East Midlands | 0.860 | 0.820 | -0.040 | 0.851 | 0.864 | 0.013 | 0.732 | 0.709 | -0.023 |
| West Midlands | 0.860 | 0.851 | -0.009 | 0.851 | 0.852 | 0.001 | 0.732 | 0.725 | -0.007 |
| North-West | 0.860 | 0.816 | -0.044 | 0.851 | 0.794 | -0.058 | 0.732 | 0.648 | -0.085 |
| Yorks | 0.860 | 0.849 | -0.011 | 0.851 | 0.826 | -0.026 | 0.732 | 0.701 | -0.031 |
| North | 0.860 | 0.783 | -0.077 | 0.851 | 0.788 | -0.064 | 0.732 | 0.617 | -0.115 |
| Wales | 0.860 | 0.777 | -0.083 | 0.851 | 0.739 | -0.112 | 0.732 | 0.574 | -0.158 |
| Scotland | 0.860 | 0.828 | -0.032 | 0.851 | 0.785 | -0.067 | 0.732 | 0.650 | -0.082 |
| Started current job within last year | | | | 0.865 | 0.676 | -0.189 | 0.722 | 0.564 | -0.158 |
| GDP growth of 1 per cent | 0.834 | 0.834 | 0.000 | 0.836 | 0.833 | -0.003 | 0.698 | 0.695 | -0.003 |
| Children Lagged | 0.837 | 0.827 | -0.011 | | | | 0.692 | 0.683 | -0.009 |
| Partner Lagged | 0.796 | 0.846 | 0.050 | | | | 0.658 | 0.699 | 0.041 |

Table 8: Marginal Probabilities

# 4 Returns to lifelong Learning

As the results above indicate, there are a number of effects present which combine to determine the return to lifelong learning. For men who upgrade, lifelong learning increases the chance of employment in the long-run and this reduces the chance that they will face the wage penalty shown in the movers' equation after unemployment. There is also the additional influence of a rise in the wage rate accruing to stayers after a lag of one year. In order to identify the combined influence of these effects, we simulate our system of equations, comparing the life-time earnings of people who have undertaken lifelong learning, with or without upgrading their qualifications, with those of people who do not.

In the simulations we assume that there are twenty thousand individuals of each type considered (initial qualification level 0 to 4; highest qualification academic or not; employed/not employed at age twenty-five). These individuals are subject to random disturbances which affect their wages as both movers and stayers and the latent variables which determine whether they are movers or stayers and whether they are employed or not. The disturbances are assumed to be normally distributed, with a covariance matrix the same as that presented above. The regional dummies are replaced by variables showing the proportion of the population aged twenty to sixty of non- self-employed men in each region. Since the *Not White* racial dummy variables are poorly defined we carry out the simulations for people who are white. We also assume that people do not have children.

The effects of lifelong learning both with and without upgrading are computed for men aged twenty-five and forty-five. We assume that future wages are discounted at a rate of 5% p.a. back to the date at which the qualifications were acquired. In table 9 we show the percentage increase in the present discounted value of wage income between acquisition of the qualification (at age twenty-five or forty) and age sixty. We show the results in two forms. First of all we show only the wage effects. These are calculated from the simulated wages generated for each period, on the assumption that these are actually earned in every period (even if affected by non-employment in the previous period). In the second group of results we take account of the influence of lifelong learning on employment, in that these figures reflect the probability of being employed in each period. Thus these figures represent the overall impact of lifelong learning.

The impact of qualification acquisition without upgrading on wage rates is generally shown to be small. The wage effect occurs as a result of the impact identified in the movers' equation. This impact declines with age because the benefits of this form of lifelong learning are realised only by movers, and the probability of being a mover falls with age. This effect is particularly marked for people qualified to level four because, as table 8 shows, the probability of such men being movers is in any case low; the non-linear nature of the probit function leads to an interaction between the qualification effect and the age effect which has a particularly powerful impact.

Upgrading of qualifications, by contrast, generates a higher return, and this is particularly marked for people initially at level 0 with only minimal qualifications. The direct effect on wages is larger than for people who do not upgrade because there is the effect shown in the stayers' equation which is in addition to the benefit of obtaining any qualification shown in the movers' equation. However, as we noted, our employment equation suggests that upgrading has an impact on employment prospects and once this is taken into account, the effect of upgrading is increased further, with the impact being particularly marked for men initially at level 0.

