Employer Learning and Schooling-Related Statistical Discrimination in Britain *

Fernando Galindo-Rueda †

December 2002

Abstract

This paper develops and tests a new model of asymmetric information in the labour market involving employer learning. In the model, I provide theoretical conditions for the identification – based on the experience and tenure profiles of estimated returns to ability and education – of employer learning about unobserved worker's productivity and statistical discrimination based on years of schooling. Using data from two British birth cohorts, estimates based on this model support the hypothesis that British employers have limited information about their workers, make inferences based on their education levels, and progressively learn about their true ability. Moreover, this learning process – particularly among blue-collar workers– favours incumbent employers relative to potential competitors (asymmetric learning). This informational advantage implies an additional distortion in the functioning of the labour market and policy evaluation rarely takes into account the informational impact of interventions and its implications for individual behaviour.

JEL Classifications: C51, C52, D82, J39, J79.

Key Words: Employer learning, statistical discrimination, asymmetric information, unobserved ability.

Approximate word count: 9,500

*First version: May 2002. Comments and suggestions from Simon Burgess, Jonathan Haskel, Steve Machin and Uta Schönberg, are gratefully acknowledged. Peter Dolton, Gerry Makepeace, Mario Fiorini, Jo Blanden and Leon Feinstein provided useful derived data and made helpful indications about further data derivations in this paper. NCDS and BCS data was provided by the Data Archive in Essex and the Centre for Longitudinal Studies. Funding from the Evidence Based Policy Fund is also acknowledged. The usual disclaimer applies.

†Department of Economics, University College London; Centre for the Economics of Education, Centre for Economic Performance, London School of Economics and IZA. Correspondence address: Centre for Economic Peformance, London School of Economics, Houghton Street, WC2A 2AE. Tel: ++44(0)2079556974. E-mail: F.Galindo-Rueda@lse.ac.uk

1 Introduction

We frequently observe how decisions must be made with very limited information. When facing this type of situation, expectations will be based on any relevant information available.[1] This applies to a very wide range of decisions such as whether to buy a second-hand car, delegate a task to someone, hire a new employee or even decide whether to get married. All these situations have in common a degree of uncertainty about the outcome and the other party involved (who may not always share the exact same objectives) having valuable information. It is no coincidence, judging from these examples, that much value is given to early signals or indicators.[2] Similarly, learning plays a major role on the way these first impressions lead to more sound judgements.

Statistical discrimination and learning often come hand in hand. In the absence of accurate information, people discriminate on the basis of characteristics that are statistically associated to what may be an important element of a decision, even if there is little direct causal relationship.[3] This statistical association is often based on the optimal behaviour of the parties involved.[4] In the classical article by Spence (1973), higher ability individuals can signal themselves as such by incurring investments (such as education) which are less costly for them than for lower ability individuals. Learning about some or all the relevant characteristics will lead to less reliance on early signals.

In the context of an employer learning about a worker's productivity, schooling appears to be a potential source of information, independently of its strict value as human capital. Individuals are often trusted with complex tasks on the basis of their certified, yet unproven, skills. As workers accumulate experience, employers may learn that the individual with a university degree may not be as productive as initially expected or that the worker who dropped out from school is highly motivated and productive. In this situation, we should expect some type of response on the firm's side.[5] Employer learning can take place in a number of different possible ways, ranging from the early stages in which CVs and previous employer's references are required, to personal interviews and in some cases even aptitude and psychometric testing. Learning would then proceed by means of work monitoring, etc.

Testing the existence of learning, signals and statistical discrimination in the labour market is important for many reasons. Educational policies and labour reforms might need to be evaluated within a context in which statistical discrimination and employer learning considerations are present. It is widely acknowledged that a labour market with these characteristics will not resemble the standard text-book competitive model which is used to foresee the impact of numerous policies. The basic motivation for this paper is to provide a sense of how such assumptions

[1]Moreover, in many circumstances substantial costs will be incurred in order to gain additional information and thus make a more accurate judgement.

[2]For example, the mileage of a car, technical reports, good-looks, etc.

[3]For example, discriminating against a job interviewee who is not wearing a suit.

[4]For example, producers of better quality goods can afford to distinguish themselves from lower quality producers by offering a longer-serving guarantee on the good sold.

[5]The obvious variable one would be interested in is wages, although other decisions such as training provision are likely to depend on results of the learning process.

can lead to misleading conclusions.

The idea of linking experience to measures of skill is not new in the literature that studies the presence of informational asymmetry in the labour market.[6] My analysis builds on previous work by Farber and Gibbons (1996) and Altonji and Pierret (1996,2001). They provide an identification strategy to test the employer learning hypothesis. If the econometrician has access to a variable "z" that is associated with an individual's productivity, but is not perfectly observed by the employer, it is possible to infer directly a learning process by observing increasing returns to this variable with experience.

Farber and Gibbons (1996) state three hypotheses following their employer learning/wage determination model: In a competitive labour market where learning occurs at the same rate for all employers. Employer learning does not imply that the coefficient on schooling changes with experience. The part of "z" which is orthogonal to the firm's information set at the beginning of the worker's career will have an increasingly larger association with wages as time passes. Wage growth will be a martingale process in the case of constant productivity, because wage changes will be strictly driven by learning of previously ignored productivity features. Farber and Gibbons's empirical analysis using residualised AFQT as the proxy for the part of "z" orthogonal to the firm's information set confirms two of these three hypotheses. The martingale proposition is rejected in favour of a more parsimonious model of human capital accumulation.

Altonji and Pierret (1996, 2001) analyse a similar model which allows to use log-wages as the dependent variable instead of wage levels. Their model allows them to simultaneously investigate the phenomenon of employer learning and statistical discrimination on education. It implies that the estimated coefficient on the interaction of (non-residualised) ability with experience will be positive, as in Farber and Gibbons (1996). But as learning takes place, the (positive) estimated returns to schooling should decrease because education will be given less credit as a discriminating variable. This hypothesis is confirmed using the same NLSY data used by Farber and Gibbons.

The paper by Bauer and Haisken-DeNew (2001) is, to my knowledge, the only one that has tried to estimate this type of employer learning model outside the US. The problem with their analysis is the lack of a fully satisfactory "z" variable in the German Socioeconomic Panel, which they use for estimation purposes. They look instead at parental education as a proxy for this variable. In fact, they find little evidence of employer learning in Germany. They also study the hypothesis that incumbent employers may have better information about a worker's ability and informally state that asymmetric learning could be inferred from the tenure wage profiles, as opposed to the standard experience interactions.

However, they provide no theory or model backing this type of implication. Under asymmetric learning, the wage determination process is not as straightforward as argued in Farber and Gibbons (1996) or Altonji and Pierret (1996, 2001). Indeed, a firm with private information

[6] Layard and Psacharopoulos (1974) and Psacharopoulos (1979) argue that under the signalling model, the estimated effect of schooling on wages should fall as labour market experience grows, following the above reasoning. This proposition, as stated by Riley (1979), does not account for the fact that a signalling equilibrium requires employers to have an unbiased expectation of the worker's ability.

on a worker will rarely pay the expected marginal productivity as the competitive pressures from other potential employers are diminished by their lack of information. The incumbent firm will trade off potential benefits of lower wages with the probability that the worker might leave. This type of model has been first analysed very recently by Schönberg (2002), who discriminates between empirical predictions associated with the symmetric and the asymmetric learning models. Estimates also based on NLSY data appear to be supportive of asymmetric learning only amongst university graduates.

In this paper I provide new evidence on employer learning in Britain. I look at both symmetric and asymmetric learning. In order to provide some structure supporting the estimation analysis, I have set up a simple model of asymmetric learning which embeds the simpler symmetric learning model. The paper is structured as follows. Section 2 reformulates the standard symmetric learning model, discusses its implications and derives the asymmetric learning model along with its empirical predictions. Section 3 describes the British cohort data used in the empirical analysis and discusses the estimation methods. This is followed by the key estimation results in section 4. Section 5 concludes.

2 Modelling and Testing Employer Learning

I will first argue that the problem of employer learning can be easily described within the standard framework of measurement error in estimation. The basic idea, explained to a larger detail and formalised in the appendix, is that a good proxy to a characteristic that is imperfectly observed by employers but is valued by them (such as ability) will be observed to have a larger impact on earned wages as employers learn about their workers total skills. This learning process is assumed to occur at an unspecified rate as experience levels increase. Conversely, the estimated return to other observable characteristics that are known to be correlated with ability should decrease over time as its informational value falls through learning of real ability.

Considering a linear first order approximation to the effect that experience has on the estimated coefficients through learning, the econometric specification would be as follows:

$$w_{it} = \gamma_0 + \mu_0 t_i + \alpha_0 s_i + \alpha_1 t_i s_i + \beta_0 z_i + \beta_1 t_i z_i + \zeta_{it}, \tag{1}$$

leading to testing the null $H_0 : \beta_1 = \alpha_1 = 0$ versus the alternative $H_1 : \beta_1 > 0; \gamma_1 < 0$.

In a richer economic model, the employer learning hypothesis could easily be generated by factors that have little to do with worker's private information on ability and employer learning. If employers train workers on the basis of their capacity to benefit from it or, alternatively, worker's skills improve faster with experience the higher their initial level of human capital, the hypotheses of non-decreasing returns and non-increasing returns to schooling would be expected. The latter contradicts the basic employer learning model result, and could also be used for identification purposes too. However, it seems extreme to impose the assumption that skills acquired through schooling do not help at all in the process of learning by doing.

The employer learning hypothesis tends to be favoured, because returns to experience are likely to be higher for more able workers and therefore overestimating β_1. The likelihood of rejection also increases because returns to experience are likely to be higher for more educated workers, attenuating the expected negative sign of γ_1.

Most existing models of employer learning, as above described, assume that information about the productivity of workers is public, independently of how incomplete such information can be. This implies that employer learning occurs at the same pace across incumbent and competitor firms. This assumption is essential in order to preserve the implication that in a competitive setting, workers will be paid their expected marginal product. It is easy to see that if an incumbent employer privately realises that a worker's skill is higher than initially expected, wages should respond only up to a point where other potential employers are willing to bid for his/her services. Because this signal is private, potential employers will not fully reward the newly discovered skills and fail to compete with the incumbent employer to the effect of driving the wage closer to the fully perceived productivity.

