Advances in Life Course Research 22 (2014) 1–14

Contents lists available at ScienceDirect

Advances in Life Course Research

journal homepage: www.elsevier.com/locate/alcr

Explaining patterns in the school-to-work transition: An analysis using optimal matching

Richard Dorsett, Paolo Lucchino *

National Institute of Economic and Social Research, United Kingdom

ARTICLE INFO

Article history:
Received 3 October 2013
Received in revised form 29 April 2014
Accepted 25 July 2014

Keywords:
School to work transition
Youth unemployment
NEET
Optimal matching

ABSTRACT

This paper studies the school to work transition in the UK with the aim of achieving a richer understanding of individuals' trajectories in the five years after reaching school leaving age. By applying the technique of 'optimal matching' on data from 1991 to 2008, we group individuals' trajectories post-16, and identify a small number of distinct transition patterns. Our results suggest that while 9 out of 10 young people have generally positive experiences post-16, the remaining individuals exhibit a variety of histories that might warrant policy attention. We assess the extent to which characteristics at age 16 can predict which type of trajectory a young person will follow. Our analysis shows that, despite the apparent heterogeneity, virtually all at-risk trajectories are associated with a relatively small set of key 'risk factors': early pregnancy; low educational attainment and self-confidence; and disadvantaged family background. These characteristics are known to be strongly correlated across individuals and raise concerns about the degree of socio-economic polarisation in the transition from school to work.

1. Introduction

Shifting social and economic conditions over the last three decades in Britain and indeed globally have diminished the centrality of the traditional route of early school leaving and rapid entry into employment (Bynner, 2001; Pollock, 2007). Trajectories have become more individualised, with educational attainment gaining an increasing importance in shaping young people's life-chances and exposing the lowest-achieving young people – often the poorest – to greater vulnerability. A large body of literature documents the social polarisation in the transition from school to work (Dickerson & Jones, 2004; Micklewright, 1989; Rice, 1999; Spielhofer, 2009).

While the effects of disadvantage on labour market outcomes are similar across countries, they are particularly marked in the UK (Ryan, 2001). Indeed, while youth unemployment hit a record high in the wake of the recent recession, the UK youth labour market had started to deteriorate as early as 2004. The reasons for this are not well understood (Goujard, Petrongolo, & Van Reenen, 2011), but there appears to be a structural problem in the transition from school to work. Some young people fail to find work after leaving school and spend a substantial amount of time Not in Employment, Education or Training (NEET). As argued by Fergusson, Pye, Esland, McLaughlin, and Muncie (2000), the experiences of many young people beyond compulsory education do not follow stable and 'traditional' trajectories, but complex ones across multiple states.

This paper studies a sample of young people reaching the end of compulsory schooling between 1991 and 2003 in the UK and traces their pathways over the following five years, covering the period up to 2008. It uses

* Corresponding author at: National Institute of Economic and Social Research, 2 Dean Trench Street, Smith Square, London SW1P 3HE, United Kingdom.
E-mail address: p.lucchino@niesr.ac.uk (P. Lucchino).

http://dx.doi.org/10.1016/j.alcr.2014.07.002
1569-4909/

a powerful statistical approach – optimal matching combined with cluster analysis – to identify groups of young people following similar pathways, capturing the full richness of individuals' experiences beyond school leaving age. In doing so, it provides an alternative to commonly used statistics that summarise outcomes at a point in time (e.g. the unemployment rate) or over a specified period (e.g. time spent unemployed in the previous year) but discard potentially interesting information on labour market dynamics (such as the order in which events occur). Using this technique, the experiences of individuals within each group can be depicted using colour-coded index plots. This gives an immediate visual insight into the patterns of transition within each, allowing us to distinguish, for example, transitory 'gap years' from deep disconnect from the labour market.

The analysis in this paper builds on earlier research that used optimal matching to study the school to work transition in the UK (Anyadike-Danes & McVicar, 2005, 2010; Halpin & Chan, 1998; Martin, Schoon, & Ross, 2008; Schoon, McCulloch, Joshi, Wiggins, & Bynner, 2001) and in a comparative perspective (Brzinsky-Fay, 2007; Quintini & Manfredi, 2009; Scherer, 2005). However, most of this literature relies on long retrospective histories which may suffer from recall bias (Paull, 2002). Brzinsky-Fay (2007) and Quintini and Manfredi (2009) are exceptions to this, but use data on youth histories only up to 2000 and 2001, respectively. We therefore add to the existing literature by considering detailed monthly histories extending to 2008 and constructed from annual survey data to minimise recall bias. Finally, in consideration of recent methodological advances in the field of optimal matching (Martin & Wiggins, 2011), we take care to ensure our use of the technique is suitably justified by theory.

As the second contribution of the paper, we identify which characteristics at age 16 can act as early predictors of unsuccessful trajectories in the labour market. While a number of papers have examined the influence of background characteristics on outcomes at later points in time, the strength of our approach is that it uses the groupings identified in the first part of the analysis to provide an insight into how background characteristics are associated with successful or unsuccessful *overall trajectories* post-compulsory schooling. Importantly, we can also use this to measure the extent of social mobility in this crucial phase in the life-course.

Our results suggest that 9 out of 10 young people experience generally successful labour market trajectories between ages 16 and 21. These are predominantly smooth transitions from education to work, or long spells of education, in some cases interrupted by a spell of employment. On the other hand, the remaining individuals exhibit a variety of histories that might warrant policy attention. Importantly, however, our subsequent analysis shows that, despite this heterogeneity, virtually all at-risk trajectories are associated with either early pregnancy or low educational attainment and self-confidence. Policy should therefore give particular attention to targeting these factors. As our observation period fully predates the recent recession, these structural obstacles will not necessarily be surmounted by a return to economic growth. Furthermore, our analysis confirms the importance of family background as a strong predictor of future labour market trajectories, thereby contributing to a significant level of socio-economic polarisation.

2. Creating a typology of school to work transitions

We explore the unfolding of school to work transitions by creating a typology of youth labour market histories (or *sequences*). This consists of two steps. Firstly, we use optimal matching techniques to construct a measure of dissimilarity between each pair of sequences (Abbot & Forrest, 1986; Sankoff & Kruskal, 1983). Secondly, we apply cluster analysis techniques to the derived measures of dissimilarity to group similar sequences together.[1]

The optimal matching algorithm performs a pairwise comparison of all individuals' sequences and, in each case, derives a measure of dissimilarity as a function of the number and type of operations on the elements of one sequence that are necessary to transform it into the other. The costs assigned to each operation determine how dissimilarity is defined in the context under study, and hence how sequences are matched. Specifying costs is important as it may influence the results that emerge. The literature does not set rigid rules on this. However, it is possible to parameterise the cost matrix to make it consistent with theoretically informed definitions of what constitutes similarity in the context under study (see Martin & Wiggins, 2011, for a review).

