CrossMark

The effect of lifelong learning on men's wages

Richard Dorsett[1] · Silvia Lui[1,2] · Martin Weale[1,3]

Abstract This paper develops a model of earnings and applies this to an examination of the effect of lifelong learning on men's 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 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, estimated by maximum likelihood. Methodologically, the results support the mover–stayer characterisation since the restrictions required for the simpler specifications popular in the literature are rejected. Substantively, simulation of the estimated model shows some statistically significant effects from acquiring qualifications of a higher level than those previously held, but not from acquiring qualifications of the same level.

Keywords Lifelong learning · Returns to education · Switching regression · Wage dynamics

✉ Martin Weale
  mweale@niesr.ac.uk

[1] National Institute of Economic and Social Research, 2, Dean Trench Street, London SW1P 3HE, UK

[2] Rijksuniversitet Groningen, Groningen, The Netherlands

[3] Queen Mary University of London, London, UK

🖄 Springer

JEL Classification C33 · C35 · I20 · J24 · J31 · J63

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 40 % 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 UK, about 30 % 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 % of new undergraduates were aged twenty-five or over; by 2007 this proportion had risen to 43 % (Higher Education Statistics Agency 1995, 2008). Using a much broader definition of lifelong learning, the UNESCO Institute for Lifelong Learning (2009) indicates that in the UK over 50 % of adults aged twenty-six to forty-five report recent participation in some form of adult learning or education, with a participation rate of 41 % for people aged 46–55 and 22 % for people aged fifty-six to sixty-five.

Individuals may participate in lifelong learning for various reasons. One motive may be a desire to progress in the labour market. However, evidence on the effectiveness of lifelong learning is mixed. In the USA, 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 erodes with time. In contrast, Ferrer and Menendez (2014) suggest that, in Canada, graduates who delay their education receive a premium relative to those who do not. Adult learning is particularly common in the UK; in 2004 more than 15 % of thirty to thirty-nine-year-olds were students, a higher level than in any other OECD country (OECD 2009). Here too, though, it is not clear that such learning brings benefits in terms of increased earnings potential. Egerton and Parry (2001) report substantial penalties for late learners, while Jenkins et al. (2002) find little evidence that qualifications gained between the ages of thirty-three and forty-two increase hourly wage growth for men. de Coulon and Vignoles (2008), on the other hand, find a positive wage effect of qualifications acquired between the ages of twenty-six and thirty-four, which varies depending on the level of qualification. Blanden et al. (2012) provide evidence of long-term positive wage impacts of lifelong learning only for women while Evans et al. (2013) discuss the more general social context of lifelong learning.

In this paper, we develop a model of men's wages that allows the evolution of individuals' wages to be influenced by whether they achieved any qualifications through lifelong learning, as well as by a range of background characteristics. We use it to explore the effects of lifelong learning. We base our analysis on the British Household Panel Survey (BHPS), a nationally representative longitudinal survey dataset spanning the period from 1991–2007.

Our modelling approach allows some people to receive a wage that is a random draw from a stationary distribution, while others have a wage that is closely related to that of the previous year. Conceptually, this is a variant of the mover–stayer model (Goodman 1961), and the econometric framework we develop builds upon earlier research applying the model to income dynamics (Dutta et al. 2001). The first group—those whose wage is a random draw—are "movers" in the sense that their position in the wage distribution is (conditionally) unrelated to their previous position. The second group are "stayers" by analogous reasoning and, intuitively, might be interpreted as having a more stable wage trajectory. Essentially, the model allows wages to be estimated using linear regression and whether the individual is a mover or a stayer (which is unobserved and is therefore identified probabilistically). In the case of movers, the regression is in levels, while for the stayers the regression is in differences.

The limited size of the sample means that it is not possible to examine separately all combinations of qualification levels before and after undertaking lifelong learning. We therefore measure initial educational attainment using the broad categories provided by the BHPS. We then distinguish those who undertake lifelong learning without moving to a higher category, from those who upgrade their educational attainment level as a result of lifelong learning. It is, however, possible that the people who gain qualifications differ in some fundamental way from those who do not. We address this by introducing control dummies which indicate people who gain qualifications at some time during the period in which they are observed. Once again, we distinguish qualifications which do not change the level of attainment from those which do.

Both cross-sectional wage equations and wage equations in first differences are nested within our more general model. Our analysis suggests that both of these common specifications should be rejected in favour of our more general model. This result compounds the findings from Dutta et al. (2001) who showed that the mover–stayer structure offered a better means of understanding income inequality in the UK than did other popular specifications.

The results further our understanding of the effectiveness of lifelong learning. Our model is sufficiently flexible to allow identification of 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.

The paper has the following structure. The next section describes our data and the pattern of lifelong learning shown by them. In Sect. 3, we set out our econometric approach. Given the multivariate nature of our model, simulation methods have to be used to show the effects of lifelong learning on earnings. These results are presented in Sect. 4. In Sect. 5, we discuss the relationship between our findings and other related work. Section 6 concludes.

2 Wages and lifelong learning in the British Household Panel Survey

The British Household Panel Survey (BHPS) started in 1991 and ran as an annual survey of each adult member of a nationally representative sample of more than 5000

households (around 10,000 individuals). Among other things, it collected information on employment status, pay, hours worked and educational attainment on a continuing basis. It was a longitudinal survey with the same individuals interviewed in each successive wave. If an individual left the original household, that individual together with all the adult members in their new household would also be interviewed. Children became eligible for interview when they reached the age of sixteen. The sample thus remained representative of the British population as it changes through the 1990s and 2000s. It has now been superseded by the larger survey, Understanding Society.

We focus on data collected from the original sample households over seventeen waves from 1991 to 2007. Members of these households were repeatedly surveyed regardless of changes to household membership. In common with most analyses of wages (see, for example, Dickens 2000; Ramos 2003; Cappellari and Jenkins 2008; Ulrick 2008; Meghir and Pistaferri 2004; Lillard and Willis 1978), we consider only employed men.[1] We limit ourselves to those aged twenty-five to sixty in order to concentrate on working lives beyond completion of the conventional period of education. Thus, for men younger than twenty-five in 1991 or older than sixty in 2007, we consider only the data they provided while in this age range. We drop observations where individuals reported themselves as self-employed because of the difficulties in defining their hourly wages. We also ignore those who provided proxy responses or whose data were incomplete while they were in this age range. Our sample is confined to those who responded in successive waves—where there was a break in response, that individual only features in our estimation sample up to the wave in which that break occurred. Finally, we trim the data to remove the observations whose reported hourly wages fall into the top and bottom 1 % of the distribution.

In our analysis, we define lifelong learning as the acquisition of any qualifications after the age of twenty-five. This age threshold was chosen in order to allow for a period to elapse following the completion of full-time education for most people. We focus on qualification acquisition rather than participation in training since this is more fully recorded in the data but also since this has merit in its own right. We look at the effects of lifelong learning since 1991 or after reaching the age of twenty-five, whichever comes later. The BHPS does not, however, tell us about people who undertook lifelong learning before the first wave of the survey in 1991.