Schönberg (2002) has recently derived the first explicit model that addresses the possibility of asymmetric employer learning. She models a two-firm bidding game for a worker's labour, where the incumbent firm has acquired privileged information about ability after one period. Induced separations are used to relate wage-tenure profiles to measures of ability in order to derive tests for asymmetric learning. My model shares the approach of using taste shocks to cause exogenous separations, but differs from the previous one in that a competitive setting (as opposed to a two-firm game) is used wherein outsider firms pay what they believe to be the real expected productivity. A continuum of worker schooling and ability types is also allowed for. Empirical predictions are shown to be very similar, with the only exception that identification of asymmetric learning from the data will only be possible when the distribution of non-pecuniary/taste shocks is such that the worker's probability of staying is sufficiently elastic to the expected wage gap between incumbent and outsider firms. If this is not the case, asymmetric learning might go completely unnoticed.

2.1 A simple asymmetric learning model

Consider a very simple specification for a worker's skill level Σ_i, that I assume translates immediately into productivity $\Sigma_i = \alpha s_i + \beta a_i + \xi_i$.

In this specification, the skill level is a linear function of the individual amount of schooling s_i, general ability a_i and an unobserved idiosyncratic component ξ_i, which is assumed independent of ability and schooling. Schooling and ability can certainly be related. For example, we can allow ability to determine the amount of schooling attained. If we invert this relationship and assume linearity again, we could specify general ability by $a_i = \phi s_i + \rho_i$, with ρ_i independent of s_i. We can then substitute this back into the skill equation and get $\Sigma_i = (\alpha + \beta\phi)s_i + \beta\rho_i + \xi_i$.

This specification again indicates the standard problem involved in identifying α and β separately. Learning in the asymmetric framework is assumed to take place as follows: Consider the case where ρ_i can be decomposed into three mutually independent terms and there are only two time periods: $\rho_i = v_{0i} + v_{1i} + v_{2i}$.

When a worker first enters the labour market, all firms can learn immediately about the true value of v_{0i}. This might happen through the interview or hiring process, or it might simply follow from education-related information about the quality of the years spent in education, which is unobservable to the econometrician. After one period, all firms, incumbent and outsiders, can learn about the value of v_{1i}. However, the incumbent firm is assumed to realise the value of v_{2i} too. This ends up with the incumbent firm having a full knowledge about the worker's skills after one period. Extreme as this might sound, the only essential aspect is that after any period of experience, the incumbent firm gets a better signal than its competitors.

For a completely inexperienced worker, the wage should be determined by the expectation of skill conditional on the commonly realised values of s_i and v_{0i}. Therefore, a worker's wage in the first period should be determined as follows:[7] $w_0 = E[\Sigma | x = 0, s, v_0] = (\alpha + \beta\phi)s + \beta v_0$, where x denotes experience.

In order to generate different values of experience in the current firm (tenure), I model separations using a very simple process. After one period, a worker receives an offer w_1^I from the current (I =incumbent/insider) employer and expects to receive an offer w_1^O from any outsider (O) firm. Bertrand-like complexities can be avoided by considering that every individual draws a value θ from a known continuous distribution Γ. This represents the money value of a taste shock relative to staying in the current firm.[8] Hence, a worker will stay if and only if the wage gap $w_1^I - w_1^O$ is higher than the money value of the taste shock θ, which for simplicity is assumed to be distributed independently from all the other variables in the model.

Let us turn now to the determination of the optimal wage offer by the incumbent employer. Since incumbents obtain full information about the worker's skill, they will try to maximise the expected value of the wage offer, which simply reduces to the product of the stay probability times the difference between real productivity and the wage paid.

$$w_1^I \in \arg\max_w (\Sigma - w)\Gamma(w - w_1^O).$$

In this setting, an incumbent employer takes its competitors' wage offers as given, and the stay probability $Pr[w_1^I - w_1^O \geq \theta] = \Gamma[w - w_1^O]$. The first order condition of this problem determines that the expected marginal benefit of a higher wage equates with the marginal cost.

$$(\Sigma - w_1^I)\Gamma'[w_1^I - w_1^O] = \Gamma[w_1^I - w_1^O].$$

The marginal benefit is stated in terms of the additional staying probability when the wage offered is below the real productivity level, whereas the marginal cost follows from the fact of having to pay a higher wage to keep the worker. Dividing both sides by $(w_1^I - w_1^O)$ a rather

[7] To simplify the notation, I will omit the individual index i from now onwards.

[8] This can refer, for example, to personality problems with colleagues or superiors or alternatively, the fact that the individual's partner has found a job in a different city.

intuitive relationship can be found.

$$\epsilon[w_1^I - w_1^O] \equiv \frac{\Gamma'[w_1^I - w_1^O]}{\Gamma[w_1^I - w_1^O]/(w_1^I - w_1^O)} = \frac{w_1^I - w_1^O}{\Sigma - w_1^I},$$

which simply indicates that the staying probability elasticity of the wage offer gap (the term on the left hand side of the equation) must equal the ratio between the wage gap and the incumbent firm's surplus.[9] This implies the wage offer:

$$w_1^I \quad = \quad w_1^O \frac{1}{1+\epsilon[w_1^I - w_1^O]} + \Sigma \frac{\epsilon[w_1^I - w_1^O]}{1+\epsilon[w_1^I - w_1^O]} \approx w_1^O \frac{1}{1+\bar{\epsilon}} + \Sigma \frac{\bar{\epsilon}}{1+\bar{\epsilon}}.$$

This shows that the incumbent firm's wage offer is a weighted average of actual productivity and the competitors' offer, where the weight given to actual productivity is positively associated with the responsiveness of the staying decision to the wage gap between incumbent and outsiders.

The determination of the optimal wage offer made by outsiders is the last element that needs to be considered. Particular attention needs to be paid to the formation of expectations about the skill of a worker observed moving between firms. Firms which make a wage offer and end up hiring an individual, need to realise that their offer has been preferred to that made by other competitors, including the incumbent firm. Winning in this competition is somehow bad news as the chosen worker may be expected to be of lower ability than someone separated from a firm for purely random reasons. In this quasi-competitive setting, the outsider's offer will satisfy:

$$w_1^O = E_{\theta,u_2}[\Sigma | s, v_0, v_1, w_1^I - w_1^O \geq \theta].$$

Using the result from the incumbent firm wage setting decision this can be rewritten as follows:

$$v_2^O = E_{\theta,u_2}[\Sigma | s, v_0, v_1, w_1^I - w_1^O \geq \theta] = E_\theta \left\{ E_{v_2} \left[v_2 | v_2 \leq \frac{1+\bar{\epsilon}}{\bar{\epsilon}} \frac{\theta}{\beta} + v_2^O | \theta \right] \right\}. \tag{2}$$

Under some regularity conditions on the distributions of θ (Γ) and v_2 it is possible to find a solution v_2^O to equation (2).[10]

This simply states that the equilibrium reward v_2^O for the unknown ability component v_2 must be equal to the expected value of this variable conditional on being smaller than the equilibrium reward plus the correction term $\Delta = (1+\bar{\epsilon})\tilde{\theta}/(\bar{\epsilon}\beta)$. The existence of a solution requires that the correction term is positive, which basically means that the nature of the taste shock must involve

[9]For simplicity, I will assume that the elasticity ϵ does not depend on the wage gap $(w_1^I - w_1^O)$. This will allow considerable simplifications as a first order approximation, it may not be so unreasonable to treat the gap effect on the elasticity as negligible compared to other effects.

[10]For illustrative purposes, it is helpful to think of a value $\tilde{\theta}$ for which the following approximation is valid: $v_2^O = E_{v_2}\left[v_2 | v_2 \leq \frac{1+\bar{\epsilon}}{\bar{\epsilon}} \frac{\tilde{\theta}}{\beta} + v_2^O \right]$. This value will rarely coincide with the mean value of θ because the conditional mean is unlikely to be a linear function. For example, under normality, this function will be concave and higher order moments would be required to improve the approximation.

on average a negative pre-disposition to stay.[11] This allows outsider firms to believe rationally that workers who end up leaving their jobs to join them are not necessarily less able workers who received very low offers from their incumbent employers.

2.2 Predictions of the asymmetric learning model

The asymmetric learning model presented above suggests a differentiated wage pattern for inexperienced workers, experienced workers who quit their original employment, and experienced workers with a higher tenure level. In a two-period model such as this, there are only three possible combinations of experience and tenure, although richer combinations could be derived from a longer horizon model with additional learning components. Wages would be given by:

$$\begin{aligned} w_0 &= (\alpha+\beta\phi)s+\beta v_0 & (3)\\ w_1^O &= (\alpha+\beta\phi)s+\beta(v_0+v_1)+\beta v_2^O & (4)\\ w_1^I &= (\alpha+\beta\phi)s+\beta(v_0+u_1)+\beta\frac{\bar{\epsilon}}{1+\bar{\epsilon}}v_2+\beta\frac{1}{1+\bar{\epsilon}}v_2^O & (5)\end{aligned}$$

In these specifications, only schooling is observable to the econometrician. If a measurement of ability z is available, measurement problems would still make it impossible to identify the parameters although inference about the features of the discrimination and learning model become possible. Linear projections of wages for each separate group on schooling and measured ability would produce the following coefficients on ability:[12]

The OLS projections on ability and education would give precise estimates that can be specified in a compact way. For a given experience level x and tenure τ, the projection on ability would be:[13]

$$\mu_z(x,\tau)=\frac{\beta Cov(z,v_0)}{Var(z-\pi s)}+x\cdot\frac{\beta Cov(z,v_1)}{Var(z-\pi s)}+\tau\cdot\frac{\bar{\epsilon}}{1+\bar{\epsilon}}\cdot\frac{\beta Cov(z,v_2)}{Var(z-\pi s)}\equiv\beta_0+\beta_1x+\beta_2\tau,$$

and estimates for the learning/discrimination model follow from the estimation of:

$$w_{it}=\gamma_{0t}+\gamma_{1t}x_{it}+\gamma_{2t}\tau_{it}+\alpha_{0t}s_i+\alpha_1x_{it}s_i+\alpha_2\tau_{it}s_i+\beta_{0t}z_i+\beta_1z_ix_{it}+\beta_2\tau_{it}z_i+\xi_{it}. \qquad (6)$$

This equation reflects the predictions made by the asymmetric employer learning model. OLS projections on the experience, tenure, ability and education variables $(x_{it},\tau_{it},z_i,s_i)$, and the pertinent interactions generalise the predictions of the two period model to a wider setting. It is obvious that with a longer period model, the number of possible experience/tenure paths grows

[11] Higher order moments in the distribution of θ should meet additional requirements, such as limits on the dimension of the variance, etc. Without imposing additional assumptions on the distributions, it is not possible to guarantee the existence of a unique equilibrium wage offer either.

[12] These would follow from: $E^*[w_0|s,z]=\mu_{s0}s+\mu_{z0}z$, $E^*[w_1^O|s,z]=\mu_{s1}^Os+\mu_{z1}^Oz$ and $E^*[w_1^I|s,z]=\mu_{s1}^Is+\mu_{z1}^Iz$, where $E^*[y|s,z]$ denotes the linear projection of y on s and z.