In analysing post-compulsory school histories, we follow Lesnard (2010) and make the following considerations. Firstly, our sequence represents the five academic years after the end of compulsory schooling, and as such is set within a clear socio-economic 'calendar'. There is a strong element of contemporaneity across sequences (e.g. summers occur at the same points in all sequences). For this reason, we retain this contemporaneity by not allowing insertions or deletions. This requires having sequences of the same length. Furthermore, the institutional set up of the further education system is likely to shape observed patterns of transition around key dates (e.g. A-level exams). To address this we use time-varying substitution costs defined as the inverse of the conditional transition probability at the specific point in the sequence, as described above. This distance measure is called the *dynamic Hamming distance*.

Having derived measures of dissimilarity, cluster analysis techniques can be used to group similar sequences together. Deciding the clustering algorithm and the number of groups requires careful consideration. We opted for the classic Partitioning Around Medoids (PAM) algorithm (Kaufman & Rousseeuw, 1990, Chapter 2), which minimises the sum of dissimilarities between each sequence and its group's medoid, and ran this for k target groups ranging from 2 to 20. Our final choice of the number of clusters was in part guided by a comparison of statistical

[1] The dissimilarity matrices were obtained using the TraMineR package (Gabadinho, Ritschard, Studer, & Müller, 2009) in R (R Development Core Team, 2008).

indices of fit and in part by the desire to describe the main patterns in the data with sufficient granularity. The average Silhouette width (Rousseeuw, 1987) and the Calinski–Harabasz index (Calinski & Harabasz, 1974) both favoured a number of groups towards the lower end of the range considered. However, a very small number of clusters failed to provide significant socio-economic insight when explored visually. Instead, we found that partitioning into a larger number of groups provided a richer picture of the dynamics at play. We ultimately selected 16 groups, as the average Silhouette width reached a local maximum of 0.416. The Calinski–Harabasz index is 644.

The combination of optimal matching with cluster analysis is a powerful statistically driven technique that can synthesise large amounts of information from complex sequences and categorise these into relatively homogenous groups. The strength of optimal matching lies in its holistic nature, as its algorithm draws on information from the full set of elements in a sequence. It therefore overcomes limitations of other commonly used statistics, which generally summarise outcomes at a point in time or over a specified period, discarding important information on labour market dynamics. Instead, optimal matching allows histories to be compared in their full dynamic richness, including the type, length, order and timing of spells. We can thus distinguish, for example, between school to work transitions characterised by short difficulties and those that are suggestive of more deep-rooted problems.

However, there is a risk that purely statistical techniques may lack sociological meaning (Elzinga, 2003; Levine, 2000; Morgan & Ray, 1995; Wu, 2000) and of being determined by arbitrary choices of the researcher (Everitt, Landau, Leese, & Stahl, 2011; Wu, 2000). For this reason, we try to make an informed choice in setting the costs, which define how the algorithm should conceive of similarity. Nevertheless, while the technique will satisfy the specified numerical optimality conditions, whether the resulting typology does in fact have an objective socio-economic significance or the extent to which this meaning may be attributed subjectively ex-post by the researcher remain open questions. We recognise this element of subjectivity and therefore caution the reader from taking our descriptions of the groups identified as absolute. However, the plausibility of the results presented below, and our confidence that these will be consistently interpreted by the majority of observers, strengthens our belief that these techniques have significant descriptive power and are capable of identifying patterns that genuinely exist in the data, and hence in society.

3. Data

We use data from the British Household Panel Survey (BHPS), a longitudinal survey which followed a nationally representative sample of households at yearly intervals from 1991 to 2008. The design of the survey is such that children within sampled households become eligible for adult interviews once they turn 16, and are interviewed annually thereafter. We focus on such children and their trajectories over the five years after they reach school leaving age. Given the extent of the BHPS, the individuals in our sample reach school leaving age between 1991 and 2003, so that the fifth anniversary of the last cohort coincides with the end of the BHPS in 2008.

We constructed a month-by-month history for each young person, following the careful methodological studies by Paull (2002) and Maré (2006). Indeed, the BHPS consists of a main questionnaire about circumstances at the time of interview and a job history module where individuals recall their employment and activity history over the previous 12–18 months. This recall period may overlap with information given at the previous interview and any inconsistencies present in this overlap need to be reconciled. We do so following the reconciliation techniques provided in Maré (2006). An advantage of focusing on those turning 16 during the survey is that we observe their full labour market histories without having to rely on long-term respondent recall. Since interviews take place roughly annually, 97% of the months covered in the life-work histories rely on recall of 14 months or less. This is an important consideration as Paull (2002) finds that individuals with the most transient behaviour, in many cases the very people of most interest, have great difficulty accurately recalling their prior experiences. Relying on recall periods of about one year keeps this potential bias to an absolute minimum.

As we are interested in status biographies covering periods of employment as well as non-employment, we follow Paull's (2002) 'main activity' definition of status. This is defined according to the individuals' own identification of their main activity from a list of 10 available choices.[2] We grouped these responses into four high-level labour market states: 'employment', 'full-time education', 'NEET – unemployed' and 'NEET – not active in the labour market'. We split the conventional definition of NEET to better understand whether different reasons for non-employment lead to distinct trajectories. Inevitably, this approach has some limitations. Firstly, there will be an element of subjectivity in the responses, which may also vary across individuals (Paull, 2002). Secondly, this measure does not capture activities carried out concurrently, such as employment and full-time education. These cases will be treated as being in only one of the two, depending on the individual's own view of which best describes their situation. For these reasons, our reconciliation analysis (not shown) finds that the histories tend to slightly overestimate educational participation and underestimate official youth employment rates although they track Department for Education NEET rates closely. Finally, we do not have information on part-time education. Overall, however, the data provide a rich description of the activities of the young people in the sample and can provide important insights into their labour market experience.

[2] These were: self-employed, employed, unemployed, retired, maternity leave, family care, full-time student, long-term sick/disabled, Government training scheme, and other.

We restrict our attention to the 1352 individuals observed for five consecutive years starting from the month they could legally leave school. As mentioned previously, having sequences of the same length is necessary when calculating the *dynamic Hamming distance*. This implies restricting our attention to individuals who are observed in the survey for the full five years. To account for possible non-randomness of remaining in the survey, we estimate a probability model of attrition within five years and restore cross-sectional representativeness by adjusting each young person's BHPS cross-sectional weight at the point they can legally leave school.

4. Results

4.1. Visualising trajectory types

As each sequence consists of 60 elements (one for each month), each taking one of four values (employment; full-time education; NEET – unemployed; and NEET – inactive), individual histories can be represented as a horizontal series of colour-coded dots. Stacking the series for all individuals in each group creates an index plot, which gives an immediate visualisation of the general labour market dynamics characterising that group. In the visualisations that follow, we assign titles to represent our interpretation of the trajectories in each group, and include the share of individuals assigned to such group. In all cases, the horizontal axis starts at the end of compulsory schooling (Y0) and covers the following five years (Y5).