Overall, there are 12,018 useable observations in the data. Table 1 shows how this sample is spread across the different waves of the survey. Due to the need to observe both qualification acquisition and wage change over the previous year, the first wave is dropped from the analysis, resulting in an estimation sample made up of 10,212 observations, relating to 1511 men.

We note some consequences of the sample specification. Because we include only consecutive responses, dropping a single year due to self-employment means that we retain observations on individuals prior to the point of first being self-employed. It may be that individuals choose self-employment because of obstacles to finding work as an employee. Alternatively, it may be that those who change jobs more frequently—and so are more likely to be movers—are also more likely to try self-employment at

---

[1] This is to avoid the complications around female labour supply, where fertility decisions are more important.

some point. In any event, dropping observations from the point of self-employment onwards may result in disproportionately discarding movers.[2] As another possibility, if our treatment of outlier wages results in dropping individuals who in fact do have large wage variation, the sample will end up under-stating the degree of earnings mobility.

As Table 1 makes clear, the data set from which we estimated our model is subject to attrition; we begin with 1806 observations and in the final wave have 211 observations. To explore the nature of attrition, we used the test described by Fitzgerald et al. (1998). We regressed initial log hourly wages on a vector of observed characteristics and also a dummy variable indicating whether someone dropped out of the panel at some point. The coefficient on this dummy was significant ($t = 3.52$), indicating that observed variables did not fully account for the link between response and the wage rate. To proceed, we constructed a variable to control for selection on unobservables. We estimated a probit model of dropout in each wave and used the generalised residual from this as a control function when estimating our model.[3] Details are provided in "Appendix 1".

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 postgraduate 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. Table 2 shows this classification. In common with other work (e.g. de Coulon and Vignoles 2008), we have treated all GCSEs[4] as being in category 1.

Table 2 also indicates the number of people gaining qualifications in our sample. These data relate to the 1511 men of our sample but cover only those qualifications gained from wave 2 onwards. While they show over seven hundred qualifications gained, a substantial proportion gained more than one qualification. Thus 1131 men did not report any qualifications, 204 gained one qualification, 82 reported two qualifications, and 94 reported three or more qualifications over the period they were observed. Much the largest category of qualifications gained is "other". However, Table 2 also shows considerable importance of City and Guilds qualifications. Sub-

---

[2] Across all waves, 10.5 % of the sample are dropped due to being self-employed, 13.4 % of workers. To provide some sense of how this might affect our results, we re-estimated the econometric model described later, excluding individuals who were self-employed at any point. The resulting estimates of lifelong learning are similar to those found when not excluding those self-employed at any point.

[3] Results available on request showed that our findings are robust to assuming that attrition is random.

[4] The General Certificate of Secondary Education (GCSE) is normally taken by children at the age of sixteen, while AS-levels are taken at age seventeen and A-levels at age eighteen. Two A-levels are the minimum qualification required for study at university although in practice most universities require three A-levels. Scotland has its own system of qualifications; these have been converted into the equivalents from the rest of the UK.

Table 1 Derivation of the sample

Wave	No. consecutive responses	Proportion	Cumulative proportion
1	1806	15.02	15.02
2	1511	12.57	27.59
3	1216	10.11	37.71
4	1040	8.65	46.36
5	902	7.5	53.86
6	799	6.65	60.51
7	722	6.01	66.51
8	642	5.34	71.85
9	586	4.87	76.73
10	529	4.4	81.13
11	466	3.88	85
12	406	3.38	88.38
13	353	2.94	91.32
14	310	2.59	93.9
15	276	2.3	96.2
16	243	2.02	98.22
17	211	1.78	100
Total	12,018	100	

degree higher education qualifications (HNC/HND or university diploma) are more common than university degrees, while not many respondents report gaining GCSEs or A-levels. Acquisition of qualifications does not, of course, mean that someone's educational attainment, as represented by the level of their highest qualification, increases. As noted in the introduction, we distinguish acquisition of qualifications without any increase in educational attainment from upgrading of educational attainment in our subsequent analysis. Those who gain "other" qualifications are treated as not having upgraded their educational attainment; thus people initially with only minimal qualifications remain at Level 0 even after gaining an "other" qualification.

Table 3 provides a summary picture of the extent of lifelong learning in terms of the qualification levels shown in Table 2. The data here relate only to the 902 men for whom we have observations for five years or more; this provides a picture of the incidence 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 first row below the transition table shows the probability of upgrading to be fairly evenly spread across qualifications levels (the somewhat smaller rate for those with level 2 qualifications is based on a small sample size). We note that those with level 5 qualifications cannot upgrade, by definition. A very different impression is formed when considering the incidence of lifelong learning without upgrading qualifications. Here there is a clear gradient. Among those with no qualifications,

Table 2 The classification of qualifications and the number gained as a result of lifelong learning in our sample

Classification	Qualification	No. gained
Level 1	Youth training cert.	2
	Trade apprenticeship	2
	Clerical or commercial	14
	City and guilds Pt i	56
	NVQ/SVQ level 1	11
	Any GCSEs	9
Level 2	City and Guilds Pt ii	28
	NVQ/SVQ Level 2	15
	Any GNVQ	13
	Any AS or A-levels	3
Level 3	City and Guilds Pt iii	14
	ONC, OND, BEC Gen Cert.	12
	NVQ/SVQ Level 3	9
	Any Higher grades	3
Level 4	HNC, HND, BEC higher cert.	13
	Nursing SEN, SRN, SCM	3
	Teaching	10
	NVQ/SVQ level 4	8
Level 5	University diploma	26
	Univ/CNAA first degree	12
	Univ/CNAA higher degree	16
Unclassified		489
Total		768

only 5 % will undertake some learning. This compares with 24 % for those with level 5 qualifications.

One might expect that those with the lowest initial qualifications would have the most to gain from lifelong learning and that therefore the prevalence would be highest for those educated to levels 0 or 1. There are a number of reasons why this might not be the case but it is possible only to speculate about them. Perhaps most obviously, people who are already reasonably well-educated may be better aware of the opportunities available to them than those who are poorly educated. But it is also possible that less-educated people may have difficulties in managing the costs associated with gaining qualifications[5] or that they may believe that their capacity to benefit from further qualifications is limited. Further, these figures show only the proportions of people who have actually gained qualifications and not those who have embarked on courses, but not completed them. For obvious reasons, surveys do not ask people about

---

[5] Costs of courses vary very greatly, so it is not possible to draw any generalisations.