[13] Something similar can be written for the projection on schooling, that would be expressed as: $\mu_s(x,\tau)=\alpha+\beta\phi-\pi\left(\frac{\beta Cov(z,v_0)}{Var(z-\pi s)}+x\cdot\frac{\beta Cov(z,v_1)}{Var(z-\pi s)}+\tau\cdot\frac{\bar{\epsilon}}{1+\bar{\epsilon}}\cdot\frac{\beta Cov(z,v_2)}{Var(z-\pi s)}\right)\equiv\alpha_0+\alpha_1x+\alpha_2\tau$. Time subscripts are omitted for presentational simplicity.

and complicates the theoretical interpretation of the coefficients.[14] Linear interactions should then be considered as a first order approximation to differences between experience levels, and also within, by different tenure magnitudes. Conditional on experience and its interactions, the tenure interaction with ability indicates the extent of additional learning by the incumbent captured by the stayer's equation that displays the correlation between the private signal v_2 and the ability variable. The key issue here is that the term $\bar{\epsilon}/(1+\bar{\epsilon})$ should be positive, otherwise retention of workers is not very sensitive to wage differences and employers do not need to match potential outside offers when $\bar{\epsilon} = 0$.

Thus, by introducing interactions in the model it is possible to capture the different components of learning. As said, asymmetric learning can be inferred from the tenure interaction provided $\bar{\epsilon}$ is positive, otherwise a positive covariance between the private signal and ability would go unnoticed.

3 Data and Estimation Methods

The limited research up to date on employer learning, particularly in Europe, can be well explained by the lack of adequate data, namely, a measure of ability which employers cannot observe. The National Child Development Study (NCDS) and British Cohort Study (BCS) are continuing longitudinal studies of two birth cohorts born in Britain between 3 and 9 March 1958 and 7 and 11 April 1970, respectively.[15] The richness of the information held about each cohort member makes this one of the few appropriate data sources for testing employer learning and statistical discrimination in the UK. Children interviewed at the ages of 10 and 11 were subject to a series of tests aiming to measure different elements of ability and numeracy and literacy skills. These test scores have proven to be good predictors of further educational qualification attainment, appear to have a direct effect on explaining wages independently of qualifications, and therefore can be considered as approximate measurements of an individual's ability.[16]

The ability measure used in this paper follows partially the methodology used in Cawley *et al* (1996,1998,2001). Test scores obtained at the age of 11 in the NCDS and the age of 10 for BCS constitute the basis for the analysis because of the proximity in terms of age across cohorts.[17] Because the tests administered in each case were not identical, even though NCDS and BCS cohort members were both measured in terms of their maths and reading ability, it is not possible to use a raw test score in the analysis. This problem is circumvented by calculating

[14] A larger number of periods imposes on the theoretical model the calculation of more equilibrium expectations of ability for movers, and additional assumptions are required about history-dependence, etc.

[15] For a detailed description of these studies, see the Centre for Longitudinal Studies website: *http://www.cls.ioe.ac.uk/cohort/cstudies.html.*

[16] This does not imply that such measured ability is innate, as it has been repeatedly shown that attainment progression (transitions in the score distributions) through childhood is very much influenced by parental and environmental characteristics.

[17] NCDS test scores at the age of 11 were (i)reading, (ii)maths ability, (iii)non verbal general ability, (iv) verbal general ability and (v)copying designs. BCS test scores at 10 include (i)maths, (ii) reading and (iii) British Ability Scale test of general ability.

the first principal component for each cohort from the set of available tests.[18] In the psychometric literature, this measure has been frequently associated with the construct "g", described as the underlying general ability or intelligence factor. In a series of papers using the NLSY data, Cawley *et al* (1996,1998,2001) find that the coefficient on "g" in the log wage regression is positive and statistically significant in all cases. They note that measured "g" has similar properties to the commonly used US military's Armed Forces Qualification Test (AFQT), which is derived from four ASVAB tests. AFQT is the measure used by Farber and Gibbons (1996) (in its residualised version) and Altonji and Pierret (1996,2001) in their analysis of employer learning.

Regardless of what is the best way to measure general intelligence, the main reason for using "g" is to enable the conversion of a set of ability variables into a single, continuous, cross-cohort-comparable variable. Using a single ability variable also allows simplification of the interpretation of the ability/experience and ability/tenure wage profiles. Accordingly, the ability measure should be considered as representing an individual's relative position in her own cohort's ability distribution.[19]

The measurement of "g" used in this paper is undoubtedly more likely to serve its purpose as a proxy for ability that cannot be fully observed by employers than variables such as parental education, as used in Bauer and Haisken-DeNew (2001). From now onwards, I will refer to it as either "g" or "z".

Labour market data from the NCDS was obtained in 1981 (age 23), 1991 (age 33) and 1999/2000 (age 41/42), whereas for BCS, adult data was collected in 1991 (age 21), 1996 (age 26) and 1999/2000 (age 29/30).[20]

Experience and schooling variables have been derived from the reconciled work histories for both cohorts. Schooling is considered as the number of years in full time education after the sixteenth birthday.[21] These magnitude, asides from recall problems, are accurate up to a month's scale. Although years of schooling may not be a good measure of an individual's education in a wage equation -given the alternative academic and vocational paths- the decision to focus on years of education can be considered more in the context of a reduced-form model, in which years of education is the 'effort input', which leads to the attainment of a given qualification. It is important to emphasise at this point that it is not returns to education that the study is

[18] More details are available on request.

[19] There is also a strong comparative advantage of using a measure of "g" obtained at the same age for all cohorts, relative to the AFQT measure in NLSY. The latter, by referring to different age levels, is differentially affected by levels of schooling and parental influence that increase with age. Residualising with respect to age may not be sufficient because parental socio-economic characteristics also shape movements in the distribution of ability. For example, a low SES child with high ability score at an early age may not score as high at a later age. Similarly, low scoring high SES children will be more likely to improve their position in the distribution.

[20] All surveys except BCS26 were conducted in the form of personal interviews, in which information was obtained about family background, socio-economic conditions, health, education and many other characteristics. The BCS26 survey differed from the rest in that a self-completion postal questionnaire was used as the survey instrument, probably because of financial constraints. Also financial constraints have affected the structure of each survey inasmuch as these surveys depended on particular sources of funding which prioritised the quality of certain types of information over other types. All surveys except BCS21 targeted the complete population of the study, in this case, each birth cohort as detailed above. BCS21 focused instead on a representative subsample of individuals.

[21] The statutory school leaving age for both cohorts.

trying to measure, but the education and ability-experience profiles. The existing literature has focused until now on using years, even in Germany for the case of Bauer and Haisken-DeNew (2001), where qualifications were converted to a year equivalent measure. Another reason for using years is to improve the comparability between and within cohorts, where the responses to qualification attainment questions proved very sensitive to how the questions were phrased.

Table 1: Sample composition by age and year

	year			
age	1981	1991	1996	2000
21		371 (BCS)		
23	3830 (NCDS)			
26			1890 (BCS)	
30				2941 (BCS)
33		2922 (NCDS)		
42				2828 (NCDS)

NOTE: Number of valid observations available for estimation in each age*year cell, following from NCDS and BCS70 adult data. Age and year groups are generically defined, as not all men in each survey were interviewed at the same time. In some cases there are several months of difference.

The sample used here is restricted to person-year observations of working men at the time of being interviewed and have a valid wage observation, ability information from the age of 10/11 and complete work histories prior to the observation date. Table 1 shows the final composition of the sample by age and year.

The older cohort (i.e. NCDS), is illustrated by the lower diagonal of observations and clearly indicates increasing drop-out figures. The younger cohort (i.e. BCS) corresponds to the upper diagonal and the pattern is different due to the reduced target sample in 1991 (which must be added to the fact that some individuals had not yet finished their education) and the reduced sample in BCS26 because of the low response to self-completion questionnaires.

This data description suggests that there are practical problems that need to be considered before attempting to test the employer learning hypothesis. These involve the potential selectivity of the data used, partly through differential attrition by age and year, and also the identification of learning separately from potential changes in the prices of skills. To deal with these issues, it is also necessary to understand how the stylised two-period model could be set within a typical estimation framework.

The symmetric learning model, in which all potential employers learn simultaneously about workers' ability can be specified econometrically as follows:

$$w_{it} = \gamma_{0t} + \gamma_{1t}x_{it} + \alpha_{0t}s_i + \alpha_1 x_{it}s_i + \beta_{0t}z_i + \beta_1 z_i x_{it} + \xi_{it}. \tag{7}$$

This parallels equation (6) and is consistent with the predictions of the learning model when the private signal to the incumbent v_2 is absent ($v_2 = 0$).[22] As noted, it is important to control for the selective composition of the samples suitable for estimation. One possible approach is to model the process that determines whether an individual enters the valid estimation sample in a given year using background childhood and family data for the earlier sweeps. If this process is adequately controlled for, it will be possible to obtain unbiased estimates of the earnings process that are relevant for the learning model. In order to instrument participation, I include a wide range of background variables which proved not to be significant if included in the main wage equation but proved significant in their prediction of participation in the survey.[23] The selection model appears to perform satisfactorily, but the selection term turns out to be not statistically significant in either of the wage equations estimated in this chapter.

The second important issue that needs to be taken account of is the possibility that the way in which the market values skills has changed over the time period under consideration. Failure to control for this could mislead us into accepting the learning model when there has been an aggregate increase in the returns to cognitive ability, possibly due to a higher demand for this type of skill or changes in the overall supply.

The root of the problem is the need to combine the need of a sufficient degree of variation in experience with the adequate controls for potential changes in returns over time. The fact that only two birth cohorts are available and there are only three wage observations at most for every individual certainly complicates the identification strategy.

Variation in experience is achieved through the combination of having observations at different ages and within the same age groups, some individuals have been interviewed in different months and they differ, conditional on a given level of schooling, in the amount of time they have been unemployed or out of the labour force for reasons other than education. The latter source of variation is not very helpful as it induces elements of endogeneity and that is the reason why the emphasis is on the age variation. With a single cohort, employer learning would be confounded with a potential increase in the returns to ability in the economy. Two cohorts are therefore the minimum data requirement but separate identification of time and learning effects is not feasible in a purely non-parametric approach.

My approach here is to allow for different returns to ability and schooling in 1981 and 2000, leaving the early and mid-nineties as the benchmark case. In practice, the only significant differences turn out to be driven by the comparison between the 1981 NCDS survey and the rest (1991 and after), with a remarkable increase in the returns to both cognitive ability and years of schooling afterwards. It is important to note that when introducing this type of control, the hypothesis of employer learning may be rejected more often than it should, as it could lead to ignoring the full extent of the learning process taking place between 1981 and 1991 in the NCDS.