Overall, we identified 16 groups, which can themselves be grouped into three high-level categories. As noted above, we find the cluster visualisations reveal fairly clean and socio-economically plausible trajectories that, we believe, would be immediately recognisable by the majority of observers. This should reduce the concern that these are determined by the choice of optimal matching or clustering approach. Nevertheless, as a sensitivity test, we re-ran the clustering using the Ward minimum variance method, which groups sequences into a target number of clusters to minimise the variance within each, and found no notable difference in the results.

The first set of groups is presented in Fig. 1. It presents index plots for the six groups depicting smooth transitions from education to work. In line with the naming given to similar groupings found in Brzinsky-Fay (2007) and Quintini and Manfredi (2009), we called these 'Express' education to work transitions, with the number in parentheses indicating the number of additional years of education before the transition occurs.

Five further groups describing predominantly educational trajectories are shown in Fig. 2, and we jointly refer to these as the 'Accumulating human capital' category. The last in this set describes individuals who stay in education throughout, while individuals in the remaining four groups also spend substantial time in education interrupted by one or two academic year(s) in employment. We call these 'Full-time education' and 'Full-time education with an employment spell', respectively.

The remaining individuals exhibit a variety of histories that might warrant policy attention. Their trajectories are depicted in Fig. 3 and, together, they make up the 'Possible cause for concern' category. Noting the small sample size of each group, we suggest the following characterisations. The groups in the middle row describe individuals experiencing some employment but developing only limited labour market attachment ('Partial recovery') or exhibiting patterns of long-term worklessness straddling unemployment and sometimes inactivity ('Long-term worklessness'). The top two groups consist of individuals in long-term inactivity from the age of 16 ('NEET from 16') or 18 ('NEET from 18'), while the last plot portrays individuals who appear to withdraw from the labour market following an apparently successful entry into employment ('Withdrawals from the labour market'). Virtually all those experiencing these last three trajectories are female and in most cases mothers by age 21. This result is confirmed when running optimal matching separately for males and females (not shown). While all other groups emerge when considering each of the two subsamples separately, these three groups are only identified for females. This point reinforces the importance of qualifying the description of these groups as giving rise to a *possible* cause for concern. In many cases, those 'NEET – inactive' trajectories may be so through voluntary choice, despite possible detrimental effects on labour market progression. However, to the extent any such choice is involuntary or constrained, there may still be a legitimate role for policy.

Looking at the share of youth assigned to each group, the results indicate that 9 out of 10 young people experience generally successful labour market trajectories (that is, they are in either the 'Express' or the 'Accumulating human capital' categories), while the remaining 1 in 10 exhibit one of the above-mentioned trajectories that might warrant policy concern.

The types of trajectories identified above show a broad agreement with previous research using similar approaches. Indeed, both Brzinsky-Fay (2007) and Quintini and Manfredi (2009) identify typologies that closely resemble some of the histories grouped above. While their analyses focus on several countries, both provide an estimate of the size of each group in the UK. As above, they find that the 'Express' pathway is by far the most common. While they too find that each of the other typologies refers to a minority of individuals, their 'problem' groups tend to be somewhat larger. This may be reconciled by the different observation period used: their sequences are defined from when the individual first leaves full-time education, while ours start from the end of compulsory schooling. We do this because we consider the choice to stay in education an integral part of one's transition from the world of education to that of work. It is possible, therefore, that individuals who we find as exhibiting a stable trajectory over the five years we consider may later move on to less fortunate trajectories once they leave the education system.

4.2. Predicting future labour market outcomes

The influence of early experiences and characteristics on later labour market outcomes is of crucial importance both to understand who is most at risk of an unsuccessful

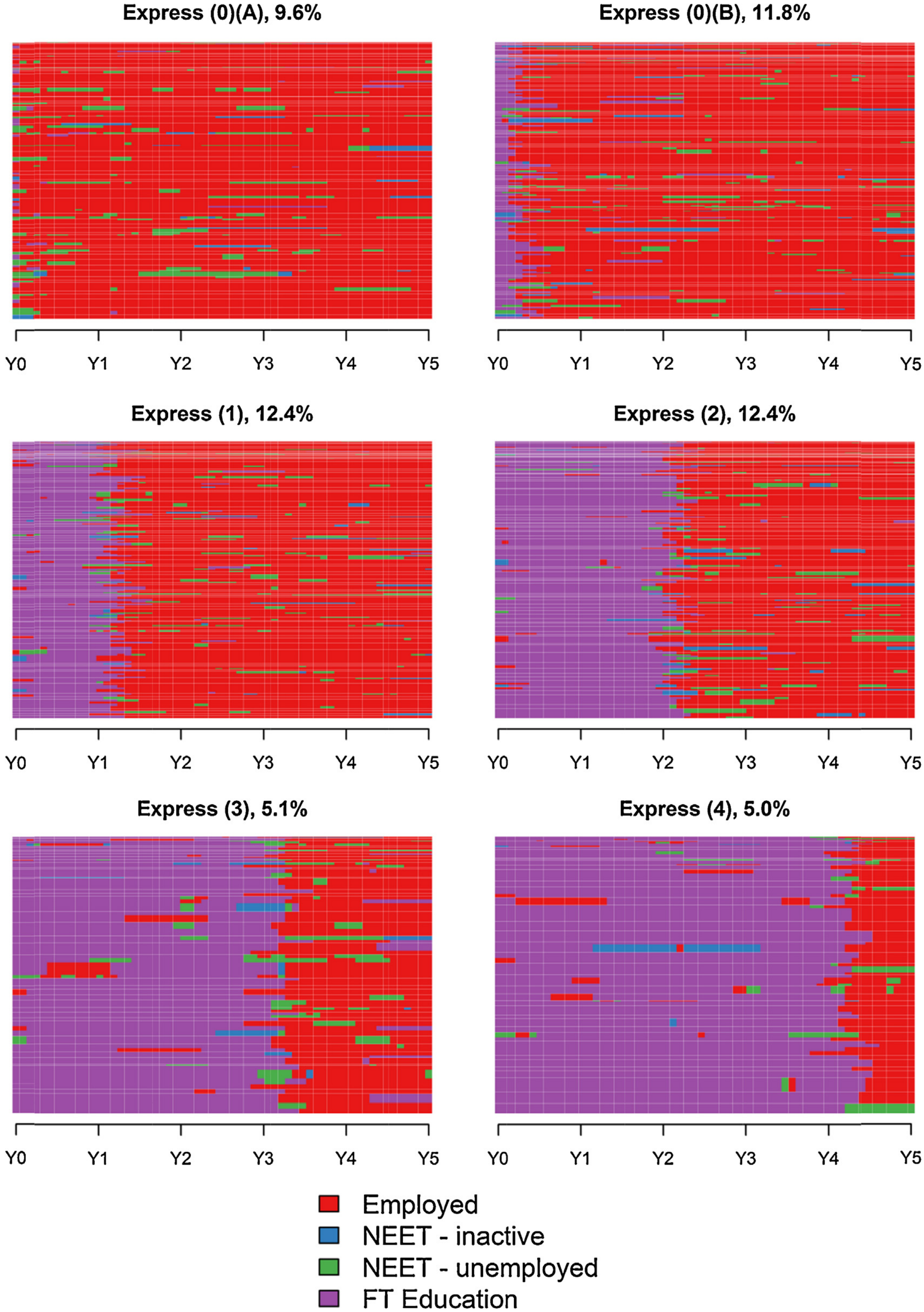

Fig. 1. 'Express' trajectories.