Table 3 Transition probabilities over a five-year window and the incidence of lifelong learning

		Initial qualification level						
		0	1	2	3	4	5	All
Percentages								
Qualification	0	95.4	–	–	–	–	–	20.5
Level	1	3.1	96.5	–	–	–	–	31.0
Five	2	0.0	0.7	98.6	–	–	–	8.1
Years	3	1.6	1.1	0.0	95.2	–	–	16.0
Later	4	0.0	0.7	0.0	1.4	94.9	–	8.8
	5	0.0	1.1	1.4	3.5	5.1	100.0	15.6
Upgrading		4.6	3.5	1.4	4.8	5.1	–	3.4
Lifelong learning		5.2	7.8	16.7	16.6	20.3	24.2	12.8
N		194	284	72	145	79	128	902

qualifications they have worked towards but failed to obtain. Finally, if the return to acquiring a particular qualification is proportional to initial earnings, then the incentives to undertake lifelong learning may be higher for those with more than minimal initial qualifications.

There is a risk that people who embark on lifelong learning may drop out of the BHPS. That might seem a substantial risk if most lifelong learning involved moving away from home to attend a college or university. But Table 2 suggests that only about 8 % of lifelong learning qualifications are university qualifications and data from the Higher Education Statistics Authority indicate that in the academic year 1999/2000 86 % of first-year students aged twenty-five and over were part-time students. Such students will not have the same reasons as full-time students to move away to go to university and are therefore much less likely to be lost from the survey. Thus, while it cannot be established definitively, it seems unlikely that the participation incidences shown in Table 3 are importantly affected by attrition. In any case, as mentioned earlier, our results are corrected for attrition due to unobserved causes.

2.2 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.[6] For all waves, we derive each employee's gross hourly wage as follows:

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

---

[6] This is a derived variable wPAYGU.

Table 4 Summary data: initial qualifications, wages and lifelong learning, 1996–2008 average

Initial education level	No lifelong learning	With qualification no upgrading	With upgrading	Total
Observations				
0	1795	204	158	2157
1	2606	477	203	3286
2	632	214	49	895
3	1211	413	63	1687
4	609	240	40	889
5	923	375		1298
All wages	7776	1923	513	10,212
0	£8.00	£9.77	£10.06	
1	£10.29	£11.09	£11.74	
2	£10.70	£10.96	£15.12	
3	£13.23	£13.46	£14.98	
4	£15.07	£14.12	£18.80	
5	£17.60	£18.41		

Pooled data 2007 prices

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 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. Lifelong learning with no qualification upgrade is associated with modestly higher wages, for all except those with qualifications at level 4. Where qualifications are upgraded as a result of lifelong learning, the apparent premium is larger, particularly for those initially with level 2 qualifications or higher.

These statistics suggest a connection between lifelong learning and earnings. But to understand whether there is indeed a return to lifelong learning, a full econometric analysis is necessary.

3 Econometric analysis

In this section, we discuss in more detail the mover–stayer model, describe the econometric approach and present estimation results.

3.1 A mover–stayer framework

The original mover–stayer model was described by Goodman (1961). In our model we describe as movers people who may receive a wage rate possibly very different from what they had earned in the previous period—they move about the wage distribution.[7] The wages of the stayers are, by contrast, closely explained by their previous wage rates. It is, however, not possible to observe whether someone is either a mover or a stayer in any period. The most one can do is infer a probability of being in one category or the other.

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. But they may also be people who have been in stagnant jobs with little prospect for progression who have the good fortune to come across more favourable labour market opportunities. Or people who have done reasonably well but still find that a better opportunity has come along. Being a mover need not be associated with a change of employer. It is perfectly possible that people will move from one post to another offering sharply better pay with 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 than 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 in the first difference of log wages.

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. The estimated model allows us to determine the probability that particular observations are those of stayers rather than movers or *vice versa*. 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); it offers a means of dealing with heterogeneity in the data.

The model encompasses the standard first differences model if all hourly wages can be explained by the stayers' equation and an equation in terms of levels if the probability of being a mover is one. In Sect. 5, we present estimates of the number of years that someone should expect to be a stayer. The model structure is such that this should be expected to depend on observable characteristics. If everyone were a mover,

---

[7] The original mover–stayer model (Goodman 1961) considered a population on which categorical data were observed. Some members, movers, were subject to a Markov process, while others, stayers, retained their initial category.

the number of years expected as a stayer would be little different from zero. On the other hand, if everyone were a stayer, this number would simply be the remainder of an individual's working life. The fact that within our 95 % confidence limits neither of these is the case supports our mover–stayer specification relative to either of these simplifications. In Sect. 4 we compare this model against pooled OLS models in levels and differences and also against a fixed effects panel model.

We now set out the components of the mover–stayer model. The choice of explanatory variables is discussed subsequently in Sect. 3.6.

3.2 Movers

For movers, wages are given by a stationary Mincerian equation

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

where y_{it} represents log hourly wages deflated by the retail price index and X_{it} is a vector of variables which influence the wage rate. Such variables include age, qualifications, lifelong learning, region of residence, log real GDP *per capita* and a measure of local unemployment. They also include the generalised residual of the probit equation to control for attrition bias ("Appendix 1"). Thus, for a mover, the wage rate is not directly related to previous wages except insofar as the variables which influence the wage of a mover have also influenced their wage on the previous occasion when they were a mover.

3.3 Stayers

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

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

It should be noted that there is no loss of generality in specifying the vector of driving variables X_{it} to be the same in both equations; provided it is general enough, differences in specification can be accommodated by restrictions on the elements of β_1 and β_2.

3.4 Switching

A respondent is a mover if the indicator variable $I_{it} = 0$ and a stayer if $I_{it} = 1$. This indicator is driven by the latent variable, I_{it}^*. The probability, P_{it} that observation y_{it} is drawn from (3) rather than (2) is driven by the latent variable,

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

with $I_{it} = 0$ if $I_{it}^* \leq 0$ and $I_{it} = 1$ if $I_{it}^* > 0$. As already noted, the indicator variable is not observed. It is possible, through the application of Bayes' theorem, to infer the probability that $I_{it} = 0$ or $I_{it} = 1$ using the density functions set out in Sect. 3.5, but since we do not make any use of such an analysis we do not pursue the matter.

3.5 Estimation strategy

The model has the following likelihood function:

$$
L_{it} = \prod_{I_{it} \in 0,1} \left\{ F\left(\varepsilon_{it} > -Z_{it}\gamma\right) f\left(u_{2it} \mid \varepsilon_{it} > -Z_{it}\gamma\right) \right.
$$

$$
\left. + F\left(\varepsilon_{it} \leq -Z_{it}\gamma\right) f\left(u_{1it} \mid \varepsilon_{it} \leq -Z_{it}\gamma\right) \right\}
$$

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

$$
\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{1\varepsilon} \\ & \sigma_2^2 & \sigma_{2\varepsilon} \\ & & 1 \end{bmatrix} \tag{5}
$$

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

Consider the case of $I_{it} = 0$. The truncated normal density is

$$
f\left(u_{1it}, \varepsilon_{it} \mid \varepsilon_{it} \leq -Z_{it}\gamma\right) = \frac{f(u_{1it}, \varepsilon_{it})}{\Phi(Z_{it}\gamma)}
$$

$$
= \frac{f(u_{1it}) f(\varepsilon_{it} \mid u_{1it})}{\Phi(Z_{it}\gamma)} \tag{6}
$$