[22] Another feature is that it allows for changes in the prices of skills (ability, education and experience (x_{it}) over time t). For simplicity, the interaction terms are not interacted with time.

[23] These include a quadratic in the ability index, parental class and education information and whether the teacher had reported that the individual, as a child, used to lie at school. Controlling for ability, the assumption of conditional independence required is more likely to be satisfied in practice than in a data set with more limited information.

Whereas the process of experience accumulation is more parsimonious (entirely deterministic in a model without unemployment), the same is not true for tenure. The probability of staying in the same firm is associated with the private signal the incumbent employer receives, and this is correlated with overall ability, $Prob[\theta \leq w_1^I - w_1^O] = \Gamma[\beta \frac{\epsilon}{1+\epsilon}(v_2 - v_2^O)]$.

In a standard OLS regression of wages on tenure, this variable is thus likely to be endogenous if the appropriate conditioning variables are not included and this will bias estimates. This is not exactly the case here, as the variables that determine tenure accumulation are accounted for through the ability variable. The idiosyncratic taste shock θ has been assumed independent of productivity and its omission does therefore not affect results.[24] These crucially depend on the quality of the ability measure used in the analysis.

4 Estimation Results

I have first considered the symmetric employer learning model, as specified by equation 7. In terms of the extended model, this corresponds to the case with the distribution of v_2 degenerate at its mean value zero.[25] Under the assumption that the ability variable z is correlated with v_1, conclusions can be drawn from the experience profiles arising from the estimated returns to ability and schooling by examining the coefficients β_1 and α_1.

Table 2: OLS estimates from the symmetric learning model

	I		II		III	
	Coef	SE	Coef	SE	Coef	SE
Schooling	0.0672	0.0029	0.0656	0.0054	0.0456	0.0060
Cognitive ability	0.1189	0.0056	0.0812	0.0131	0.0533	0.0125
Experience	0.0612	0.0034	0.0571	0.0038	0.0479	0.0051
ExperienceSq/100	-0.1200	0.0101	-0.1089	0.0107	-0.0931	0.0144
Schooling*Exp/10			0.0025	0.0044	0.0058	0.0046
Ability*Exp/10			0.0256	0.0081	0.0179	0.0077
Mills ratio					0.0025	0.0211
Training					0.0748	0.0069
R-squared	0.3216		0.3225		0.3875	

NOTE: Pooled OLS estimates, with standard errors robust to within-individual correlation. Dependent variable: Log gross real hourly wage. (Prices=Jan 2001). All specifications include controls for different returns to ability and schooling in 1981 and year dummies. Specification III also includes industry, occupation, firm size and marital status dummy variables. The Mills ratio refers to the derived variable $\frac{\phi(\lambda_j e_i)}{\Phi(\lambda_j e_i)}$, where $\Phi(\lambda_j e_i)$ denotes the probability of a cohort member being interviewed in sweep τ, ϕ the associated normal density and e_i is a vector of background characteristics used to estimate this process.

Table 2 contextualises estimates from the learning model with a typical human capital regres-

[24] Although z is a proxy for real ability, it is important to note that the predictions of the model are based on the actual OLS projections. Hence, the validity of interacting ability and education with tenure hinges on the adequacy of the theoretical model to reproduce the data generating process for tenure.

[25] If employer learning takes place, the distribution of v_1 (i.e. the ability information not available at the beginning of an individual's career but further revealed to all firms) is not degenerate.

sion which includes years of schooling and experience, as shown in specification I, plus a measure of cognitive ability. Estimates in specification II suggest that much of the cognitive ability effect is associated with experience in the labour market. The schooling*experience interaction is positive, although not significantly different from zero.[26] This result proves to be robust to the inclusion of additional explanatory variables, including the sample selection correction, which does not appear to have any impact on the estimated wage process.[27] Overall, the symmetric employer learning model hypothesis on the ability term is accepted (increasing returns to g with experience), although the estimates do not fully confirm the accompanying hypothesis of a negative coefficient on the schooling-experience interaction. There seem to be signs of employer learning within a marginally richer model of human capital accumulation.

The next step is to understand better the nature of this learning process. The model states that without private signals to incumbent employers, $\alpha_2 = \beta_2 = 0$ in equation 6. This is so because there is no private realisation of the v_2 ability signal and all further realisations are of the v_1 type, i.e. common to all employers. However, under asymmetric learning, $\alpha \leq 0$ and $\beta_2 \geq 0$.[28]

Table 3: OLS estimates from the asymmetric learning model

	I		II		III	
	Coef	SE	Coef	SE	Coef	SE
Schooling	0.0666	0.0029	0.0661	0.0054	0.0461	0.0060
Cognitive ability	0.1177	0.0056	0.0800	0.0131	0.0525	0.0125
Experience	0.0554	0.0035	0.0498	0.0039	0.0435	0.0052
ExperienceSq/100	-0.1090	0.0107	-0.0961	0.0113	-0.0852	0.0149
Tenure	0.0116	0.0017	0.0156	0.0019	0.0097	0.0018
TenureSq/100	-0.0315	0.0081	-0.0387	0.0081	-0.0243	0.0077
Schooling*Exp/10			0.0148	0.0054	0.0165	0.0055
Ability*Exp/10			0.0246	0.0093	0.0216	0.0088
Schooling*Tenure/10			-0.0249	0.0054	-0.0211	0.0051
Ability*Tenure/10			0.0012	0.0088	-0.0073	0.0081
Mills ratio					0.0025	0.0211
Training					0.0718	0.0070
R-Squared	0.3256		0.3282		0.3901	

NOTE: Same specifications as in table 2, apart from the tenure terms displayed.

Table 3 looks at the asymmetric learning hypothesis. Introducing interactions, the positive

[26]This pattern of increasing returns to ability with experience is also confirmed by looking at the age profiles for the estimated returns in each one of the available surveys. Returns to schooling appear to increase over age but a much lower rate and for the NCDS cohort members over thirties it completely disappears. Results available on request.

[27]I also considered a more general specification by interacting the selection correction term with year dummies. This produced no significant effect for any of the estimates presented in the paper.

[28]The differences in the tenure profiles will be driven by the magnitude of the responsiveness of staying decisions to the insider-outsider wage offer gap. With a low responsiveness, incumbent firms do not need to reward workers for skills that would not be valued (in expectations) by outsiders and learning does not translate into a stronger relationship with the ability proxy z ("g") used in this study. When the 'staying' elasticity to the offer gap is high, private learning can be more easily inferred from a differentiated tenure profile.

effect of schooling and ability appears to increase significantly with total labour market experience. Interestingly, the effect of schooling falls strongly with tenure, as the asymmetric learning model with high stay-response elasticity would suggest. Thus it looks as if conventional schooling becomes a less important determinant of wages the longer an individual stays employed in the same firm. The opposite result does not appear to follow from the ability/tenure interaction, which is insignificant. One possible explanation is that private learning refers to a variable not related to the cognitive ability index, such as motivation, capacity to work hard, etc.

One possible research avenue to understand better the employer learning process is to separate workers according to their occupational classification. Two groups have been made in this case: white and blue collar workers. The estimates related to the symmetric learning model are reproduced in table 4. Looking at the most basic specification for white collar workers, both implications from the employer learning model are satisfied: $\beta_1 > 0$ and $\alpha_1 < 0$.[29]

Table 4: OLS estimates from the symmetric learning model: By occupation class

	White collar workers			Blue collar workers		
	I	II	III	I	II	III
Schooling*Experience/10	-0.0142	-0.0191	-0.0140	0.0321	0.0143	0.0163
	(0.0052)	(0.0059)	(0.0067)	(0.0127)	(0.0183)	(0.0178)
Ability*Experience/10	0.0219	0.0107	0.0117	0.0078	0.0020	-0.0046
	(0.0101)	(0.0124)	(0.0119)	(0.0080)	(0.0109)	(0.0104)
Mills ratio			0.0008			0.0009
			(0.0354)			(0.0245)
Training			0.0843			0.0657
			(0.0105)			(0.0087)
R-Squared	0.3211	0.3228	0.3776	0.1752	0.1765	0.2596
Observations	7286	7286	7286	7476	7476	7476

NOTE: Pooled OLS estimates, with standard errors robust to within-individual correlation within parentheses. Dependent variable: Log gross real hourly wage. (Prices=Jan 2001). Specification I includes year dummies, ability and a quadratic polynomial in schooling and experience. Specification II also includes controls for different returns to ability, schooling and experience in 1981. Specification III additionally incorporates industry, occupation, firm size and marital status dummy variables.

With regard to blue collar workers, there seems to be little evidence about any type of dependence of experience profiles on either ability or schooling.[30] Comparing white and blue collar estimates, it is also interesting to notice that the returns to schooling are decreasing for white collar workers and increasing for blue collars. The magnitude of the returns to experience is also considerable. White collar workers appear to have much steeper profiles. Additionally, predictions based on individual characteristics seem to perform much better for white collars[31],

[29]However, after controlling for possible different returns to ability and schooling in 1981, the positive effect of experience on the returns to ability becomes statistically insignificant, although remains positive. It is important to note that, given the limitations of the data, this type of control will tend to diminish the evidence supporting employer learning. This could thus be considered as a lower bound on the estimated learning effects.

[30]The only exception is the most basic specification I.

[31]There is considerably less variation in terms of schooling amongst blue collar workers, many of which have no reported spells of full time education after their 16^{th} birthday.

although job related characteristics have a higher explanatory power for the group of blue-collar workers.

Table 5: OLS estimates from the asymmetric learning model: By occupation class

	White collar workers			Blue collar workers		
	I	II	III	I	II	III
Schooling*Experience/10	-0.0050	-0.0099	-0.0054	0.0352	0.0186	0.0217
	(0.0063)	(0.0069)	(0.0076)	(0.0171)	(0.0221)	(0.0211)
Ability*Experience/10	0.0274	0.0160	0.0204	-0.0014	-0.0073	-0.0080
	(0.0127)	(0.0148)	(0.0141)	(0.0093)	(0.0119)	(0.0113)
Schooling*Tenure/10	-0.0177	-0.0169	-0.0145	-0.0099	-0.0113	-0.0142
	(0.0060)	(0.0060)	(0.0058)	(0.0172)	(0.0172)	(0.0160)
Ability*Tenure/10	-0.0117	-0.0105	-0.0154	0.0229	0.0234	0.0105
	0.0144	0.0144	0.0137	0.0110	0.0110	0.0102
Mills ratio			-0.0021			0.0012
			0.0357			0.0245
Training			0.0852			0.0602
			0.0106			0.0088
R-Squared	0.3232	0.3247	0.3792	0.1898	0.1910	0.2653
Observations	7236	7236	7236	7476	7476	7476

NOTE: Pooled OLS estimates, with standard errors robust to within-individual correlation within parentheses. Dependent variable: Log gross real hourly wage. (Prices=Jan 2001). All specifications include level ability, a quadratic polynomial in schooling, experience and tenure and year dummies. Specification II incorporates controls for different returns to ability, schooling, experience and tenure in 1981. Specification III also includes industry, occupation, firm size and marital status dummy variables.