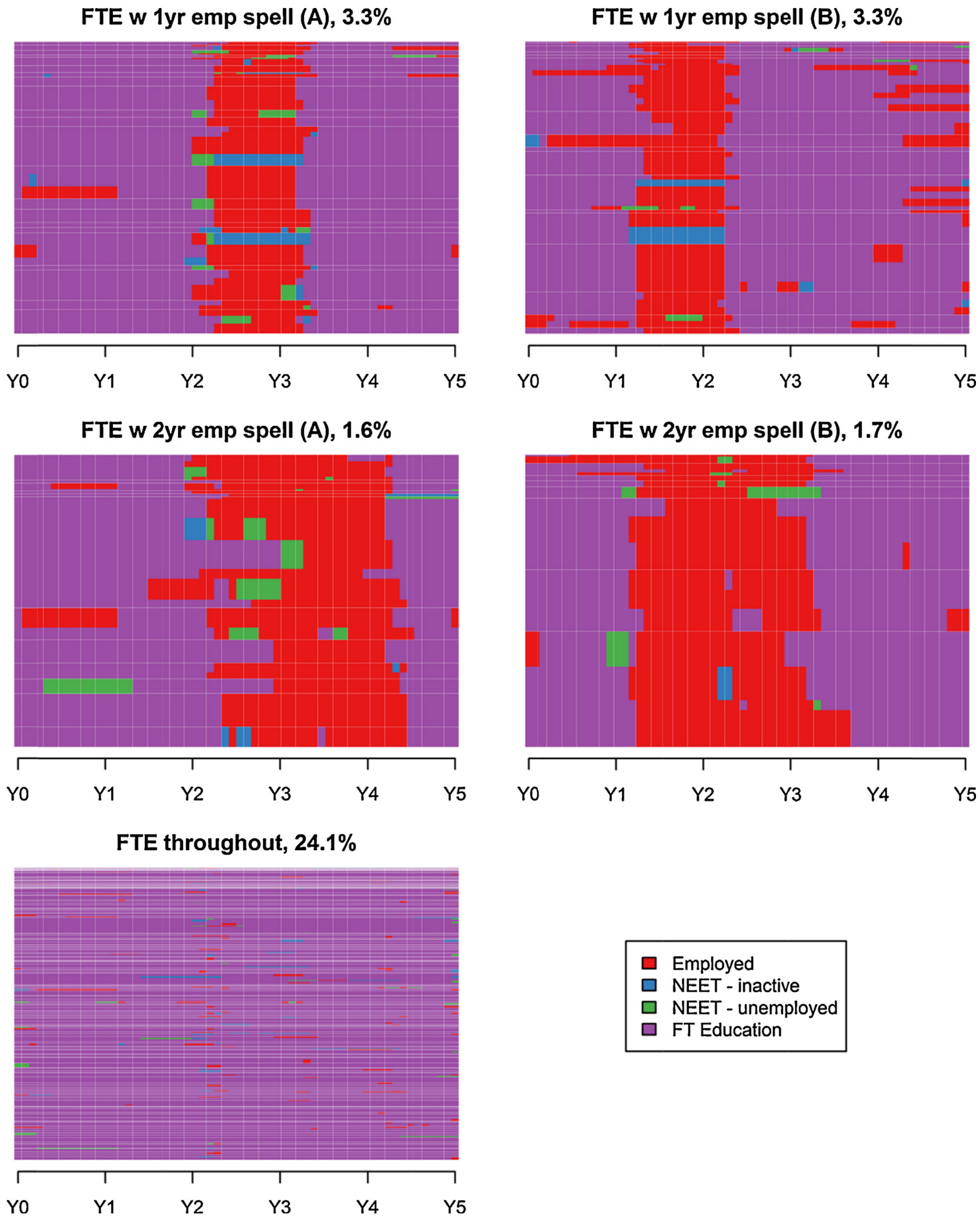

Fig. 2. Human capital trajectories.

transition and the potential for individual mobility given one's starting conditions. The existing literature on the issue, however, has tended to describe outcomes as measured at a specific point in time rather than in a more holistic manner. By making use of our identified typology of trajectories, we can test for statistical correlations between characteristics at age 16 and nature of an individual's *overall trajectory* over the five years after compulsory schooling.

We use statistical techniques to understand the influence of distinctive characteristics at age 16 on an individual's future labour market trajectory. Note that the emphasis here is on identifying predictive markers at age 16. We are therefore not concerned with identifying causal links as much as we are about establishing robust correlations. The aim is, partly, predictive. However, this should not be interpreted as a practical tool for policy targeting. This would go beyond the scope of this paper