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

$$
f\left(u_{1it} \mid \varepsilon_{it} \leq -Z_{it}\gamma\right) = \frac{f(u_{1it}) \int_0^{-Z_{it}\gamma} f(\varepsilon_{it} \mid u_{1it}) \, d\varepsilon_{it}}{\Phi(-Z_{it}\gamma)} \tag{7}
$$

noting that

$$
f(\varepsilon_{it} \mid u_{1it}) \sim N\left(\left(\frac{\rho_{1\varepsilon}}{\sigma_1}(y_{it} - X_{it}\beta_1)\right), \left(1 - \rho_{1\varepsilon}^2\right)\right) \tag{8}
$$

where $\rho_{1\varepsilon} = \frac{\sigma_{1\varepsilon}}{\sigma_1}$. We can then write

$$
f\left(u_{1it} \mid \varepsilon_{it} \leq -Z_{it}\gamma\right) = \frac{\Phi\left(-\frac{Z_{it}\gamma + \frac{\rho_{1\varepsilon}}{\sigma_1}(y_{it} - X_{it}\beta_1)}{\sqrt{1 - \rho_{1\varepsilon}^2}}\right) \phi\left(\frac{y_{it} - X_{it}\beta_1}{\sigma_1}\right) / \sigma_1}{\Phi(-Z_{it}\gamma)} \tag{9}
$$

Similarly, the case of $I_{it} = 1$ results in

$$f\left(u_{2it} \mid \varepsilon_{it} > -Z_{it}\gamma\right) = \frac{\Phi\left(\frac{Z_{it}\gamma + \frac{\rho_{2\varepsilon}}{\sigma_2}\left(\Delta y_{it} - X_{it}\beta_2\right)}{\sqrt{1-\rho_{2\varepsilon}^2}}\right)\phi\left(\frac{\Delta y_{it} - X_{it}\beta_2}{\sigma_2}\right)/\sigma_2}{\Phi\left(Z_{it}\gamma\right)} \qquad (10)$$

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

$$L_{it} = \prod_{I_{it} \in 1,2, J_{it}=1} \left\{ \Phi\left(-\frac{Z_{it}\gamma + \frac{\rho_{1\varepsilon}}{\sigma_1}\left(y_{it} - X_{it}\beta_1\right)}{\sqrt{1-\rho_{1\varepsilon}^2}}\right)\phi\left(\frac{y_{it} - X_{it}\beta_1}{\sigma_1}\right)/\sigma_1 \right.$$
$$\left. + \Phi\left(\frac{Z_{it}\gamma + \frac{\rho_{2\varepsilon}}{\sigma_2}\left(\Delta y_{it} - X_{it}\beta_2\right)}{\sqrt{1-\rho_{2\varepsilon}^2}}\right)\phi\left(\frac{\Delta y_{it} - X_{it}\beta_2}{\sigma_2}\right)/\sigma_2 \right\} \qquad (11)$$

Equation (11) shows that the likelihood function is, for each observation, a weighted average of the contributions to the likelihood which would arise with pooled equations in levels and differences, respectively. The weights, however, depend on unobserved characteristics. The model is estimated using maximum likelihood 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, therefore, we maximise a log pseudolikelihood.

3.6 Variables used in the analysis

The main variables of interest are those that relate to lifelong learning. Among those who acquire new qualifications, we distinguish between those whose highest level of qualification is increased as a result (that is, they upgrade) and those whose highest level is left unchanged (not upgraded). We define dummy variables *Gains Qualification: No Upgrade* and *Gains Qualification: Upgrade* accordingly. These take the value 1 from the wave in which the qualification is acquired onwards. Someone who gains a qualification without upgrading and then subsequently gains one with upgrading will be indicated by the first dummy until their upgrade, when they are indicated by the second dummy. Someone who upgrades and then gains a further qualification retains the dummy which results from their initial upgrading.

Beyond this, theory has little to say about what might be included as explanatory variables in Eqs. (2)–(4) and our strategy is therefore to include variables to control for sources of variation within our sample which may be correlated with gaining qualifications as a result of lifelong learning. All equations include the following variables: qualification level when first observed; a dummy variable indicating whether

the highest qualification at that time was academic[8] or not; age; whether from an ethnic minority group or not; partnership status (couple vs. single adult household), the presence of children (represented by a 0/1 dummy variable); region (using dummies to indicate the region within Britain people live in); GDP or its change as an indicator of the state of the economy; and local unemployment relative to the national rate. The switching equation 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. Intuitively, people are more likely to be movers if the gap between interviews is long than if it is short and those with a recent job change are more likely to have experienced a wages shock that would be likely to classify them as movers. These variables are excluded from the movers' and stayers' equations.

The effect of rising overall prosperity is controlled for by including the growth rate of GDP in Eq. (3), the wage equation for stayers. 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; we use GDP growth to represent this. By contrast, we expect the wage rate of movers 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. Both the level and the change in log GDP are included in the switching Eq. (4). We also include a variable *Regional Unemployment deviation*. This captures the extent to which the local unemployment rate differs from the national average, and so provides a measure of the relative strength of the local economy.

The variables mentioned in this subsection are either exogenous (age, ethnic group, wave of survey) or relate to an earlier time period in order to reduce concerns about endogeneity.

Lastly, we include two types of variables in an attempt to control for selection effects. *Generalised Residual,* the generalised residual of the probit equation which explains attrition (see "Appendix 1"), is included to control for sample selection. To control for selection into learning (the possibility that individuals who participate in lifelong learning might differ in some way from those who do not), we include indicators of whether individuals obtain qualifications at some time during the period for which we have data. *Qualifies Sometime: No Upgrade* and *Qualifies Sometime: Upgrade* indicate, respectively, qualification acquisition at some point, with and without upgrade. These are exclusive; someone who first gains a qualification without upgrading and then a further qualification with upgrading will be indicated only by the *Qualifies Sometime: Upgrade* dummy.

We do not explicitly consider men with multiple jobs and how these might alter our conclusions. In fact, multiple employment is quite rare; when first observed, only 8 % of those in work held more than one job. Interestingly, holding a second job was more common among the more highly qualified (13 % among those with level 5 qualifications, compared to 6 % among those with no qualifications).

---

[8] An academic qualification is one which is normally taken in a school or university. Thus academic qualifications are GCSEs, AS and A-levels, and their Scottish equivalents or university degrees and diplomas.

3.7 Parameters of the mover–stayer model

We explored the system of equations in two forms. The unrestricted equation includes all four dummy variables associated with gaining qualifications in all three equations. We find, however, that they are highly insignificant in the stayers' equation; the Wald test statistic does not reject the restriction of setting these effects to be zero $\chi_4^2 = 2.36.(p = 0.67)$. We therefore present results both for the unrestricted model and a restricted model in which these terms are set to zero. The parameters of both equations are presented in Table 5, and it is clear that the restriction has very little influence on the other parameters in the model. Since the restrictions are so easily accepted, our subsequent discussion is limited to the restricted model shown in columns four to six of Table 5. The Wald test for *Generalised Residual* ($\chi_3^2 = 4.6$, $p = 0.21$) suggests that we can reject the hypothesis that significant sample selection effects are present in the model and the variable is plainly not significant in any of the three equations. Similarly, the control dummies (*Qualifies sometime: no upgrade)* and (*Qualifies sometime: upgrade)* are not significant in any equation, suggesting selection into lifelong learning may not significantly bias results.