Turning now to the more general asymmetric learning model, there also seems to be a clearly differentiated pattern between white and blue collar workers. After controlling for experience, there is no evidence of a positive effect of tenure on the estimated returns to ability, although as argued before for the whole sample, the returns to schooling appear to fall with tenure. As for blue collar workers, there is some indication of increasing returns to cognitive ability with tenure for the specifications without job characteristics. The sign on the schooling interaction is negative (although insignificant), as predicted by the asymmetric learning model. It is also interesting to note that returns to tenure are almost twice those for white collar workers.

In conclusion, there seems to be some reason to support the learning hypothesis in its asymmetric form amongst blue collar workers, although this learning may not so much refer to the cognitive ability index used in this paper. It is more obvious to say that firms do initially discriminate on the basis of schooling amongst this group. The fact that the evidence supporting learning for blue collar workers disappears once we control for firm characteristics raises new questions, such as whether firms decide to train workers after they learn about them, and whether the type of training received makes them less likely to move elsewhere because of largely untransferable job-specific skills. This type of effect would be unlikely to be reflected in wages and therefore little inference could be made from the experience and tenure profiles.

I examine this hypothesis in table 6. The theoretical structure devised for wage determination may not directly translate into a similar set of training incentives for firms.[32] I estimate the reception of training as a function of the same characteristics used in the earnings regressions. For the combined sample, there are positive (although decreasing) effects of schooling, experience and tenure on training incidence. Cognitive ability plays an important role too. The important aspect here is the decreasing effect of schooling on the incidence of training as tenure increases, with the opposite effect holding for the incidence of ability. When introducing additional controls, the schooling interaction effect loses importance, but it becomes clear that previously laid-off workers are less likely to receive training.

Table 6: Firm-sponsored training in current job: Probit estimates

	All		White collar		Blue collar	
	I	II	I	II	I	II
Schooling*Experience/10	-0.0397	-0.0188	-0.0282	-0.0249	0.0933	0.0688
	(0.0141)	(0.0145)	(0.0177)	(0.0180)	(0.0366)	(0.0383)
Ability*Experience/10	0.0087	0.0060	-0.0451	-0.0494	0.0802	0.0585
	(0.0232)	(0.0239)	(0.0359)	(0.0362)	(0.0337)	(0.0350)
Schooling*Tenure/10	-0.0219	-0.0183	-0.0294	-0.0246	0.0135	0.0281
	(0.0149)	(0.0153)	(0.0168)	(0.0170)	(0.0458)	(0.0473)
Ability*Tenure/10	0.0533	0.0428	0.1147	0.1109	-0.0072	-0.0278
	(0.0261)	(0.0269)	(0.0394)	(0.0396)	(0.0382)	(0.0400)
Laid-off movers		-0.0654		-0.0347		-0.0896
		(0.0374)		(0.0541)		(0.0514)
Voluntary movers		-0.0021		0.0843		-0.0914
		(0.0286)		(0.0399)		(0.0411)
Log-likelihood	-9848.66	-9054.97	-4772.49	-4658.07	-4821.92	-4290.76
Pseudo-R2	0.0351	0.1128	0.0437	0.0667	0.0285	0.1355
Observations	14762	14762	7286	7286	7476	7476

NOTE: Binary dep.var.=1 if individual received formal training in current job. Coeffs and rob.standard errors reported. Voluntary movers:=report having left previous job to earn more or to get a better one. Laid-off movers:= reported dismissal or laid off from their previous job or fixed term expired, not renewed by employer. Ability, schooling, experience and tenure square polynomials also included.

The learning effect on the incidence of training amongst white collar workers is very strong, judging from the opposite signs of the α_2 and β_2-equivalent coefficients. In this case, voluntary movers are also found to be more likely to receive training. For blue collar workers, α_1 and β_1 are both positive (experience interactions). The tenure profiles in this case do not reveal much, but it is quite remarkable to see how being a "mover" reduces the chances of receiving training, independently of the context (laid-off versus voluntary job changers). Although it is possible that these results actually reflect some sort of inverse causation, a reasonable interpretation is that employer learning is present in many parts of the labour market and that it operates in many different ways. The effects on training indicate that it is not possible to ignore the process of human capital accumulation, which will have very different implications depending on the specificity of the human capital. Asymmetric learning shows that even perfectly transferable human capital may not be so because of the informational advantage that an incumbent firm

[32]Unfortunately it is not possible to distinguish between general and firm specific training.

may have with respect to other potential employers.

5 Concluding Remarks

This paper has discussed the empirical implications of a model of asymmetric employer learning about a worker's real productivity. This framework encompasses existing models on symmetric learning and also provides some foundations for existing informal tests of asymmetric learning. Additionally, it has provided new evidence on the existence of labour market discrimination based on schooling levels and on the process of employer learning about male workers' ability in Britain. This could also explain the increasing reliance by employers on personality questionnaires and related forms of psychometric testing, as documented in Jenkins (2001).

Broadly speaking, there is support for the hypotheses of ability learning and schooling-related discrimination particularly for white collar workers, although human capital accumulation in the workplace seems to play an equally important role *vis à vis* explaining the observed ability and schooling experience-related wage profiles. Such process of skill accumulation can also be driven by the existence of employer learning. Estimates of the probability of receiving training from the employer coincide with this view.

With regard to the evidence on incumbent employers having privileged learning opportunities, there seems to be some mild support for the hypothesis that this is the case for blue-collar, less educated workers. Asymmetric learning for this group implies that employers will have an even stronger monopsonistic power over their employees. This suggests that further efforts should be made to devise schemes of skill certification for less educated British workers. A dual apprenticeship scheme might be a feasible mechanism.

Throughout this paper, ability has been widely identified with the more specific concept of cognitive ability. It is left for further research to investigate learning on the related concept of social intelligence.[33]

Together, these results combined present a richer description of the British labour market which substantially differs from the standard view amongst many economists and policy makers. In order to prevent unintended distortions, educational reforms and interventions aimed at improving the skill level of workers need to foresee their likely impact on the informational value associated to qualifications. Changes in the informational content of signals in an economy will have a wide range of repercussions on the human capital investment efforts made by individuals and on the way expectations are formed by employers about their employees' real skills.

References

[1] Altonji, J. and Pierret, C. (1996). "Employer Learning and the Signaling Value of Education", NBER Working Paper No. 5438, National Bureau of Economic Research: Cambridge,

[33] Estimates using an index of social adjustment measured while in school appeared not to be significant, but showed the right signs associated with the learning hypothesis. These results are available on request.

MA.

[2] Altonji, J. and Pierret, C. (2001). "Employer Learning and Statistical Discrimination", *The Quarterly Journal of Economics*, 116 (1), pp. 313-351.

[3] Bauer, T.K. and Haisken-DeNew, J.P.(2001). "Employer Learning and the Returns to Schooling", *Labour Economics*, 8, pp. 161-180.

[4] Cawley, J., Heckman, J. and Vytlacil, E. (2001). "Three Observations on Wages and Measured Cognitive Ability", *Labour Economics,* 8.

[5] Cawley, J., Heckman, J. and Vytlacil, E. (1998). "Cognitive Ability and the Rising Return to Education", NBER Working Paper No. 6388. National Bureau of Economic Research: Cambridge, Ma.

[6] Cawley, J., Conneely, K., Heckman, J. and Vytlacil, E. (1996). "Measuring the effects of cognitive ability", NBER Working Paper No. 5645. National Bureau of Economic Research: Cambridge, Ma.

[7] Farber, H.S. and Gibbons, R. (1996). "Learning and Wage Dynamics", *The Quarterly Journal of Economics*, 111, pp. 1007-1047.

[8] Jenkins, A. (2001). "Companies' Use of Psychometric Testing and the Changing Demand for Skills: A Review of the Literature", Centre for the Economics of Education Discussion Paper No. 012, June: London.

[9] Layard, R. and Psacharopoulos, G. (1974). "The Screening Hypothesis and the returns to education", *Journal of Political Economy*, 82, pp. 985-998.

[10] Psacharopoulos, G. (1979). "On the Weak versus the Strong Version of the screening hypothesis", *Journal of Political Economy*, 82, pp. 985-998.

[11] Riley, J.G. (1979). "Testing the Educational Screening Hypothesis", *Journal of Political Economy*, 84, pp. 279-252.

[12] Schönberg, U. (2002). "Testing for Asymmetric Information in the Labour Market", unpublished manuscript.

[13] Spence, A.M. (1973). "Job Market Signaling", *The Quarterly Journal of Economics*, 87, pp. 355-379.

A Appendix: Employer Learning and Measurement Error

Let us assume that employers pay workers according to their productivity $Y_i \equiv \exp(y_i)$, which I define as a very simple function of a worker's ability a_i and years of education s_i.

$$y_{it} = \gamma + \alpha s_i + \beta a_{it} + \kappa_{it} + \xi_{it}.$$

This model assumes that education can have a positive effect on productivity independently of ability a_i by allowing α to be positive. A pure signalling model of education would imply $\alpha = 0$ and $\beta > 0$. At the other extreme, a pure human capital model would be characterised by $\alpha > 0$ and $\beta = 0$, although education attainment could be partially determined by ability. To simplify the exposition, I will assume that ξ_i is independent of both a_i and s_i.

In this setting, it is also possible to allow an individual's productivity to vary with experience t, representing a skill improvement through learning by doing, etc. This process of skill improvement is modelled as homogeneous and perfectly deterministic for all workers and perfectly observable by the firm: $a_{it} = \mu + a_{it-1}$, where μ is a constant. Let us denote a_{i0} as an individual's ability at the beginning of her working career. ξ represents an element of productivity which is observable by the firm but not by the econometrician. It may relate to the quality of the machinery the worker will use. The variable κ represents worker characteristics which are unobservable to both the firm and the econometrician at any moment of her career. Therefore it is impossible for a firm to infer with complete certainty the correct value of a worker's ability from past output realisations.