In table 8 we see that the marginal effect of upgrading on the probability of employment is 5.5 percentage points, whereas in table 9 we see that, for someone initially at level 0, the employment effect raises the return to upgrading from about 9% to about 21-22%, depending on the age at which this takes place. There are two factors behind this apparent discrepancy. First of all, the employment rate for such men is low. The model generates a rate which peaks at 77% for men in their early thirties and falls to 42% by age 60. Thus a 5.5 percentage point increase in the probability of employment is equivalent to a 7% increase in the number of men employed in their thirties, increasing to a 13% increase in the number of men employed by the age of 60; it is these latter proportions which affect the return to qualifications. Secondly, the marginal probability is calculated for men age 30. As the base-line probability of employment decreases with age, so that the value of the underlying latent variable falls back from around 1 to around 0, so the impact of the term in *Ever Upgraded* in the employment equation on the probability of employment is increased. These two effects, taken together lead to the impact of upgrading on employment raising the return for people initially at level 0 by 12-13 percentage points.

# 5 Comparison with Other Approaches

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, 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 would be

| | Initial Attainment Level | Age 25 | | Age 40 | |
|---|---|---|---|---|---|
| | | Wage Effect Only | Total Effect | Wage Effect Only | Total Effect |
| No Upgrading | 0 | 6.0% | 6.0% | 5.8% | 5.6% |
| | 1 | 5.9% | 5.9% | 5.4% | 5.3% |
| | 2 | 5.6% | 5.6% | 4.9% | 4.8% |
| | 3 | 5.5% | 5.4% | 4.6% | 4.6% |
| | 4 | 4.7% | 4.7% | 1.8% | 1.6% |
| Upgrading | 0 | 9.3% | 21.0% | 9.0% | 21.7% |
| | 1 | 8.9% | 14.4% | 9.3% | 16.3% |
| | 2 | 8.6% | 12.3% | 8.9% | 14.1% |
| | 3 | 8.9% | 12.2% | 9.1% | 13.5% |

Table 9: Returns to lifelong Learning for Men

statistically insignificant. We can reject the hypothesis that all the non-constant terms, together with the relevant co-variances, are insignificant strongly, with $\chi^2_{26} = 268.95$ as the Wald statistic computed from the restricted model.

This test also allows us to reject the idea that the data are fully described by the first-difference model of section 3.2. As we noted in section 3.3, it is impossible for people always to be stayers unless they are always fully employed. However 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 statistic 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, despite our findings in section 3.2, it is not appropriate to ignore employment selection effects.

While these observations suggest that we cannot undertake formal testing of our model relative to earlier work, we can, nevertheless comment on how our results differ from other findings; the most appropriate study for comparison is that by Blanden et al. (2008) since they also worked with the British Household Panel Survey. They find greater age differences than we do; those we find arise in both the switching equation and in the influence of qualification acquisition on the probability of employment. Blanden et al. (2008) do not make the distinction we made between acquisition without upgrading and upgrading (although they do produce some results which relate the effects of lifelong learning to initial qualifications). They do, however, find a return to lifelong learning for young men of 8% after four years; they regard this as a long-run effect while we show influences only from upgrading. A greater difference is that they find a negative return to lifelong learning for men aged 50 and older, while our model suggests that they enjoy the benefits of a greater chance of employment. Thus a general conclusion is that we find more substantial evidence than do Blanden et al. (2008) for economic benefits accruing to men as a result of lifelong learning. We note, however, that such a comparison is complicated by the fact that Blanden et al. (2008) use a somewhat different definition of lifelong learning.

# 6 Conclusions

In this paper we have investigated the effect of lifelong learning on men's earnings using data from the British Household Panel Survey. We have done this using a model of wage evolution structured round a switching regression, with the conventional approach to such a regression equation being extended to take account of endogenous employment effects. Our results offer a more positive view of the effects of lifelong learning on hourly earnings than has been found by other researchers. We find that qualifications which result in men's educational status being upgraded have a clear effect on their earnings and these are considerable amplified once one takes account of employment effects.

Methodologically, the results demonstrate that studies that omit to address the employment decision risk introducing selection bias. Also, they are unable to capture an important role of lifelong learning in increasing employment. The existence of two regimes for wage determination is strongly supported by the results and this structure permits a more nuanced understanding of the role of lifelong learning than is possible under the more usual approach of assuming a single wage equation. 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 characterstics of people who, at some point, undertake lifelong learning.