The equation for movers shows that a qualification gained without upgrading enhances the wage of a mover by 0.11 log units, while upgrading of education status raises the wage by 0.17 log units. Both are significant at a 5 % level. Thus we see clear effects of lifelong learning.

The dummies for the initial level of educational attainment show a pattern broadly commensurate with other studies. Dickson (2013) suggests an effect of about 10 % per year of study. We show a clear effect of level 1 qualifications, notwithstanding that the people who gain these usually have studied for no longer than people who have no qualifications. The coefficient for level 5 education, 0.48, is commensurate with the idea that a degree is achieved at age twenty-one, while someone with no qualifications will probably have left school at sixteen. The picture is, however, complicated by two factors, which offset each other to some extent. People with no qualifications, by definition, do not have academic qualifications. On the other hand, all level 5 qualifications are classified as academic. This incurs a penalty (albeit not significant) of 0.075 log units shown in the movers' equation but delivers a faster rate of income growth to people when they are not movers. Moreover, adding the relevant educational dummy to the academic dummy, the switching equation implies that people with academic qualifications at level 5 are appreciably more likely to be stayers than are those with lower levels of qualifications. Thus simulation techniques, of the type which we use to establish the benefits of lifelong learning, would be needed to establish the returns to the different levels of educational attainment. The equation shows that the movers' wages are increasing in real GDP, but the elasticity is surprisingly low.

The results for the stayers' equation suggest few identifiable influences on the rate of growth of wages. In particular, lifelong learning does not appear to increase wages for stayers. People with academic qualifications can look forward to a growth rate over 1 % faster than those who do not have such qualifications, but otherwise nothing is significant at a 5 % level. While it might be possible to impose further zero restrictions on this equation, so as to form a more clearly specified notion of what influences the wage growth of stayers, that is outside the scope of this paper.

The switching equation is defined so that the larger the value of the latent variable, the more likely someone is to be a stayer rather than a mover. Gaining lifelong qualifications makes it significantly more likely that someone will be a mover. The probability of being a stayer is positively related to the initial level of educational attainment, and those with academic qualifications are more likely to be stayers. Not surprisingly, we can see that a recent job is more likely to be associated with a move in the wage distribution. More broadly, the fact that the probability of being a stayer increases with age and educational status is consistent with the idea that qualifications help people find good job matches and that so too does the passage of time.

While these coefficients show significant effects from lifelong learning, we should not rush to conclude that there are significant effects on discounted incomes. Men have to "move" in order to realise the benefits of their lifelong learning qualifications, and without some sense of the frequency with which this happens, it is not possible to say whether the effects shown in Eq. (2) will translate into effects on wages with similar levels of significance. Obviously the uncertainty present in the other coefficients of the model will influence this. Thus, to establish whether lifelong learning has significant effects on wages, it is necessary to resort to simulation with successive draws of parameter values made with reference to the covariance structure of the parameter set as a whole. We present such simulations for both the restricted and unrestricted models in the subsequent section.

A separate question arises about the generality of the switching model relative to the simpler alternatives, either that everyone is a mover with hourly earnings fixed by an equation in levels, or everyone is a stayer with hourly earnings explained by an equation in differences. The former is the case if the latent variables generated by the coefficients of Eq. (4) are large and negative, while the latter is the case if the coefficients are large and positive. The fact that some of the coefficients are themselves statistically significant does not answer this question. We therefore defer it to Sect. 5 where we simulate our model using repeated draws of coefficients from the distribution behind those of Table 5. The simulated values of the latent variables address this question directly.

Table 5 also shows the variance–covariance structure of the system. The standard error of the movers' equation, σ_1, is 0.32, while that of the stayers' equation, σ_2, is 0.13, reflecting the underlying structure of the model that the wage conditional on being a stayer is much less variable than the wage conditional on being a mover. The correlations, $\rho_{1\varepsilon}$ and $\rho_{2\varepsilon}$, relate to the correlations between the residual of the equation in question and that of the switching equation (Eq. 4). Hence, the tendency to be a mover (conditional on observed characteristics) is correlated with a tendency to have lower wages. Conversely, the tendency to be a stayer is correlated with a higher rate of wage growth. Together, these findings imply that stayers are likely to earn higher wages.

4 Comparison with other models

There are three models with which it is worth comparing our results. As we noted in Sect. 3.5, the likelihood function is a weighted combination of the contributions that

Table 5 Parameters of the mover–stayer model

		Unrestricted			Restricted		
		Mover	Stayer	Switching	Mover	Stayer	Switching
Gains Qualification	No upgrade	0.108	-0.001	-0.272	0.107		-0.273
		(0.045)	(0.004)	(0.124)	(0.043)		(0.124)
	Upgrade	0.166	0	-0.125	0.165		-0.126
		(0.065)	(0.010)	(0.196)	(0.063)		(0.194)
Initial Qualification	Level 1	0.215	-0.004	0.058	0.215	-0.004	0.058
		(0.052)	(0.005)	(0.128)	(0.052)	(0.005)	(0.128)
	Level 2	0.306	-0.005	0.145	0.307	-0.006	0.143
		(0.045)	(0.005)	(0.139)	(0.044)	(0.005)	(0.139)
	Level 3	0.337	-0.001	0.512	0.338	-0.002	0.512
		(0.055)	(0.005)	(0.138)	(0.054)	(0.005)	(0.138)
	Level 4	0.404	0.007	1.057	0.404	0.007	1.056
		(0.124)	(0.005)	(0.193)	(0.123)	(0.005)	(0.193)
	Level 5	0.482	0.003	0.975	0.483	0.002	0.971
		(0.180)	(0.006)	(0.256)	(0.179)	(0.006)	(0.255)
	Other	0.053	0.004	0.092	0.054	0.004	0.091
		(0.046)	(0.003)	(0.117)	(0.046)	(0.003)	(0.117)
Academic qualification		-0.075	0.013	0.398	-0.075	0.013	0.399
		(0.052)	(0.004)	(0.121)	(0.052)	(0.004)	(0.121)
Age		0.016	-0.001	0.078	0.016	-0.001	0.078
		(0.015)	(0.002)	(0.041)	(0.015)	(0.002)	(0.041)
$Age^2/100$		-0.017	0.001	-0.073	-0.017	0.001	-0.074
		(0.017)	(0.002)	(0.047)	(0.017)	(0.002)	(0.047)
Public Sector		0.059	-0.005		0.059	-0.005	
		(0.039)	(0.003)		(0.038)	(0.003)	
Not White		0.081	-0.002	-0.156	0.081	-0.002	-0.157
		(0.109)	(0.009)	(0.284)	(0.108)	(0.009)	(0.284)
ln GDP			-0.019	-2.71		-0.02	-2.69
			(0.142)	(2.265)		(0.142)	(2.260)
Δ ln GDP		0.157		0.74	0.158		0.742
		(0.141)		(0.387)	(0.141)		(0.386)
Regional unemp. deviation		0	0	0.023	-0.001	0	0.024
		(0.009)	(0.001)	(0.023)	(0.009)	(0.001)	(0.023)
Wave Gap				-0.009			-0.01
				(0.021)			(0.020)
Recent Job				-0.668			-0.669
				(0.071)			(0.071)
Generalised Residual		0.035	0.027	0.692	0.033	0.027	0.69
		(0.210)	(0.023)	(0.453)	(0.209)	(0.022)	(0.452)
Qualifies sometime	No upgrade	-0.023	0	0.07	-0.023		0.072
		(0.040)	(0.004)	(0.109)	(0.040)		(0.108)
	Upgrade	0.054	0.007	-0.106	0.062		-0.124
		(0.054)	(0.008)	(0.183)	(0.054)		(0.180)
	σ_i	0.322	0.132		0.322	0.131	
		(0.013)	(0.003)		(0.013)	(0.003)	
	$\rho_{i\varepsilon}$	-0.290	0.191		-0.291	0.196	
		(0.118)	(0.061)		(0.118)	(0.061)	
N				10212			10212
Log-likelihood				3556.6			3555.9