Assuming that all firms have identical knowledge about a worker's productivity, competition determines that the worker will be paid her expected productivity at each point of her career. Consider that Ω_{it} is the information set of a firm that employs worker i with experience t. Under the assumption of symmetric information across firms, this information set will be common to all and the expectation of a worker's overall productivity will be identical, thus implying:

$$w_{it} = \gamma + \alpha s_i + \log E[\exp(\beta a_{it})|\Omega_{it}] + \frac{1}{2}\sigma_\kappa^2 + \xi_{it}.$$

Expectations about a worker's ability can be decomposed into a deterministic experience-related component and a component relating to the expectation on the initial ability.

$$\begin{aligned} E[\exp(a_{it})|\Omega_{it}] = E[\exp(a_{i0} + \mu t)|\Omega_{it}] &= \exp(\mu t) E[\exp(a_{i0}|\Omega_{it}] = \\ &= \exp(\mu t + E[a_{i0}|\Omega_{it}]) \cdot E[\exp(v_{it})|\Omega_{it}]. \end{aligned}$$

A firm's expectations at time t about a worker's original ability will be rational and $z_{it}^e = E[a_{i0}|\Omega_t]$, with the initial ability decomposed as $a_{i0} = E[a_{i0}|\Omega_t] + v_{it}$. This implies that $E[v_{it}|\Omega_t] = 0$ for all values of t. However, as the information on a worker's ability becomes more accurate through learning from past output realisations, the variance of the expectation error will diminish, i.e. $\sigma_{v_t}^2 > \sigma_{v_{t+1}}^2$. Wage determination will then imply:

$$w_{it} = \gamma + \frac{1}{2}\sigma_\kappa^2 + \frac{\beta^2 \sigma_{v_t}^2}{2} + \mu\beta t + \alpha s_i + \beta z_{it}^e + \xi_{it}.$$

A crucial element of the identification of the learning process is based on the relationship between Ω_t and the econometrician's information set Ω_t^{Ect}. It is strictly necessary that there is a part of the latter's information set not included in the employer's information set. To formalise the learning process, imagine that the true initial ability can be decomposed into a series of mutually independent additive components $a_{i0} = \sum_J \varsigma_{ij}$. Consider now that a firm's initial knowledge of a worker's ability is $z_{i0}^e \equiv E[a_{i0}|\Omega_0] = \sum_{j \in J_0^f} \varsigma_{ij}$. Learning implies that more terms become revealed to the firm and the expectation becomes $z_{it}^e \equiv E[a_{i0}|\Omega_t] = \sum_{j \in J_t^f} \varsigma_{ij}$. An empirically derived proxy for ability could be represented as $z = \sum_{j \in J_0^{ec}} \varsigma_{ij}$. If all the members of J_0^{ec} are contained in J_0^f, it is impossible to find out anything about employer learning. Otherwise, with excluded components at an initial stage, learning tells us that the correlation between $E[a_{i0}|\Omega_t]$ and z_i will increase. Thus, the econometrician's proxy for expected ability becomes more accurate. This implies that the measurement error problem in the behavioural relationship between log wages and schooling and expected ability is less important.

Because the bias persists, there is little hope that the marginal productivity of ability can be separately identified from that of schooling. But this is not the aim of the exercise. Evidence of employer learning can support the existence of a value in schooling other than human capital, which would play a substantial role in the information set Ω_t. Even though this might be true in the absence of effective employer learning, it is very unlikely that employer learning takes place about a worker's ability information he or she is completely unaware of.[34] The learning hypothesis suggests that measurement error will prevail but diminish as the worker becomes more experienced. This leads to a range of hypothesis that can be tested. A prediction of the model is that estimates of the effect of expected ability should be less attenuated as experience increases. An interaction term of z_i with experience would identify this effect given the previous assumptions. Thus, changes in the estimated returns to ability and schooling will have the following probability limits.

$$plim(\hat{\beta}_{t+1} - \hat{\beta}_t) = \beta \left(\frac{\sigma^2_{u_{t-1}}}{\sigma^2_{u_{t-1}} + \sigma^2_{z^e_{t-1}}(1 - \rho^2_{z^e_{t-1}s})} - \frac{\sigma^2_{u_t}}{\sigma^2_{u_t} + \sigma^2_{z^e_t}(1 - \rho^2_{z^e_t s})} \right), \tag{8}$$

$$plim(\hat{\alpha}_{t+1} - \hat{\alpha}_t) = \beta \left(\frac{\sigma_{z^e_t}}{\sigma^2_s} \frac{\sigma^2_{u_t}}{\sigma^2_{u_t} + \sigma^2_{z^e_t}(1 - \rho^2_{z^e_{t-1}s})} - \frac{\sigma_{z^e_{t-1}}}{\sigma^2_s} \frac{\sigma^2_{u_{t-1}}}{\sigma^2_{u_{t-1}} + \sigma^2_{z^e_{t-1}}(1 - \rho^2_{z^e_{t-1}s})} \right). \tag{9}$$

It is possible to show that the estimated coefficient on ability z_i will increase with experience. The measurement error of real expected ability will decrease as explained above. Similarly, the variance of firm's expected ability will increase and its correlation with schooling will diminish, thus implying an unambiguous increase in the returns to ability with experience.

As for the coefficient on schooling, the opposite will hold. We should expect a reduction in the estimated returns to schooling with experience. As the contractual relationship unfolds, the variance of v_i will obviously go down, and so will the variance of the error made by the econometrician when considering z_i instead of z^e_i. The coefficient on z_i will be less under-estimated and the coefficient on years of education less over-estimated. Then, if we consider the estimated coefficients at different stages $T = 0$ and $T = 1$, we will expect to find that $\hat{\beta}_o(1) > \hat{\beta}_o(0)$ and $\hat{\alpha}_o(1) < \hat{\alpha}_o(0)$.

[34] At least if we are willing to presume rational behaviour on her part. It is indeed possible that individuals have a *wrong understanding* of what their true skills are.

B Appendix: Not intended for publication

B.1 General details on data used

Following the completion of the latest surveys carried out in 1999/2000 in two of the main British longitudinal cohort studies, NCDS and BCS70, for the first time it is possible to undertake a full comparison of these two cohorts in terms of their early progression into the labour market after the legally established minimum school leaving age.

The National Child Development Survey, although originally devised as a Perinatal Mortality Survey of the cohort born in Britain in the week 3-9 of March 1958, went on to record information on the children and their families and school environments at the ages of 7, 11 and 16. Further follow up studies were undertaken in 1981 (NCDS4), 1991 (NCDS5) and 1999/2000 (NCDS6). Each survey contained retrospective questions on the individual's previous track record of economic activities, qualifications, earnings and many other aspects of her life.

The British Cohort Study of 1970 is a longitudinal cohort study of British children born between 5 and 11 April 1970. The 1999/2000 questionnaire provides the most detailed available source of information on these individuals as adults, although a survey based on a postal questionnaire was carried out in 1996 and a reduced sample was interviewed in 1992. The 1999/2000 questionnaire was structured identically to the sixth follow-up of the National Child Development Survey (NCDS6), although referring to different time periods in the respondents' lives.

Childhood data in the 1970 cohort was obtained through the British Birth Study (1970), the 1975 Child Health and Education Studies (in 1975 and 1980). The Youthscan Study interviewed cohort members at the age of 16 (i.e. 1986). Childhood data was collected from very different sources. Parents would be interviewed and asked to report on certain aspects of their child's behaviour as well as on related family and their own individual characteristics.

Producing work history data was probably the most complicated task. Substantial effort was devoted to creating a reconciled version of work history data, available from a very limited number of surveys, which are subject to attrition, with data collected through radically different methods and with generally poor measures of the transition from school to the labour market. This work involved lots of cross-validation exercises, comparing longitudinal reports and through the use of fertility and education-related data. The reconciled data provides a record of an individual's main economic activity status in every month since her/his sixteenth birthday. Experience measures were computed from this activity vector for each survey.[35]

With respect to the earnings data, I have used cleaned data cleaned by members of the Institute of Fiscal Studies whenever possible. Gross hourly earnings deflated to prices as of January 2001 have been converted to hourly wages using reported and checked usual working hours. This applies for all follow-ups in the NCDS and the BCS at age 30. Wage data for the BCS reduced sample at 21 was generated by following this same pattern. Only net earnings were reported in the Postal Self-Completion Questionnaire administered to BCS cohort members in

[35] The guidance of Peter Dolton and Gerry Makepeace was essential for carrying out the construction of individual work histories.

1986. Given the secondary importance of this follow-up because its reduced response rates and the lack of net data on earnings for most of the other surveys, I chose to convert such earnings to gross amounts using the limited available information. Mario Fiorini, from IFS, provided me with a code that simulated the tax system in 1986 and converted earnings from net to gross. We invoked the assumption of revealed preference in choosing working hours to resolve the non-monotonicity problem in inverting the theoretical relationship between gross earnings and tax paid.

Other data on job characteristics was similarly defined across surveys, although reports on training received from employer are more difficult to compare.

The indicator of behavioural adjustment is used as a proxy for non-cognitive abilities. It is based on the Bristol Social Adjustment Guide (BSAG). Teachers were asked to indicate whether sample children scored in a series of tests comprising various syndromes. These syndromes refer to the child's capacity to adjust to different social environments and circumstances and reflect aspects such as hostility and depression. The antisocial index follows from adding up the number of items underlined by the teachers and then standardising the total score. Higher values represent lower adaptability. NCDS cohort members were tested at age 11 and BCS children at 10.

The tests which have been used to derive measures of cognitive ability are as follows: For NCDS at age 11, children were administered five tests, namely, (i) 'General ability test' (Douglas), with verbal and non-verbal sections, (ii) 'Reading comprehension test' (NFER special), (iii) 'Arithmetic and Maths test' (NFER special), (iv) 'Copying designs test'. Moving on to the BCS children, they were assessed at age 10, when they were administered (i) 'British ability tests', (ii) 'Shortened Edinburgh reading test' and (iii) 'Youngs Maths test'.

The wage data collected in the surveys corresponds in all cases but one to gross wage earnings.[36] Unfortunately, in the BCS26 questionnaire individuals were asked about their net earnings, which implies a strong comparability problem. This paper uses a tax simulation model developed by Mario Fiorini (Institute of Fiscal Studies) that uses all available information in the questionnaire to compute the expected tax paid by every individual and thereby retrieve the original gross wage.[37]

Gross hourly wages are deflated to January 2001 prices. Table 7 shows average wages and standard deviations for each age/year group. Average wages appear to grow with age/time for each cohort and the same applies to the dispersion as measured by the standard deviation.