These findings offer perhaps a stronger indication of the value of lifelong learning than is visible from the other studies considered in this paper. As noted in the Introduction, it is common for government policy to encourage lifelong learning. 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


Blanden, J., Buscha, F., Sturgis, P. & Urwin, P. (2008), Earnings and Occupational Status Returns to Lifelong Learning. University of Surrey.

Cappellari, L. & Jenkins, S. (2008), 'Estimating Low Pay Transition Probabilities accounting for Endogenous Selection Mechanisms', *Journal of the Royal Statistical Society, Series C* pp. 165–186.

Dickens, R. (2000), 'The Evolution of Individual Male Earnings in Great Brtain: 1975-95', *Economic Journal* **110**, 27–49.

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.

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.

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. CEE Discussion Paper No 19, Centre for the Economics of Education.

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.

Lillard, E. & Willis, R. (1978), 'Dynamic Aspects of Earnings Mobility', *Econometrica* **46**, 985–1012.

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

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.

Ramos, D. (2003), 'The Covariance Strucutre of Earnings in Great Britain', *Economica* **70**, 353–374.

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: Full model and testing results

We show the parameters of the unrestricted model in table 10 . The unrestricted results suggest a number of possible but not very well determined, influences. In the movers' equation we find that the *Ever Acquired* terms have effects which are positive and of plausible magnitude but which are not statistically significant for any of the three age groups. We can see significant effects associated with upgrades to educational status in the stayers' equation after a lag of one year. This coefficient, of 0.059 is obviously representing the same influences as were shown in the simple first-difference model of equation 5, but the coefficient is now higher and its statistical significance has increased. *Ever Upgraded 25-34* has a negative coefficient in the switching equation which is statistically significant at the ten per cent level and the influence of *Ever Upgraded 50-60* is of similar magnitude although not significant . However the impact of *Ever Upgraded 35-49* is of opposite sign. We also see, in the employment equation, that the effect of an upgrade influences the probability of employment in a manner which is statistically significant for men aged 35-49 at the ten per cent level. There are weaker and statistically insignificant effects for younger and older men. However, had the three age groups not been distinguished, the coefficient on upgrading would be statistically significant and the restriction is easily accepted as we show below. Separately the statistically significant negative coefficient on $Upgrade_{t-1}$ and the insignificant coefficient of similar magnitude on $Upgrade_t$ suggest that in the short-term upgrading has an adverse effect on employment. Of course the overall impact each year after upgrading is given after adding to these coefficients the term in *Ever Upgraded* for the relevant age group.

We develop the preferred specification set out in table 6 by imposing restrictions sequentially on the unrestricted model. We have considered related restrictions in sequence rather than simply restricting coefficients in order of increasing statistical significance so as to arrive at a parsimonious model. Thus, when considering the terms in *Acquired* and *Upgraded* we have tested jointly that these are zero, except where, as in the employment equation, there was strong evidence that some were not. Similarly, where there seemed the possibility that restricting the age-dependent terms *Ever Acquired* and *Ever Upgraded* to be equal across the different age bands, we have imposed this restriction without imposing the additional zero restriction, unless the age terms were individually highly insignificant.

The standard errors and covariances of the residuals of the different equations are shown in table 11. The z-statistics show that three of the five covariances are statistically significant.

Table 12 shows the $\chi^2$ tests for the restrictions as imposed sequentially in deriving the preferred specification. We note that the Wald test of all fifty-eight restrictions imposed jointly is accepted, but only with a p-value of 0.14. Given this, one might have some doubts on the validity of the individual restrictions. However, if only restrictions 1 to 10 are imposed, the p-value for the Wald test is 0.54, suggesting that the validity of restriction 11 is doubtful. The $\chi^2$ statistic is raised by 15.5 with the loss of three degrees of freedom, suggesting, on the basis of successive testing of the unrestricted model, the extra restriction is invalid. On the other hand, when considered with restrictions 1 to 10 already imposed, the test is accepted with the p-value of 0.16 shown in the table and on these grounds we impose the restriction. We note, however that the properties of the model are not materially affected if restriction 11 is not imposed.