Regional dummies and constant terms are also included. Standard errors are clustered.

Table 6 Likelihoods of alternative models

Model	Log-likelihood	Deg. of freedom k	No. data N_1	No. clusters N_2	BIC$_1$	BIC$_2$
Levels	−3917	29	10212	1511	8100	8045
Differences	2622	29	10212	1511	−4976	−5032
Fixed effects panel	5088	1535	10212	1511	3994	1061
Mover–stayer Model	3556	93	10212	1511	−6255	−6432

$BIC_i = -2 Log\ likelihood + k \log(N_i).\quad i = 1, 2$

would be made, observation by observation, to pooled models in log differences and in log levels. Since the weights are designed to accommodate the data, this procedure should be expected to result in a higher likelihood than would be found if the model were estimated in levels or in differences alone.[9] The third relevant model is a fixed effects panel model in log levels. Table 6 shows that, comparing the log likelihoods, the mover–stayer model outperforms the pooled models in levels and differences, but it itself outperformed by the fixed effects panel model. Some adjustment is, however, needed for the number of degrees of freedom. The Bayesian Information Criterion (BIC) offers a means of taking account of this. It is, however, not clear how many observations there are with panel data. If all the observations in each cluster coincided, then the number of observations would be the number of clusters rather than the number of data points. Stata Manual (2014) suggests that using the number of data points, N_1, in the BIC calculation is the least favourable approach as far as the fixed effects model is concerned, while using the number of clusters, N_2, offers the most favourable approach, but that all models should be compared using the same number of "observations". Table 6 shows that on either basis the mover–stayer model is preferred to the other three models.

We also show in Table 11 of "Appendix 3" the results of estimating the fixed effects panel model. This shows significant effects when upgrading to level 1 or to levels 3, 4 or 5. The restricted nature of the dynamics means, however, that the model has nothing to say on whether the effects of upgrading might become attenuated with age. The effects of age in the switching equation of our model and the role that switching plays in realising the benefits of increased educational attainment mean that our model can address this.

5 Results: Returns to lifelong learning

In Table 7, we show the returns to lifelong learning generated by the restricted model from Table 5. Table 7 shows the percentage increase in discounted expected earnings from the age at which the qualification is acquired (either thirty or forty-five) to age sixty; a discount rate of 2 % per annum is used. The results are generated by

---

[9] It should be noted that the model in differences is not simply the model in levels. In the former, variables explain growth in wage rates, while, in the latter, they explain the level of wage rates.

repeated simulation of the experiences of a panel of 10,000 men over the life course. One thousand simulations were carried out in order to provide the indicators of the distribution of the expected return. For each simulation, the parameters of the model were drawn randomly from a multivariate normal distribution whose means are those shown in the tables and whose variance is given by the estimated covariance matrix of the parameters. In these random draws, there is a risk that the resulting covariance matrix of the shocks to the three equations of the model is not positive definite. Draws with this property were replaced by new draws. The table also shows, in the final column for each case, the proportion of simulations in which the discounted expected earnings declined following the acquisition of qualifications.

The model is nonlinear, and the effects therefore depend not only on the initial level of education, but also on whether the highest qualification of the men in question is academic and on where they live.

Results are presented for men showing the returns as functions of their initial level of education, the age at which they gain lifelong qualifications and whether they upgrade their qualification level or not. The broad pattern is that the impacts of lifelong learning are lower than the relevant coefficients in Table 5 suggest. The reason for this is that our model suggests men have to "move" in order to realise the benefits of lifelong learning, and the probabilities of such moves are fairly low. As a check, we simulated the model with the probability of a move set to 1, and, as expected, the simulations which result showed the returns implied by the parameters of the *Gains Qualification* terms in the movers' equation.

The estimates of the effects of gaining qualifications do not differ greatly between the unrestricted model ("Appendix 2", Table 10) and the restricted model. Thus a man aged thirty who is initially educated to level 2 and who upgrades his qualification level, is estimated to gain 12.0 % in pay when the unrestricted model is used, and 12.9 % when the restricted model is used. At age forty-five the effects are 6.8 and 5.2 %. With the restricted model, we can see that, for men aged thirty whose highest qualification is not academic and who upgrade their qualifications, there is a significant return at a 95 % significance level[10] for those educated at up to level 3 before gaining their qualification. For those educated at level 4, the return is significant at 90 %. For those with academic initial qualifications, however, the returns are lower and are not statistically significant. If the qualification is not gained until age forty-five, none of the returns is significant at a 95 % level and only for men initially educated to level 1 or 2 with a non-academic qualification is the return significant at the 90 % level.

These differences can be understood from the fact that in order to benefit from gaining qualifications men have to experience the sort of shock which leads to them having their salary given by the "movers" equation; it is here that significant effects on wages from lifelong learning are found. It follows that men who have a long expected time as "stayers" are less likely to benefit significantly from their additional

---

[10] The confidence intervals are calculated from the simulations; we do not make the assumption that the returns are normally distributed. The lower limit of the 95 % confidence interval of the estimate of the return is given by the 2.5 percentile of the ranked returns. We show in Table 7 the proportion of simulations which result in a reduction in the discounted wage and, when this is more than 2.5 %, the estimate is not significant at a 95 % level. When it is more than 5 % the estimates are not significant at a 90 % level.