The right side of table 7 illustrates the problems underlying selective attrition in the data. The cognitive ability index has zero mean and unit standard deviation in the sample where it is computed. Adult observations included in the sample are more likely to have been drawn from the upper part of the ability distribution. This is due to a number of factors. First, the differentiated attrition pattern implies that more able individuals are more likely to remain part

[36] In many instances, individuals are also asked about their net wages.

[37] This tax simulation model imputes the married couples allowance to the male partner, given the lack of information on the partners' earnings that would allow to understand the couple's optimal use of the allowance.

Table 7: Log wages and cognitive ability in NCDS/BCS70

	Log wages				Cognitive ability: g			
	year				year			
age	1981	1991	1996	2000	1981	1991	1996	2000
21		1.8225				.0053		
		(.3833)				(.8913)		
23	1.7429				.0752			
	(.3502)				(.9661)			
26			1.9795				.2801	
			(.4689)				(.9441)	
30				2.1264				.1477
				(.4756)				(.9937)
33		2.2179				.1994		
		(.4329)				(.9538)		
42				2.3324				.1950
				(.5142)				(.9560)

NOTE: Average log real gross hourly wage and average cognitive ability index. Standard deviation within parentheses.

of the study as they are interviewed as adults. Continuity is also important, as individuals discontinuously interviewed will have longer spells of missing work history data and hence have dropped-out of the sample used here. Second, higher ability tends to support participation in the labour market and so the probability of obtaining a valid wage observation.

For NCDS, the sample seems to be not far initially from the whole population average nil value for g, although successive follow-ups improve the ability composition, making the sample less representative of the male population. In BCS, the smaller size sample at age 21 appears to reproduce the ability characteristics of the cohort, but the very opposite conclusion follows from the examination of BCS26. It seems in the last case that the use of a self-completion questionnaire strongly discouraged less able individuals. At age 30, I obtain an ability composition comparable to that at age 33 in the NCDS.

B.2 Principal components as a measure of g

This appendix briefly describes the extraction of principal components to derive a comparable index of cognitive ability for both cohorts. A first principal component is defined as the linear combination g_1 of the data matrix X, i.e. $g_1 = Xa_1$, such that it produces the best possible overall fit to the different variables x_k in X.

Assuming g_1 known, a hypothetical regression of vector x_k on g_1 would produce an estimation error $e_k = [I - g_1(g_1'g_1)^{-1}g_1']x_k$. Because a particular regression on a given variable is identical to the infinite number of variables which are proportional to this one, a normalisation is required. Therefore it is imposed that $g_1'g_1 = 1$. This implies that $e_k = [I - g_1g_1']x_k$ and a measure of the error made in approximating variable x_k is given by the sum of squared errors $e_k'e_k$. The

Table 8: Summary statistics

Variable	Mean	Std. Dev.	Min	Max
Real gross hourly wage	8.879	5.656	0.520	99.911
Schooling (years)	1.560	2.086	0.000	7.750
Cognitive ability	0.162	0.965	-3.114	2.520
Experience (years)	12.356	6.702	0.000	26.167
Tenure (years)	6.276	6.022	0.000	26.083
Training (EPT) in current job	0.471	0.499	0.000	1.000
Married/with partner	0.589	0.492	0.000	1.000
Occupation				
Professional	0.067	0.249	0.000	1.000
Intermediate	0.297	0.457	0.000	1.000
Skilled non manual	0.130	0.336	0.000	1.000
Skilled manual	0.296	0.456	0.000	1.000
Semi-skilled	0.104	0.305	0.000	1.000
Unskilled	0.021	0.142	0.000	1.000
Firm size				
1-10 emp	0.138	0.344	0.000	1.000
11-24 emp	0.129	0.335	0.000	1.000
25-99 emp	0.231	0.421	0.000	1.000
100-499 emp	0.229	0.420	0.000	1.000
500+ emp	0.198	0.399	0.000	1.000
Year dummies				
Year 1981	0.259	0.438	0.000	1.000
Year 1991	0.223	0.416	0.000	1.000
Year 1996	0.128	0.334	0.000	1.000
Year 2000	0.390	0.488	0.000	1.000

NOTE: Total of 14762 person-year observations, corresponding to 8163 different individuals. Sample restricted to individuals whose main occupation is paid work, with a gross hourly wage ranging between 50 pence and 100 pounds. Further restrictions involve valid work history information (less than six months missing from 16^{th} birthday to interview date) and ability test completion at 10 (BCS) or 11 (NCDS).

objective then is to minimise the aggregate error $\sum_{k \in K} e_k' e_k = trace(X'[I - g_1 g_1']X)$ subject to the normalisation constraint.

Since the objective is to retrieve the optimal a_1, it is now necessary to substitute into the objective function in order to derive g_1. The problem becomes:

$$min_{\{a_1,\lambda\}} trace(X'X) - trace(X'Xa_1a_1'X'X) - \lambda(1 - Xa_1a_1'X'),$$

where the first term can be ignored. This problem can be re-written as:

$$max_{\{a_1,\lambda\}} a_1'(X'X)(X'X)a_1 + \lambda(1 - a_1'X'Xa_1),$$

which leads to the first order condition:

$$(X'X)a_1 = \lambda a_1.$$

This is basically the definition of a_1 as the characteristic vector of matrix $X'X$. If values in the score matrix X have been previously standardised, $X'X$ is simply the correlation matrix. Changes of origin and scale in our problem are necessary in order to make data comparable across cohorts. The second order condition is immediately satisfied because $X'X$ is a positive definite matrix.

Now the problem is to ascertain which is the sought-after characteristic vector. Standard results from matrix algebra show that the number of possible eigenvectors is identical to the rank of matrix $X'X$. This will coincide with the number of tests administered if there is no perfect collinearity. To solve this, notice that after substitution from the f.o.c., the objective that needs to be minimised is given by:

$$a_1'(X'X)(X'X)a_1 = a_1'\lambda^2 a_1 = \lambda^2.$$

Then the highest eigenvalue associated with the correlation matrix will determine the appropriate characteristic vector. The remaining $K-1$ principal components can be easily calculated using the remaining eigenvalues and imposing that $g_m' g_n = 0$, for all $m \neq n$, i.e. they must be mutually orthogonal.

The principal components methodology appears to be a natural approach in order to use for deriving an index of cognitive ability for each survey.

Two main problems affect the derivation of an ability index in this cohort data. On the one hand, the tests used for each different cohort are not identical, although there is some overlapping on the attributes they intend to measure, such as maths and reading. On the other hand, there is the issue of whether an ability index can encompass all the information contained in a series of different test scores.

To minimise these problems, I followed a series of steps. I first removed within-cohort variation purely due to the fact that children have been tested in different months.[38] Deriving residuals from a regression of each test score variable on time was meant to put respondents' scores on a same level of comparison.[39] Using the residualised scores, the full set of principal components is calculated for each cohort.

The principal component extraction results are summarised in table 9.

Differences between the batteries of tests administered in NCDS and BCS may imply that g might mean different things for each cohort. In fact, there are more tests available for NCDS, which has an effect on the proportion of the observed total variation explained by the first principal component. Copying designs is frequently described as less correlated with cognitive ability, begin associated with more primary-type of abilities, such as visual and motor co-ordination. I have checked how robust g is to the exclusion of this test from the battery, and I have further experimented by aggregating verbal and non-verbal ability into a single value by adding them

[38]Those tested earlier have on average lower scores because their cognitive capacities are less developed than their peers interviewed months later.

[39]For example, all BCS respondents around the age of ten interviewed in different months become comparable this way.

Table 9: Principal component derivation of g

Pcpal Compt.	Eigenvalue	Standard Deviation	Cumulative variance	Original variables	Eigenvector (1st comp.)
NCDS-Age 11					
1	3.4256	1.85	0.69	Copying designs	0.26
2	0.8252	0.91	0.85	Verbal ability	0.50
3	0.3549	0.59	0.92	Non verbal ability	0.48
4	0.2280	0.47	0.96	Maths ability	0.49
5	0.1663	0.40	1.00	Reading ability	0.46
BCS-Age 10					
1	2.4784	1.57	0.82	Maths test	0.57
2	0.2668	0.51	0.91	Reading	0.58
3	0.2547	0.50	1.00	British Ability Scale	0.58

NOTE: Principal components calculated from age-residualised test scores, within every age-cohort group. Last two columns represent each test's association with the first principal component.

up.

Table 10: Alternative factor decompositions: NCDS11

g(4): Four scores only.			
Factor order	Contr.to var.	Variable	*g(4)* loadings
g(4)= 1	0.811	Verbal ab.	0.51
2	0.089	Non-verbal ab.	0.49
3	0.057	Maths	0.50
4	0.042	Reading	0.48
g(3): Three scores only.			
Factor order	Contr.to var.	Variable	*g(3)* loadings
g(3)= 1	0.844	Ver+non ver.ab.	0.58
2	0.091	Maths	0.58
3	0.064	Reading	0.56
Correlation matrix of alternative g's			
	$g(3)$	$g(4)$	$g(5)$
$g(3)$	1		
$g(4)$	0.9951	1	
$g(5)$	0.9862	0.9913	1

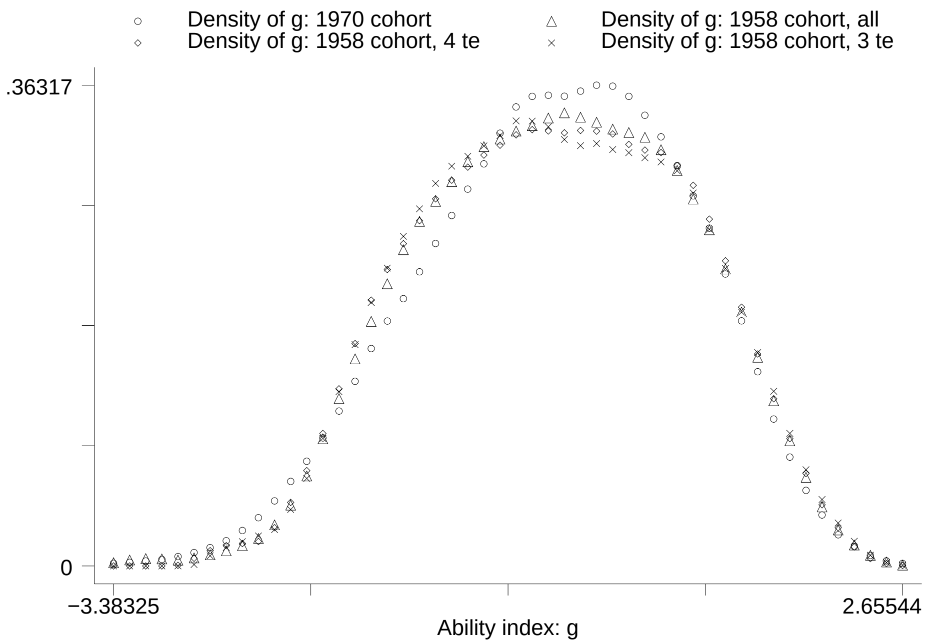
NOTE: Principal component analysis of reduced sets of ability scores in NCDS11. Proportion of variance explained $= \lambda_i / \sum_j \lambda_j$, where λ_j is the j^{th} eigenvalue. Factor loadings proceed from eigenvector associated to first eigenvalue in each decomposition. The correlation matrix presents correlations between g(3), g(4) and the first principal component using the five variables1.