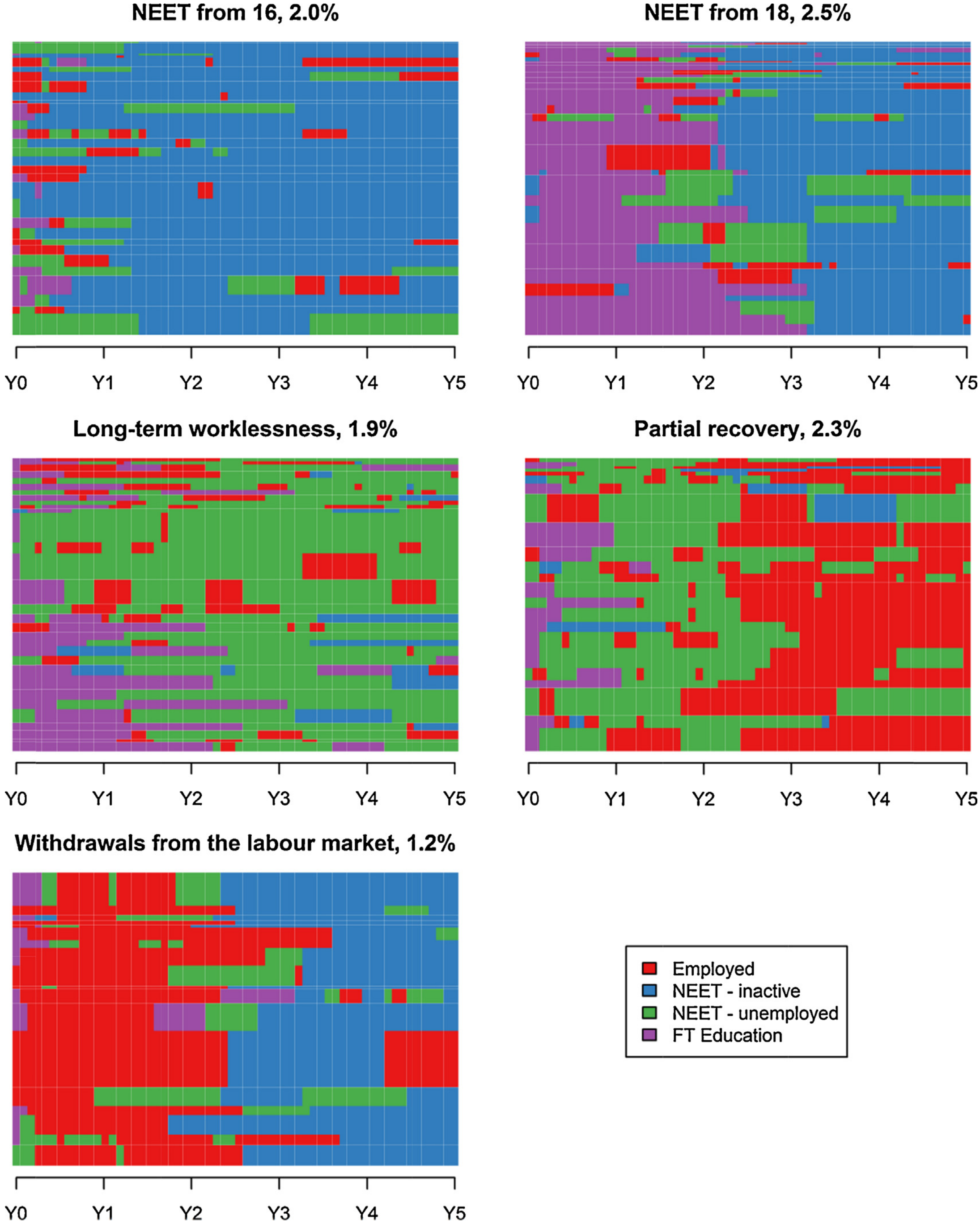

Fig. 3. Possible cause for concern trajectories.

and the accuracy that can be achieved from the general-purpose survey data we use here. Rather, the aim is to evaluate the strength of the link between starting conditions and future outcomes, thereby gauging the scope for social mobility over this period of the life-course.

With this in mind, our model includes a fairly standard set of observable individual characteristics, particularly in relation to education (gender, ethnicity, grades and receipt of grants for education); and family background (parental and sibling employment, parental qualifications, housing tenure and income quartile). We control for local labour market conditions by including the deviation of the district level 16–64 year old claimant count rate from the national mean, and net out cohort and time effects by including the individual's year of birth as a categorical variable.[3] We also include a set of covariates to attempt to capture the increasing influence of personal and attitudinal characteristics such as self-confidence and motivation in shaping the school to work transition. These factors are notoriously difficult to measure, and the chosen covariates should only be interpreted as proxies. These include relative age (via month of birth); health condition and disabilities; responses to questions on attitudes and subjective well-being from the reduced version of the General Health Questionnaire module included in the BHPS; and smoker status. The later is chosen as an indirect measure of a fairly disparate set of personal circumstances, including risky behaviours, lifestyle, stress, depression/distress, self-esteem, attitudes, and health concerns, which have been found to be correlated with adolescent smoking (Tyas & Pederson, 1998).

We used a multinomial logit model to estimate the probabilities of belonging to each of the three high-level categories mentioned above: 'Express' transition into work, 'Accumulating human capital' or 'Possible cause for concern'.[4] Due to the small sample size of those in each of the 'concern' pathways, we had to treat these as a single category rather than analyse each group individually. Prediction of cluster membership is strongest when clusters are made up of very similar sequences. While our clustering outcome is relatively homogenous overall, the index plots do reveal a small number of sequences that may not fit exactly within their cluster. We ensure our results are robust to the exclusion of such sequences by re-running the model without these borderline cases. We do this in three different ways: following Piccarreta (2012), we drop the 5% of sequences that are most distant from their cluster medoids; we drop sequences with silhouette width less than or equal to zero; or we reassign such sequences to their 'neighbouring' cluster (Rousseeuw, 1987). The results that follow are robust to all three approaches.

As is common with non-linear estimators, the magnitude of the effect of a given variable cannot be read directly from the estimation coefficients but instead needs to be calculated at a given set of values of the explanatory variables. It is therefore possible to estimate the percentage point change in the probability of an individual with a given characteristic (as opposed to some reference value for that same characteristic) entering a specific trajectory. Average marginal effects can be obtained by averaging these estimates across all individuals in the sample. These are presented in Table 1. For example, the table indicates that the probability that an individual whose most highly educated parent holds a degree enters a human capital trajectory is, on average, just under 25 percentage points higher than for an otherwise identical individual whose parents' highest qualifications are at most GSCEs graded D–G.

A number of demographic and background characteristics emerge as strong predictors of subsequent labour market pathways. As suggested in the previous section, being female is associated with a higher probability of being in the 'Possible cause for concern' group. Non-white youth are found to be more likely to enter a human capital trajectory.[5] Similarly to what is found in Crawford, Dearden, & Meghir (2010), those born between September and December, who are therefore the oldest in their year, exhibit more successful transitions to employment and fewer experiences of unsuccessful trajectories. Importantly, school attainment and family background (parental qualifications and housing tenure) emerge as the particularly strong predictors in the model. These results are consistent with evidence indicating that high-achieving and advantaged individuals will tend to move successfully along the available structured pathways, particularly in relation to the education system (Andrews & Bradley, 1997; Bynner, 2001; Dickerson & Jones, 2004; McVicar & Rice, 2000; Rice, 1999). Interestingly, when included alongside the above background and environmental characteristics, life-limiting health conditions or disabilities, receipt of education grants, family income and parental employment emerge as having at most weak predictive power.

The effect of the labour market status of older siblings is less straightforward to interpret. On the one hand, human capital theory would predict that having siblings will lead to lower investment in education as family resources are spread more thinly (Becker & Lewis, 1973). Where significant, our results are generally consistent with this prediction and other empirical evidence on the issue (Björklund, Eriksson, Jäntti, Oddbjørn, & Österbacka, 2004; Hanushek, 1992). However, part of this effect could be driven by a strong correlation between parental unobserved heterogeneity and fertility decisions, and alternative estimation techniques have in fact questioned these results (Angrist, Lavy, & Schlosser, 2010; Cáceres-Delpiano, 2006). Furthermore, our results differentiate according to the labour market status of the older siblings and hint at a correlation across sibling status, possibly evoking a role-model effect.

Individual trajectories are found to not be statistically associated with the local claimant count rate at age 16. While perhaps counter-intuitive, this is in line with the wider evidence on this issue, which has been mixed and, where present, impacts have mainly concerned the outcomes of young males with lower qualifications (Meschi, Swaffield, & Vignoles, 2011; Rice, 1999).