| N=12888 | Mover | | | Stayer | | | Switching | | | Employment | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | coeff. | s.e. | | coeff. | s.e. | | coeff. | s.e. | | coeff. | s.e. | |
| Acquired$_t$ | -0.015 | 0.070 | | -0.003 | 0.014 | | -0.173 | 0.206 | | -0.206 | 0.168 | |
| Acquired$_{t-1}$ | -0.009 | 0.067 | | -0.012 | 0.010 | | -0.143 | 0.166 | | -0.129 | 0.142 | |
| Acquired$_{t-2}$ | -0.015 | 0.062 | | -0.012 | 0.010 | | -0.204 | 0.167 | | -0.092 | 0.139 | |
| Acquired$_{t-3}$ | 0.047 | 0.078 | | -0.010 | 0.010 | | -0.217 | 0.179 | | -0.137 | 0.135 | |
| Acquired$_{t-4}$ | 0.035 | 0.073 | | 0.006 | 0.010 | | -0.079 | 0.163 | | -0.089 | 0.107 | |
| Upgraded$_t$ | -0.126 | 0.111 | | -0.001 | 0.023 | | 0.383 | 0.337 | | -0.580 | 0.286 | ** |
| Upgraded$_{t-1}$ | 0.095 | 0.096 | | 0.059 | 0.026 | ** | 0.027 | 0.325 | | -0.596 | 0.249 | ** |
| Upgraded$_{t-2}$ | -0.083 | 0.100 | | -0.003 | 0.021 | | 0.285 | 0.318 | | -0.336 | 0.233 | |
| Upgraded$_{t-3}$ | -0.166 | 0.105 | | -0.005 | 0.021 | | 0.148 | 0.285 | | -0.009 | 0.250 | |
| Upgraded$_{t-4}$ | -0.070 | 0.121 | | -0.036 | 0.026 | | -0.073 | 0.268 | | -0.125 | 0.187 | |
| Ever Acquired 25-34 | 0.054 | 0.070 | | 0.013 | 0.008 | | 0.223 | 0.191 | | -0.068 | 0.213 | |
| Ever Acquired 35-49 | 0.049 | 0.066 | | -0.004 | 0.005 | | -0.079 | 0.159 | | 0.064 | 0.181 | |
| Ever Acquired 50-60 | 0.091 | 0.090 | | 0.003 | 0.007 | | 0.238 | 0.200 | | 0.233 | 0.224 | |
| Ever Upgraded 25-34 | 0.008 | 0.112 | | -0.011 | 0.022 | | -0.599 | 0.339 | * | 0.219 | 0.391 | |
| Ever Upgraded 35-49 | 0.046 | 0.100 | | 0.012 | 0.013 | | 0.284 | 0.301 | | 0.489 | 0.285 | * |
| Ever Upgraded 50-60 | 0.045 | 0.138 | | -0.006 | 0.017 | | -0.469 | 0.313 | | 0.247 | 0.333 | |
| Orig Qual 1 | 0.126 | 0.046 | *** | 0.003 | 0.005 | | 0.188 | 0.125 | | 0.114 | 0.121 | |
| Orig Qual 2 | 0.239 | 0.050 | *** | -0.005 | 0.006 | | 0.406 | 0.132 | *** | 0.289 | 0.153 | * |
| Orig Qual 3 | 0.268 | 0.060 | *** | 0.008 | 0.005 | | 0.585 | 0.125 | *** | 0.373 | 0.123 | *** |
| Orig Qual 4 | 0.474 | 0.071 | *** | 0.016 | 0.005 | *** | 0.976 | 0.145 | *** | 0.321 | 0.131 | ** |
| Orig Qual other | 0.028 | 0.034 | | 0.000 | 0.003 | | 0.035 | 0.088 | | -0.103 | 0.092 | |