Table 10 analyses changes in the derived g. It suggests that measured g is quite robust to these alternative specifications. Additionally, it must be noticed that with a three scores specification, BCS and NCDS are far more comparable in terms of (a) Number of scores used,

(b) Correlation between g and the original scores, (c) Proportion of total variation explained.

I have also checked whether the ability distributions in BCS and NCDS are significantly different from each other. Kernel density estimates in figure 1 basically overlap with each other, indicating they're almost identical.

Figure 1: Ability indexes at age 11/10



NOTE: Kernel density estimates of cognitive ability index distribution for BCS and NCDS (under three alternative specifications specified in table 10).

About whether it is valid to collapse all ability scores into a single index, there is an efficiency/unbiasedness trade-off. Because separate scores are so much correlated with each other, it is highly plausible that there is an underlying commonality to all of them, as suggested by the high proportion of variance explained by g in BCS and NCDS (three scores), above 80 percent, which is much higher than the 33 percent that would be implied by completely independent variables. Multi-collinearity at this scale may well remove all evidence from employer learning. Measurement error of g is also likely to attenuate ability-related estimates. In any case, this is a better approach than relying on maths scores which are not directly comparable either.[40]. In terms of explained wages, excluded components play a much minor role compared to g in both cohorts, particularly for BCS. Thus, this construct of g appears to be most suitable measure of cognitive ability that can be used from the available data.

[40]For example, emphasis on arithmetics is not identical, etc...

B.3 Controlling for differential sample selection

As noted, it is important to control for the selective composition of the samples suitable for estimation. One possible approach is to model the process that determines whether an individual enters the valid estimation sample in a given year using background childhood and family data for the earlier sweeps. If this process is adequately controlled for, it will be possible to obtain unbiased estimates of the earnings process that are relevant for the learning model.

Table 11: Determinants of participation in estimation panel

	NCDS-23	NCDS-33	NCDS-42	BCS-21	BCS-26	BCS-30
Ability index	.059	.140	.118	-.021	.215	.151
	(.028)	(.029)	(.029)	(.031)	(.019)	(.018)
Ability i. squared	-.063	-.049	-.037	-.109	-.037	-.015
	(.013)	(.014)	(.014)	(.023)	(.014)	(.013)
Lies at school	-.211	-.249	-.256	-.057	-.089	-.062
	(.052)	(.055)	(.055)	(.057)	(.039)	(.032)
Father's Social Class						
Professional	-.044	.191	.111	-.268	.276	.276
	(.101)	(.103)	(.103)	(.178)	(.117)	(.113)
Intermediate	.061	.229	.114	-.149	.243	.172
	(.066)	(.067)	(.067)	(.123)	(.084)	(.080)
Skilled non-manual	.157	.245	.067	-.165	.291	.324
	(.076)	(.076)	(.067)	(.134)	(.091)	(.087)
Skilled manual	.124	.184	.136	-.065	.209	.178
	(.052)	(.054)	(.054)	(.101)	(.072)	(.067)
Semi-Skilled manual	.400	.282	.231	-.059	.070	.087
	(.163)	(.160)	(.160)	(.118)	(.086)	(.080)
Unskilled	.163	.259	.198	-.370	-.203	-.021
	(.068)	(.069)	(.069)	(.181)	(.120)	(.105)
Pseudo-R2	0.04	0.05	0.05	0.02	0.04	0.02
Log-likelihood	-4797.3	-4625.3	-4594.2	-1351.9	-3523.9	-3941.7
Observations		NCDS: 7230			BCS: 5815	
Proportion	52.95	40.40	39.07	6.38	32.50	50.59

NOTE: Dependent variables equals one if individual is present in each adult survey and , given childhood information is available. Other controls include region at birth, parent's education level and school type. Reference father's social class group is father not working/missing. Probit coefficients reported. Standard errors displayed within parentheses.

Consider the probability that an individual i from cohort c participates in survey j as a function of a vector of characteristics $e_{i,c}$ such as $\Phi[\lambda_j e_{i,c}]$. According to the symmetric learning model, the expected log wage of this individual in survey j conditional on being part of that survey and providing valid data is determined by:

$$E[w_{i,c,j}|x_{ij}, s_i, z_i, present] = \gamma_j + \mu_j x_{ij} + \alpha_{0j} s_i + \alpha_1 x_{ij} s_i + \beta_{0j} z_i + \beta_1 z_i x_{ij} + E[\xi_{ij}|present],$$

where $E[\xi_{ij}|present] = \sigma_j \frac{\phi[\lambda_j e_{i,c}]}{\Phi[\lambda_j e_{i,c}]}$ under the assumption that $\Phi[\cdot]$ corresponds to the standard normal distribution cumulative function and σ is the covariance between the random component of participation and the heterogeneity term in the wage equation ξ_{ij}. Thus, in order to account for unobservable determinants of participation that might be correlated with the wage process, I include additional terms consisting of the estimated selection term, known as the 'Mills Ratio' $\frac{\phi[\lambda_j e_{i,c}]}{\Phi[\lambda_j e_{i,c}]}$. In order to instrument participation, I include a wide range of background variables which proved not to be significant if included in the main wage equation but proved significant in their prediction of participation in the survey. These include a quadratic in the ability index, parental class and education information and whether the teacher had reported that the individual, as a child, used to lie at school. Controlling for ability, the assumption of conditional independence required is more likely to be satisfied in practice than in a data set with more limited information. Table 11 displays the coefficients of key variables in the selection equation, suggesting a sufficient degree of correlation of the excluded variables with the participation outcome. It is interesting to see that adult attrition is not monotonic in ability, with individuals at the extremes of the distribution less likely to be part of the final estimation sample. Anticipating some of the wage results, the selection term turns out to be not statistically significant in either of the wage equations estimated in this chapter.

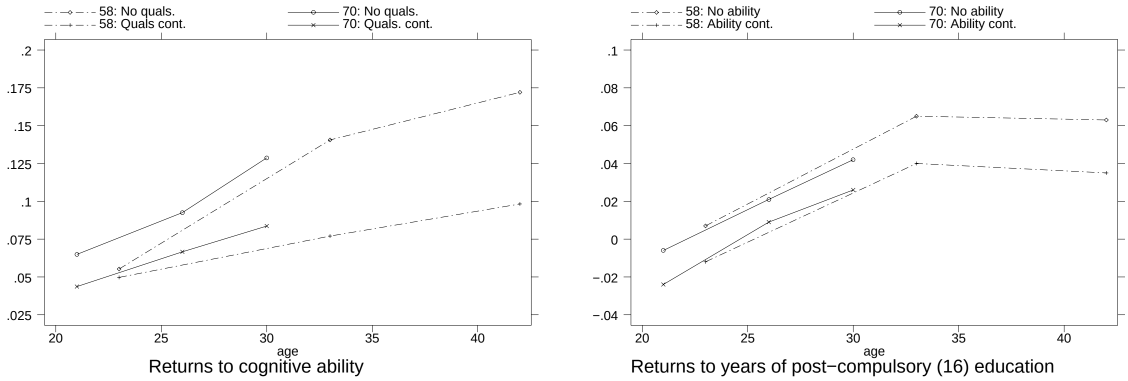
B.4 Returns to ability and schooling over the lifetime

Variation in potential experience amongst identically educated individuals is mostly driven by age differences. It is therefore interesting to investigate how returns to ability and education vary over the lifetime. I carry out a graphical interpolation exercise that can be considered as a less restrictive imposition on the data than an arbitrary functional form. Figure 2 displays the point estimates of returns to g reported on its left-hand-side. Returns estimates for the same cohort are joined with a straight line. The figure on the right hand side documents the estimated returns to years of full time education after the age of compulsory schooling.

The severity of the missing data problem cannot be denied. There is no common age support for age values higher than 30 and younger than 23. Having only two cohorts, time effects are restricted to twelve-year long intervals (the difference in birth dates for both cohorts) that must refer to relatively young workers (ages 23 to 30). However, there are important features that emerge from such limited data, such as a clear life-cycle pattern of increasing returns to ability (over the set of observed age values) which is common to both cohorts.

The left part of figure 2 shows that the fitted ability returns/age profile for the later 1970 cohort always exceeds that for the 1958 NCDS cohort. I discuss the two specifications separately. Firstly, without qualification controls, the gap is marginally wider when the cohorts are in their early twenties, which corresponds to a comparison of the early 1980s with the early 1990s. According to the literature, this is the period when returns to education grew faster (Gosling *et al* (2000)). Thus, in order to make inferences about changes in this period one must rely on two relatively young cohorts, when returns to both education and ability are at their lowest. The gap in ability returns between the cohorts actually becomes smaller as age grows and the time comparison must refer to the late nineties versus the late eighties. This coincides with the

Figure 2: Returns to cognitive ability and years of education over lifetime: By cohort



NOTE: Returns to measure of cognitive ability displayed by cohort and two specifications: (1) Including parental social class and education controls. (2) As in (1), including detailed individual qualifications. Returns to years of post compulsory (after 16) full time education displayed by cohort and two specifications. (1) With parental controls. (2) As in (1), including measures of cognitive and non-cognitive ability. All returns are based on OLS regressions of log hourly wages for working males in each survey.

deceleration in the upward trend in returns to education which was reported during the nineties.

After controlling for education, the gap in ability returns between the cohorts increases monotonically and the returns can be well fitted with a straight line. Cognitive ability is increasingly rewarded in the labour market as an individual grows older, and this pattern has become marginally stronger in the 1970 cohort.

I also display estimated returns to schooling by cohort under alternative specifications (right hand side). With and without ability measures, returns to schooling appear to increase at least until individuals are in their mid-thirties. For very young individuals, returns to education are not significantly different from zero because more educated individuals have only just joined the labour market and are compared to more experienced workers. Neglecting ability data leads to higher estimates of the returns to education. No significant differences over time can be found between cohorts for either specification (with and without ability controls).

In terms of the symmetric learning model, the implication from these figures is that β_1 is unambiguously positive, suggesting the existence of employer learning. However, α_1 is also positive although its appears to become zero after the early thirties for the NCDS cohort.