We also attempt to examine the relationship between 'soft' personal characteristics (such as attitude, self-confidence and motivation) and labour market trajectories. Such non-cognitive skills have been found to affect wages, years of schooling, future employment status, job type and levels of supervision on the job (Waddell, 2006). Drawing

[3] Note that the structure of our dataset implies that cohort and calendar year are collinear and cannot be identified separately.

[4] The estimations were carried out using the software STATA 12.

[5] Unfortunately, it was not possible to analyse sub-groups within the non-white population due to the small sample size.

Table 1
Age 16 marginal effects on future trajectory outcomes. Change in probability of entering the named trajectory when exhibiting a given characteristic compared to the reference value.

	Express	Human capital	Possible cause for concern
Individual characteristics			
Sex (ref: males)			
Female	−0.06 [0.025]*	−0.01 [0.023]	0.07 [0.014]***
Ethnicity (ref: white)			
Non-white	−0.18 [0.059]**	0.18 [0.056]***	−0.01 [0.030]
School attainment (ref: 5 + GCSE A*–C)			
1–4 GCSE A–C	0.32 [0.034]***	−0.34 [0.030]***	0.02 [0.018]
GCSE D-G	0.19 [0.037]***	−0.26 [0.034]***	0.07 [0.019]***
No qualifications	0.06 [0.051]	−0.22 [0.048]***	0.15 [0.033]***
Receipt of grants for education (ref: no)			
In receipt	0.00 [0.061]	−0.05 [0.056]	0.05 [0.033]
Month of birth (ref: May–Aug)			
Jan–Apr	0.08 [0.031]**	−0.08 [0.028]**	−0.01 [0.017]
Sept–Dec	0.09 [0.031]**	−0.06 [0.029]*	−0.03 [0.016]*
Health (ref: no limitations)			
Health limits daily activities	0.10 [0.059]	−0.08 [0.056]	−0.01 [0.024]
Smoker (ref: non-smoker)			
Smoker	0.04 [0.035]	−0.12 [0.033]***	0.09 [0.019]***
Count of 'negative' GHQ responses	−0.02 [0.005]**	0.02 [0.005]**	0.00 [0.003]
Parental and background characteristics			
Parental qualifications (ref: GCSE D–G)			
High (degree)	−0.19 [0.050]***	0.25 [0.045]***	−0.06 [0.032]
Medium (>GCSE A–C)	−0.09 [0.034]**	0.14 [0.031]***	−0.05 [0.017]**
Employment of household head (ref: not employed)			
In employment	0.06 [0.035]	−0.03 [0.033]	−0.03 [0.017]
Sibling labour force status (ref: no siblings)			
Empl/training	0.08 [0.034]*	−0.08 [0.031]**	0.01 [0.017]
NEET	0.01 [0.071]	−0.08 [0.067]	0.07 [0.035]*
FT student	0.04 [0.038]	−0.01 [0.033]	−0.03 [0.022]
Housing tenure (ref: owned)			
Social rented	0.09 [0.039]*	−0.12 [0.037]**	0.03 [0.018]
Private rented	0.24 [0.063]***	−0.22 [0.056]***	−0.01 [0.030]
Household income quartile (ref: top)			
Bottom	−0.02 [0.045]	−0.01 [0.041]	0.04 [0.025]
Second	0.06 [0.038]	−0.09 [0.033]**	0.03 [0.023]
Third	0.04 [0.038]	−0.06 [0.033]	0.02 [0.023]
Labour market context			
Local area claimant count rate dev (16–64)	0.00 [0.007]	0.00 [0.007]	0.01 [0.004]
Cohort dummies (year of birth)	YES	YES	YES
Number of observations		1282	
Log likelihood		−825	

Standard errors in brackets.
* Significant at the 5% level.
** Significant at the 1% level.
*** Significant at the 0.1% level.
In all groups of categorical values, the coefficient on at least one category for one outcome is statistically significant, implying the same for categorical variable as whole.

on responses to the reduced version of the General Health Questionnaire module included in the BHPS, which covers questions on attitudes and subjective well-being, we construct a count variable indicating the number of 'negative' responses given by our sample of 16-year-olds to the eight questions in the survey.[6] Results appear to confirm that self-confidence and motivation problems show lasting associations with future outcomes, increasing the probability of 'Accumulating human capital' but reducing the probability of being in the 'Express' category. Our model also considers the proxy variable of being a smoker at age 16. This is found to be associated with a strong reduction in the probability of entering a human capital trajectory and an increase in the risk of falling within the 'Possible cause for concern' category. Interpreting these estimates is complex. While undoubtedly not the direct impact of smoking per se, these are suggestive of the possible detrimental effects of lifestyles and attitudes that have been found to be correlated with adolescent smoking. The inclusion of these variables may also contribute to capturing the impact on labour market

[6] Results are robust to an alternative specification that uses factor analysis to construct variables capturing the pattern of variation in responses to these questions.

Table 2
Age 16 and youth marginal effects on future trajectory outcomes. Change in probability of entering the named trajectory when exhibiting a given characteristic compared to the reference value.