| High Qual Academic | 0.019 | 0.041 | | 0.005 | 0.003 | | 0.252 | 0.096 | *** | 0.054 | 0.094 | |
| Age | 0.047 | 0.016 | *** | -0.004 | 0.002 | ** | 0.140 | 0.033 | *** | 0.115 | 0.027 | *** |
| Age$^2$ /100 | -0.044 | 0.020 | ** | 0.003 | 0.002 | | -0.155 | 0.042 | *** | -0.183 | 0.032 | *** |
| Not White | -0.089 | 0.070 | | -0.004 | 0.008 | | -0.211 | 0.189 | | -0.105 | 0.175 | |
| London | 0.060 | 0.054 | | 0.010 | 0.006 | * | -0.089 | 0.162 | | -0.306 | 0.137 | ** |
| South-West | -0.128 | 0.058 | ** | -0.003 | 0.005 | | -0.082 | 0.131 | | 0.082 | 0.136 | |
| East Anglia | -0.121 | 0.073 | * | 0.010 | 0.007 | | 0.041 | 0.166 | | 0.075 | 0.182 | |
| East Midlands | -0.127 | 0.052 | ** | 0.005 | 0.005 | | 0.060 | 0.139 | | -0.264 | 0.149 | * |
| West Midlands | -0.105 | 0.046 | ** | 0.000 | 0.005 | | 0.007 | 0.133 | | -0.076 | 0.131 | |
| North-West | -0.109 | 0.052 | ** | 0.000 | 0.005 | | -0.255 | 0.128 | ** | -0.289 | 0.135 | ** |
| Yorks \Humb. | -0.234 | 0.047 | *** | 0.004 | 0.005 | | -0.132 | 0.129 | | -0.069 | 0.145 | |
| North | -0.128 | 0.058 | ** | -0.003 | 0.006 | | -0.289 | 0.141 | ** | -0.466 | 0.135 | *** |
| Wales | -0.165 | 0.063 | *** | 0.003 | 0.006 | | -0.488 | 0.156 | *** | -0.498 | 0.146 | *** |
| Scotland | -0.143 | 0.058 | ** | 0.004 | 0.005 | | -0.298 | 0.148 | ** | -0.207 | 0.156 | |
| Employed at Start | -0.070 | 0.087 | | -0.004 | 0.013 | | 0.857 | 0.265 | *** | 1.828 | 0.089 | *** |
| Ln GDP | 0.529 | 0.112 | *** | | | | | | | | | |
| Δ ln GDP | | | | 0.313 | 0.211 | | -1.985 | 3.696 | | 0.815 | 2.103 | |
| Newly Employed | -0.321 | 0.140 | ** | | | | -10.000 | - | | | | |
| Wave Gap | | | | | | | -0.053 | 0.024 | ** | 0.016 | 0.012 | |
| Recent Job | | | | | | | -0.722 | 0.090 | *** | | | |
| Children Lagged | | | | | | | | | | -0.075 | 0.074 | |
| Partner Lagged | | | | | | | | | | 0.297 | 0.078 | *** |
| Gen. Residual | -0.006 | 0.024 | | 0.000 | 0.005 | | -0.109 | 0.079 | | 0.083 | 0.055 | |
| Sometime Acquired | -0.051 | 0.040 | | 0.002 | 0.004 | | 0.094 | 0.107 | | 0.223 | 0.145 | |
| Sometime Upgraded | 0.074 | 0.053 | | 0.003 | 0.009 | | -0.033 | 0.178 | | -0.186 | 0.183 | |
| Constant | -1.125 | 0.608 | * | 0.095 | 0.041 | ** | -2.402 | 0.806 | *** | -2.229 | 0.570 | *** |