Table 7 The Returns to lifelong learning for men acquiring qualifications at ages 30 and 45

	Non-academic			Academic		
	Mean (%)	SD	% < 0 (%)	Mean (%)	SD	% < 0 (%)
Age 30						
No upgrade						
0	4.0	0.044	17.9	−0.3	0.044	49.5
1	4.9	0.042	13.2	0.9	0.038	40.0
2	5.2	0.041	10.7	1.5	0.036	31.4
3	3.9	0.035	12.2	0.9	0.026	33.7
4	1.6	0.021	18.4	0.1	0.012	39.0
5	2.6	0.026	12.6	0.6	0.013	30.1
Upgrade						
0	12.8	0.069	2.5	8.4	0.070	11.4
1	13.0	0.067	2.0	8.6	0.062	7.8
2	12.9	0.065	1.4	8.4	0.058	5.6
3	9.9	0.055	1.6	5.4	0.042	6.7
4	4.9	0.036	3.0	1.9	0.022	9.5
Age 45						
No upgrade						
0	1.1	0.031	35.0	−2.3	0.026	84.2
1	1.8	0.030	25.6	−1.6	0.022	77.5
2	2.0	0.027	21.3	−1.1	0.020	71.5
3	0.4	0.019	37.1	−1.0	0.012	84.5
4	−0.4	0.009	68.3	−0.5	0.006	92.6
5	0.0	0.011	50.2	−0.4	0.006	87.4
Upgrade						
0	7.0	0.052	7.5	2.1	0.047	31.0
1	7.1	0.049	4.8	2.3	0.039	25.0
2	6.8	0.047	4.2	2.3	0.036	23.9
3	3.5	0.032	9.2	0.7	0.021	35.5
4	0.7	0.015	25.1	0.1	0.011	52.1

qualifications than are those who are early movers. Moving is a stochastic phenomenon, but the switching equation allows us to work out the probability that someone is a stayer and thus, in our simulated panel, the expected time that they wait before a move. The effects of age on the probability of staying are clearly positive, suggesting that men are likely to have to wait increasingly long for a move as they age. Similarly, the higher is the qualification level, the more likely it is that men will be stayers. Furthermore, men with academic qualifications are more likely to be stayers than men with non-academic qualifications. Table 8 shows the expected time as a stayer, i.e. before a move, for men initially aged thirty and forty-five with different types of qualifications; the effects described are clearly visible. This is calculated from the probability of being a stayer at each age, conditional on not having been a stayer earlier.

Table 8 Expected number of years as a "stayer"

	Non-academic		Academic	
	Mean	SD	Mean	SD
Age 30				
No upgrade				
0	4.0	1.6	9.1	3.3
1	4.5	2.0	9.9	2.9
2	5.4	2.1	11.6	3.5
3	11.0	3.6	18.8	3.6
4	21.3	4.0	26.5	2.7
5	19.7	4.9	25.7	2.9
Upgrade				
0	3.8	1.9	8.7	3.9
1	4.4	2.4	9.5	3.9
2	5.2	2.6	11.1	4.5
3	10.5	4.4	17.9	4.8
4	20.5	4.8	25.9	3.4
Age 45				
No upgrade				
0	7.5	1.8	10.9	1.8
1	8.0	1.9	11.4	1.6
2	8.7	1.9	12.0	1.7
3	11.8	1.6	14.1	1.0
4	14.6	0.9	15.5	0.5
5	14.2	1.3	15.4	0.5
Upgrade				
0	7.1	2.0	10.5	2.1
1	7.6	2.2	11.0	1.9
2	8.3	2.3	11.6	2.1
3	11.4	2.0	13.8	1.3
4	14.4	1.1	15.4	0.6

The reason for this difference is easy to understand by looking at the coefficients of the switching equation in Table 5. The effect of gaining qualifications on the growth rate of stayers' earnings is to add in very considerable uncertainty around a zero, or in the case of upgrading, small positive average value. The implication of this is that gaining a qualification makes the growth rate of earnings much more uncertain than it was in the absence of the qualification and, as a result, the effect on the discounted future wage is much less certain. The zero restriction rules out this increased uncertainty.

A more general issue is how our estimates of returns to learning compare with those of other researchers. Blanden et al. (2012) failed to find a return to lifelong learning

for men. However, de Coulon and Vignoles (2008), working with the British Cohort Study, found effects of qualification acquisition on the wages of men aged twenty-six to thirty-four ranging from 10 to 30 % depending on the qualification gained but with the lowest returns for level 3 qualifications. They did not distinguish men who upgraded their qualification levels from those who did not do so.[11] While they did not investigate the relationship between the returns to qualifications and initial qualification level, they did suggest that the effect of gaining an NVQ2 level qualification was higher for people with low ability than for the population as a whole, a finding consistent in broad terms with our own results. More generally, the effects shown are rather more powerful than our own. They therefore do not suggest that our results are implausibly large.

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 around a switching regression. This model distinguishes two wage processes; some men receive wages close to those received in the previous year, while others receive a wage which is related to their educational attainment, age and other characteristics but which is not directly related to their previous earnings. The switching equation determines the probability that each of these processes is relevant.

We find that raising educational attainment affects directly the wages of those whose earnings are not directly related to those of the previous period with the consequence that the benefit of increased attainment is not gained until someone experiences the sort of random shock which means that his wages are no longer necessarily close to those of the previous period. Such shocks are more common for younger than older men and also for those whose educational attainment is not very high and whose highest qualification is not academic. The consequence of these influences is that we find an increase in educational attainment boosts significantly the earnings of men aged 30 who are educated to level 3 or less and whose highest qualification is not academic. For men who are aged forty-five when they acquire their qualifications, the returns, although lower, remain significant at a 10 % level for those initially educated to levels 1 or 2. Our model suggests that the reason for this is that older men are more likely than younger men to remain as "stayers", with earnings closely related to their previous earnings. This means that policies to promote lifelong learning by older men need to be combined with policies which ensure that such men have a greater opportunity than currently seems to be the case to take advantage of any qualifications they gain.

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. It carries with it the implication that a single equation approach is mis-specified. It should also of course be noted that it is perfectly possible that the returns

---

[11] They did include initial qualification level as a control variable, but this is not sufficient to distinguish upgrading from simply acquiring a qualification.

of qualifications to men who do not gain qualifications are different from those for men who do. As with all studies of this type, it is not possible to explore that issue.

Our ability to explore the effects of different types of qualification is limited by the data. However, 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.

Appendix 1: The effects of attrition

Sample attrition can arise due to survey non-response or to individuals being excluded for any of the reasons discussed in the text (other than ageing out of the sample). A probit model was used to estimate the probability of attriting in the next survey wave. Table 9 shows the estimated parameters. The score vector from this estimation provides the generalised residuals (equivalently, the inverse Mills' ratio). These are then included in the main model to control for the possibility that the unexplained component of attrition may be correlated with the residuals of any of the equations of

Table 9 Coefficients of the probit model of attrition

Variable	Coefficient/SE
Age	−0.061
	(0.017)
$Age^2/100$	0.086
	(0.020)
Initial qualification	
Level 1	0.041
	(0.044)
Level 2	−0.028
	(0.064)
Level 3	−0.018
	(0.052)
Level 4	−0.036
	(0.064)
Level 5	0.092
	(0.054)
Children	0.061
	(0.038)
Partnered	−0.08
	(0.046)
Interviewer Change	
	0.36
	(0.036)
N	11642
Log-likelihood	−4181.8

Regional dummies, wave dummies and a constant are also included

our system. Included in the probit model is a variable showing whether the interviewer changes between survey waves. This is likely to affect response because panel members may feel more comfortable about responding to a familiar interviewer. The variable is not included in our main model and so acts as an instrumental variable to help with identification.