	Express (subsample)	Express (subsample + youth variables)	Human capital (subsample)	Human capital (subsample + youth variables)	Possible cause for concern (subsample)	Possible cause for concern (subsample + youth variables)
Individual characteristics						
Sex (ref: males)						
Female	−0.016 [0.041]	−0.058 [0.042]	−0.052 [0.037]	−0.052 [0.038]	0.068 [0.023]**	0.11 [0.026]***
Ethnicity (ref: white)						
Non-white	−0.29 [0.078]***	−0.26 [0.081]**	0.228 [0.080]**	0.211 [0.080]***	0.062 [0.058]	0.049 [0.060]
School attainment (ref: 5 + GCSE A*–C)						
1–4 GCSE A–C	0.334 [0.051]***	0.325 [0.052]***	−0.353 [0.047]***	−0.33 [0.048]***	0.019 [0.026]	0.005 [0.025]
GCSE D–G	0.193 [0.063]**	0.163 [0.065]*	−0.297 [0.058]***	−0.265 [0.061]***	0.104 [0.037]**	0.102 [0.036]**
No qualifications	0.036 [0.076]	0.049 [0.076]	−0.204 [0.070]**	−0.189 [0.070]**	0.168 [0.058]**	0.14 [0.055]*
Receipt of grants for education (ref: no)						
In receipt	−0.025 [0.096]	−0.034 [0.095]	−0.068 [0.086]	−0.045 [0.089]	0.093 [0.060]	0.079 [0.057]
Month of birth (ref: May–Aug)						
Jan–Apr	0.09 [0.048]	0.089 [0.048]	−0.081 [0.044]	−0.072 [0.044]	−0.01 [0.027]	−0.016 [0.027]
Sept–Dec	0.152 [0.052]**	0.158 [0.051]**	−0.098 [0.048]*	−0.09 [0.048]	−0.053 [0.026]*	−0.068 [0.025]**
Health (ref: no limitations)						
Health limits daily activities	0.062 [0.092]	0.07 [0.093]	−0.065 [0.084]	−0.044 [0.087]	0.003 [0.048]	−0.026 [0.041]
Smoker (ref: non-smoker)						
Smoker	−0.102 [0.060]	−0.132 [0.060]*	−0.068 [0.057]	−0.028 [0.059]	0.17 [0.042]***	0.16 [0.036]***
Count of ‘negative’ GHQ responses	−0.009 [0.009]	−0.007 [0.009]	0.016 [0.008]	0.017 [0.009]	−0.007 [0.005]	−0.01 [0.004]*
Parental and background characteristics						
Parental qualifications (ref: low)						
High (degree)	−0.134 [0.077]	−0.139 [0.077]	0.14 [0.068]*	0.111 [0.070]	−0.006 [0.046]	0.028 [0.047]
Medium (>GCSE A–C)	−0.091 [0.057]	−0.082 [0.059]	0.084 [0.054]	0.054 [0.056]	0.006 [0.025]	0.027 [0.023]
Employment of household head (ref: not employed)						
In employment	0.017 [0.058]	0.022 [0.058]	0.005 [0.055]	0.005 [0.055]	−0.022 [0.025]	−0.028 [0.026]
Sibling labour force status (ref: no siblings)						
Empl/training	0.116 [0.053]*	0.101 [0.054]	−0.123 [0.049]*	−0.129 [0.049]**	0.007 [0.027]	0.028 [0.028]
NEET	0.031 [0.109]	0.029 [0.108]	−0.137 [0.105]	−0.13 [0.106]	0.105 [0.058]	0.101 [0.050]*
FT student	0.063 [0.060]	0.064 [0.058]	−0.049 [0.052]	−0.052 [0.052]	−0.014 [0.035]	−0.011 [0.034]
Housing tenure (ref: owned)						
Social rented	0.115 [0.063]	0.096 [0.063]	−0.152 [0.058]**	−0.149 [0.059]*	0.037 [0.031]	0.054 [0.032]
Private rented	0.343 [0.079]***	0.294 [0.101]**	−0.347 [0.051]***	−0.344 [0.055]***	0.003 [0.060]	0.05 [0.087]
Household income quartile (ref: top)						
Bottom	−0.09 [0.071]	−0.098 [0.071]	0.024 [0.065]	0.021 [0.065]	0.066 [0.039]	0.077 [0.041]
Second	0.063 [0.059]	0.064 [0.059]	−0.087 [0.052]	−0.081 [0.053]	0.024 [0.033]	0.017 [0.032]
Third	0.009 [0.058]	0.012 [0.057]	−0.044 [0.051]	−0.049 [0.050]	0.035 [0.033]	0.037 [0.033]

Labour market context						
Local area claimant count rate dev (16–64)	0.027 [0.016]	0.026 [0.016]	−0.02 [0.015]	−0.02 [0.015]	−0.008 [0.008]	−0.006 [0.008]
Cohort dummies (year of birth)	YES	YES	YES	YES	YES	YES
Youth attitudes and experiences						
Truancy (ref: none)						
Plays truant		0.169 [0.051]***		−0.181 [0.047]***		0.012 [0.026]
Discipline (ref: none)						
Expelled/suspended or has vandalised property		0.102 [0.050]*		−0.015 [0.045]		−0.086 [0.037]*
Index of agreement with traditional gender roles		0.01 [0.024]		0.007 [0.023]		−0.017 [0.012]
Means a lot to do well at school (ref: agrees)						
Does not agree		0.039 [0.107]		−0.019 [0.105]		−0.02 [0.032]
Bullying (ref: does not worry about it)						
Is worried about it		−0.083 [0.045]		0.012 [0.042]		0.07 [0.027]**
Number of observations			530			
Log likelihood (subsample; subsample + youth variables)			−342; −324			

Standard errors in brackets.
* Significant at the 5%level.
** Significant at the 1% level.
*** Significant at the 0.1% level.
Columns labelled ‘subsample’ re-estimate the model in Table 1 on the subsample of individuals for whom we have sufficient data to estimate the extended model which includes additional youth characteristics. Such model is then estimated in the columns labelled ‘subsample + youth variables’.

choices of personal traits that would otherwise be unaccounted for in the model and could therefore bias the estimated effect of other variables. It is thus interesting to see that, even when including non-cognitive and personality traits, family background, grades and gender still remain significant predictors of future labour market trajectories.

These results reinforce existing evidence indicating that outcomes are determined in part by factors, such as educational attainment, over which the individual has at least some influence, and others, such as family background and gender, which are predetermined. We can use the statistical model to gauge the extent to which starting conditions may lead individuals down a somewhat inescapable path. We define the model's predicted outcome for each individual as that which is assigned the highest predicted probability out of the three outcomes considered. By this definition, around 68% of cases are predicted correctly. Just over half of the 'Possible cause for concern' group is predicted correctly by the model, while an equal number are incorrectly assigned to this group. These estimates shed light on the significant weight of structural inequalities at age 16 in influencing later outcomes, and the extent they reproduce themselves over time. Furthermore, the effects of school attainment and parental education and employment will combine and reinforce each other as individuals exhibit more than one such 'risk factor'. Indeed, the actual degree of polarisation is likely to be higher than can be inferred from Table 1. Using the model results, we estimate that, while virtually no young males with at least 5 GCSEs at A*–C at age 16 and living with highly educated and employed parents will enter a 'Possible cause for concern' trajectory, this will be the case for almost one in four young males obtaining no GCSEs at 16 and living with unemployed parents holding low qualifications. While high-achieving females are also almost free from the risk of an unsuccessful trajectory, that will be the outcome for almost one in two disadvantaged females. Note however that, in this case, the role of family background may be confounded by the high incidence of early pregnancies in this group.

We further validate this result by testing whether the predictive power of environmental factors is reduced when additional information about the young person is included. To do this, we consider the responses given by our sample members when they were aged between 10 and 15 years and therefore eligible for the BHPS youth questionnaire. This allows us to introduce previous experiences of truancy, bullying, and disciplinary issues at school as well as responses to attitudinal questions on gender roles and schoolwork into our statistical model. Unfortunately, this additional information comes at the cost of a reduced sample size of 530 individuals. This is primarily due to the fact the youth questionnaire was only introduced in wave 4 of the BHPS and that the same questions were not asked every year. We therefore also re-estimate the main specification on the reduced sample, allowing us to distinguish changes in the estimates that are due to the changes in sample from ones caused by the new variables.