* Significant at 10% ** Signficant at 5% *** Significant at 1%

Table 10: Unrestricted Model Parameters

| | Coeff. | Std. Err | z-stat |
|---|---|---|---|
| ln $\sigma_1$ | -0.963 | 0.041 | -23.54 |
| ln $\sigma_2$ | -1.995 | 0.021 | -93.83 |
| arctanh $\rho_{1\eta}$ | -0.648 | 0.203 | -3.19 |
| arctanh $\rho_{2\eta}$ | -0.078 | 0.118 | -0.66 |
| arctanh $\rho_{1\varepsilon}$ | 0.272 | 0.361 | 0.75 |
| arctanh $\rho_{2\varepsilon}$ | 0.178 | 0.055 | 3.25 |
| arctanh $\rho_{\eta\varepsilon}$ | 0.589 | 0.239 | 2.47 |

Table 11: Standard Errors and Correlations in the Unrestricted Model

| No. | Restriction | d.f | $\chi^2$ | p |
|---|---|---|---|---|
| 1 | Acquired$_t$.. Acquired$_{t-4}$ all equations | 20 | 11.21 | 0.94 |
| 2 | Upgraded$_t$..Upgraded$_{t-4}$ Movers, Stayers and Switching excl Upgraded$_{t-1}$ in Stayers' Equation | 14 | 18 | 0.21 |
| 3 | Ever Acquired 25-34=Ever Acquired 35-49 =Ever Acquired 50-60=0 Stayers' Equation | 3 | 4.99 | 0.17 |
| 4 | Ever Upgraded 25-34=Ever Upgraded 35-49 =Ever Upgraded 50-60=0 Stayers' Equation | 3 | 0.5 | 0.92 |
| 5 | Ever Upgraded 25-34=Ever Upgraded 35-49 =Ever Upgraded 50-60=0 Movers' Equation | 3 | 0.71 | 0.87 |
| 6 | Ever Acquired 25-34=Ever Acquired 35-49 =Ever Acquired 50-60=0 Switching Equation | 3 | 3.92 | 0.27 |
| 7 | Upgraded$_{t-3}$ =Upgraded$_{t-4}$=0 Employment Equation | 2 | 1.73 | 0.42 |
| 8 | Ever Acquired 25-34=Ever Acquired 35-49 =Ever Acquired 50-60 Movers' Equation | 2 | 0.55 | 0.76 |
| 9 | Ever Upgraded 25-34=Ever Upgraded 35-49 =Ever Upgraded 50-60 Employment Equation | 2 | 0.97 | 0.62 |
| 10 | Ever Acquired 25-34=Ever Acquired 35-49 | 3 | 5 | 0.17 |
| 11 | Ever Upgraded 25-34=Ever Upgraded 35-49 | | | |
| | =Ever Upgraded 50-60 Switching Equation | 3 | 5.1 | 0.16 |
| | All restrictions imposed on unrestricted model | 58 | 69.9 | 0.14 |
| | All restrictions except 11 imposed on unrestricted model | 55 | 53.4 | 0.53 |

Table 12: Tests of Parameter Restrictions

# B Appendix: The Decision to Undertake lifelong Learning

In table 13 we show the results of the ordered probit regression which determines whether people undertake lifelong learning without upgrading their qualifications or with upgrading 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 men undertaking lifelong learning is not age-dependent. Being employed at the start of the survey or at age 25 (*Employed 1991)* increases the chance of undertaking lifelong learning.

| | | | |
|---|---|---|---|
| | Number of obs= | | 12896 |
| | LR chi2(36)= | | 254.54 |
| | Prob > $\chi^2$= | | 0 |
| | Pseudo $R^2$= | | 0.0311 |
| | Log likelihood | | -3962.8 |
| | Coef. | Std. Err. | z |
| Orig Qual 1 | 0.021 | 0.063 | 0.340 |
| Orig Qual 2 | 0.254 | 0.064 | 3.950 |
| Orig Qual 3 | 0.103 | 0.059 | 1.740 |
| Orig Qual 4 | -0.040 | 0.063 | -0.630 |
| Orig Qual other | 0.223 | 0.038 | 5.940 |
| High Qual Academic | 0.030 | 0.043 | 0.710 |
| Age lagged | 0.001 | 0.015 | 0.060 |
| Age$^2$ lagged | 0.000 | 0.000 | -1.260 |
| Not White | -0.080 | 0.096 | -0.840 |
| London | -0.020 | 0.071 | -0.280 |
| South-West | -0.002 | 0.065 | -0.040 |
| East Anglia | 0.049 | 0.090 | 0.550 |
| East Anglia | 0.000 | 0.064 | 0.010 |
| East Midlands | 0.076 | 0.064 | 1.180 |
| North-West | 0.021 | 0.062 | 0.350 |
| North-West | -0.002 | 0.064 | -0.030 |
| Yorks & Humberside | 0.117 | 0.068 | 1.720 |
| North | 0.122 | 0.073 | 1.670 |
| Wales | -0.171 | 0.075 | -2.280 |
| $\Delta$ ln GDP | 0.551 | 3.074 | 0.180 |
| Wave Gap | 0.003 | 0.017 | 0.160 |
| Children Lagged | -0.059 | 0.040 | -1.470 |
| Partner Lagged | 0.068 | 0.044 | 1.560 |
| 1995 | 0.020 | 0.087 | 0.230 |
| 1996 | -0.022 | 0.089 | -0.240 |
| 1997 | -0.054 | 0.093 | -0.580 |
| 1998 | -0.016 | 0.093 | -0.170 |
| 1999 | 0.084 | 0.095 | 0.890 |
| 2000 | 0.070 | 0.088 | 0.790 |
| 2001 | -0.037 | 0.093 | -0.400 |
| 2002 | 0.050 | 0.092 | 0.550 |
| 2003 | 0.118 | 0.092 | 1.280 |
| 2004 | 0.106 | 0.093 | 1.130 |
| 2005 | 0.000 | 0.098 | 0.000 |
| 2007 | 0.106 | 0.608 | 0.170 |
| Employed 1991 | 0.197 | 0.052 | 3.770 |
| Cut 1 | 1.420 | 0.344 | 4.134 |
| Cut 2 | 2.340 | 0.345 | 6.790 |

Table 13: The Decision to Undertake lifelong Learning: Ordered Probit Results