Appendix 2: Returns to lifelong learning from the unrestricted model

Table 10 shows the results of the simulation for the version of the model in which the learning terms in the "stayers" equation are not restricted to zero. The coefficients on the learning terms are very close to zero and poorly determined (Table 5). It is

Table 10 Returns to lifelong learning from the unrestricted model

	Non-academic			Academic		
	Mean (%)	SD	% < 0 (%)	Mean (%)	SD	% < 0 (%)
Age 30						
No upgrade						
0	3.9	0.050	20.4	−0.7	0.064	52.0
1	4.7	0.050	17.0	0.3	0.062	48.0
2	4.9	0.051	17.2	0.7	0.063	45.1
3	3.1	0.058	30.6	−0.1	0.067	50.1
4	0.6	0.067	47.2	−1.1	0.074	55.3
5	1.5	0.066	40.4	−0.6	0.072	53.6
Upgrade						
0	12.1	0.102	10.4	7.8	0.143	28.2
1	12.2	0.103	11.4	7.9	0.140	27.1
2	12.0	0.108	11.7	7.7	0.143	28.3
3	9.2	0.132	24.2	5.2	0.162	39.0
4	4.8	0.161	39.7	2.5	0.180	46.6
Age 45						
No upgrade						
0	0.7	0.042	43.5	−3.0	0.043	74.8
1	1.4	0.041	36.5	−2.3	0.042	72.0
2	1.4	0.041	34.2	−1.8	0.041	67.3
3	−0.3	0.040	53.8	−1.7	0.039	66.1
4	−1.1	0.038	62.6	−1.2	0.039	62.8
5	−0.7	0.039	58.2	−1.2	0.038	62.7
Upgrade						
0	5.8	0.086	22.8	1.2	0.098	44.7
1	6.0	0.083	22.2	1.3	0.094	43.7
2	5.5	0.085	23.8	1.4	0.093	43.7
3	2.5	0.087	37.9	0.3	0.092	48.5
4	0.4	0.090	48.4	−0.1	0.093	50.5

therefore to be expected that there is little impact on the mean effects and that the results are much less well determined.

Appendix 3: A fixed effects model

Table 11 shows the results of estimating a fixed effects model. In this model, each man's initial educational attainment is absorbed into the individual-specific fixed effect. The term "No upgrade" shows the effect of acquiring a qualification which does not result in any change in the attainment level; the coefficients by the other terms show the effects of acquiring a qualification at this level during the course of the survey, when

Table 11 Parameters of a fixed effects model

Variable		Coefficient/SE
Generalised residual		−0.131
		(0.031)
Gains qualification	No upgrade	−0.011
		(0.010)
Upgrades to	Level 1	0.101
		(0.031)
	Level 2	0.008
		(0.031)
	Level 3	0.08
		(0.033)
	Level 4	0.207
		(0.034)
	Level 5	0.192
		(0.044)
	Other	0.004
		(0.012)
Other variables	Academic qualification	−0.043
		(0.026)
	Age	0.045
		(0.006)
	Age^2	−0.042
		(0.004)
	Public sector	−0.026
		(0.008)
	ln GDP	0.387
		(0.147)
	Regional unemp. deviation	0
		(0.002)
	N	10212
	Log-likelihood	5087.9

Regional dummies and a constant term are also included

it is at a level higher than previously. Upgrading to level 1 or to level 3 or higher has a significant effect on the hourly wage, while acquiring a qualification without upgrading does not.

References

Blanden J, Buscha F, Sturgis P, Urwin P (2012) Measuring the returns to lifelong learning. Econ Educ Rev 31:501–514

Cappellari L, Jenkins S (2008) Estimating low pay transition probabilities accounting for endogenous selection mechanisms. J R Stat Soc Ser C 57:165–186

de Coulon A, Vignoles A (2008) An Analysis of the Benefit of NVQ2 Qualifications acquired at Age 26–34. In: Centre for the economics of education discussion paper No. 106

Dickens R (2000) The evolution of individual male earnings in Great Britain: 1975–95. Econ J 110:27–49

Dickson M (2013) The causal effect of education on wages revisited. Oxf Bull Econ Stat 75:477–498

Dutta J, Sefton J, Weale M (2001) Income distribution and income dynamics in the United Kingdom. J Appl Econ 16:599–616

Egerton M, Parry G (2001) Lifelong debt: rates of return to lifelong learning. High Educ Q 55:4–27

Evans K, Schoon I, Weale MR (2013) Can lifelong learning reshape life chances. Br J Educ Stud 61:25–47

Ferrer A, Menendez A (2014) The puzzling effect of delaying schooling on canadian wages. Can Public Policy 40:197–208

Fitzgerald J, Gottschalk P, Moffitt R (1998) An analysis of sample attrition in panel data: the Michigan Panel Study of Income Dynamics. J Hum Resour 33:251–299

Goodman L (1961) Statistical methods for the mover-stayer model. J Am Stat Assoc 56:841–868

Higher Education Statistics Agency (1995) Students in Higher Education Institutions, pp 1994-1995

Higher Education Statistics Agency (2008) Students in Higher Education Institutions, pp 2007–2008

Holmlund B, Liu Q, Skans O (2008) Mind the gap? Estimating the effects of postponing higher education. Oxf Econ Pap 60:683–710

Jenkins A, Vignoles A, Wolf A, Galindo-Rueda F (2002) The determinants and effects of lifelong learning. Appl Econ 35:1711–1721

Light A (1995) The effect of interrupted schooling on wages. J Hum Resour 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

OECD (2009) Education today 2009: the OECD perspective, Organisation for Economic Co-operation and Development

Purcell K, Wilton N, Elias P (2007) Hard lessons for lifelong learners? age and experience in the graduate labour market'. High Educ Q 61:57–82

Quandt R (1958) The estimation of parameters of a linear regression obeying two different regimes. J Am Stat Assoc 53:873–880

Ramos D (2003) The covariance structure of earnings in Great Britain. Economica 70:353–374

Stata Manual (2014) Bic note. http://www.stata.com/manuals13/rbicnote.pdf#rBICnote

Ulrick S (2008) Using semi-parametric methods in an analysis of earnings mobility. Econ J 11:478–498

UNESCO Institute for Lifelong Learning (2009) In: Global report on adult learning and education, 2009. http://www.unesco.org/uil/en/UILPDF/nesico/GRALE/gralen_en.pdf