Results for these estimations are presented in Table 2. Firstly, it is interesting to note that the main specification estimated on the smaller sample delivers very similar point estimates to the results in Table 1. This suggests that the observable characteristics of the sub-sample are not systematically different from those in the full sample, and that the two are similarly representative of the youth population. Also, as one would expect, many variables lose statistical significance as the standard errors on the reduced sample estimation are larger. The addition of the youth questionnaire variable improves the explanatory power of the model (the likelihood ratio test of their joint significance has a p-value of 0.0003). However, most of the additional characteristics are not statistically significant. These characteristics are: whether the young person was ever expelled or suspended or ever vandalised property; an index picking up agreement with the statements "the family suffers if woman has full-time job", "the husband should earn, the wife stay at home" and "the man should be head of the household"; and (dis)agreement with the statement "it means a lot for me to do well at school". On the other hand, truancy is significantly associated with a lower probability of entering a human capital trajectory and a higher probability of an 'Express' transition into work, suggesting a mismatch between the individual and the school system rather than more complex personal issues. Vulnerability to bullying is associated with lower 'Express' transitions to work and a higher probability of being in the 'Possible cause for concern' category. Most notably, however, the addition of these characteristics does not contradict the overall results obtained from the main specification. The role of background and environmental factors is confirmed, indicating the presence of a strong polarisation across socio-economic contexts that may cast its effect across generations.

5. Conclusion and policy considerations

This paper used optimal matching combined with cluster analysis to identify groups of young people who are broadly similar with regard to their experiences beyond school-leaving age. It also used multinomial logit models to assess the extent to which characteristics and circumstances at age 16 can predict which group young people are likely to fit into.

The analysis was prompted in part by the recognition that there is no single school to work transition; rather, individuals increasingly vary in the pathways they follow post-16. Our results provide further confirmation of this complex pattern of transitions. The strength of our approach is that, through the use of optimal matching, young people's experiences can be compared in their full richness. This is crucial: while few young people follow identical pathways, optimal matching allows the degree of similarity to be quantified so that people with broadly comparable trajectories can be identified.

The results illustrate that the school to work transition is successful for most young people, at least over the five-year period considered in this study. The fact that more than half our sample appears to achieve a successful entry into employment augurs well for continued self-sufficiency in the longer run. Roughly a third of the sample spends most of the five-year observation period in full-time education, possibly with a short employment spell. For these individuals, the transition into employment is not observed. However, the investment in their human capital is likely to stand them in good stead relative to their less well-qualified peers.

The remainder of the sample – roughly one in ten – is observed to follow pathways that are a 'Possible cause for concern' for policymakers. It is important to bear in mind that our observation period is entirely before the recent recession. As such, our results reflect structural issues and cannot be explained by cyclical factors. It is not the case that a return to economic growth will necessarily help these young people. Indeed, the type of trajectories of sustained detachment described here are not solely driven by the lack of job opportunities but appear to be determined by a more multidimensional distance between the individual and their environment. Bringing these young people back into the labour market is likely to require not just access to jobs (for example, via the Government's Youth Contract) but rather a more holistic and joined-up approach spanning across the relevant youth services (Barnes, Green, & Ross, 2011).

Effective policy must be sensitive to the fact that there is considerable heterogeneity among individuals who do not appear to manage a successful school to work transition. Despite this, two factors appear to characterise almost all of these trajectories. Virtually all cases in 'NEET from 16', 'NEET from 18' and 'Withdrawals from the labour market' groups are associated with early pregnancies. On the other hand, as was confirmed by the statistical analysis, individuals in the remaining groups share low educational attainment (grades) and signals of possible low self-confidence ('negative' GHQ responses, smoking and fear of bullying).

Moreover, the observed labour market patterns indicate that unsuccessful outcomes often start at key decision points in a youth's educational career (particularly at the end of compulsory schooling and at the end of two further academic years), suggesting this could be because of a poor decision taken at that point in time. Clear and accessible knowledge of options post-16 is therefore essential in minimising the risk of 'fractured transitions' – ending one activity without securing a stable outcome in the next (Coles, 1995; Furlong and Cartmel, 2004). This possibly highlights how effective career advice and job-search assistance programmes might be used to facilitate successful employee-employer matches.

Finally, however, the results also remind us of the extent to which such detachment from the labour market is only partly explained by individual choices and characteristics. Indeed, the statistical analysis confirmed the importance of family background characteristics as strong predictors of future labour market trajectories. These are known to be strongly correlated across individuals. As such, our results ring true with other evidence highlighting the significant, and possibly increasing, level of socio-economic polarisation characterising the transition from school to work. Policies helping youth from disadvantaged backgrounds to achieve a more successful transition will serve the broader aim of increasing social mobility.

Acknowledgement

The authors are grateful for the support received by the Nuffield Foundation (grant ref EDU/39082). The opinions expressed are those of the authors, and do not necessarily reflect those of the Foundation.

References

Abbot, A., & Forrest, J. (1986). Optimal matching methods for historical sequences. *Journal of Interdisciplinary History, 16*(3), 471–494.

Andrews, M., & Bradley, S. (1997). Modelling the transition from school and the demand for training in the United Kingdom. *Economica, 64*(255), 387–413.

Angrist, J., Lavy, V., & Schlosser, A. (2010). Multiple experiments for the causal link between the quantity and quality of children. *Journal of Labor Economics, 28*(4), 773–824.

Anyadike-Danes, M., & McVicar, D. (2005). You'll never walk alone: Childhood influences and male career path clusters. *Labour Economics, 12*(4), 511–530.

Anyadike-Danes, M., & McVicar, D. (2010). My brilliant career: Characterizing the early labor market trajectories of british women from generation X. *Sociological Methods & Research, 38*(3), 482–512.

Barnes, M., Green, R., & Ross, A. (2011). *Understanding vulnerable young people: Analysis from the longitudinal study of young people in England. 118. Research Report*. Department for Education.

Becker, G. S., & Lewis, H. G. (1973). On the interaction between the quantity and quality of children. *Journal of Political Economy, 81*(2), S279–S288.

Björklund, A., Eriksson, T., Jäntti, M., Oddbjørn, R., & Österbacka, E. (2004). Family structure and labor market success: The influence of siblings and birth order on the earnings of young adults in Norway, Finland, and Sweden. In M. Corak (Ed.), *Generational income mobility in North America and Europe*. Cambridge University Press.

Brzinsky-Fay, C. (2007). Lost in transition? Labour market entry sequences of school leavers in Europe. *European Sociological Review, 23*(4), 409–422.

Bynner, J. (2001). British youth transitions in comparative perspective. *Journal of Youth Studies, 4*, 5.

Cáceres-Delpiano, J. (2006). The impacts of family size on investment in child quality. *Journal of Human Resources, 41*(4), 738–754.

Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics – Theory and Methods, 3*(1), 1–27.

Coles, B. (1995). *Youth and social policy: Youth citizenship and young careers*. UCL Press.

Crawford, C., Dearden, L., & Meghir, C. (2010). *When you are born matters: The impact of date of birth on educational outcomes in England*. Institute for Fiscal Studies.

Dickerson, A., & Jones, P. (2004). *Estimating the impact of a minimum wage on the labour market behaviour of 16 and 17 years old*. Low Pay Commission.

Elzinga, C. H. (2003). Sequence similarity. *Sociological Methods & Research, 32*(1), 3–29.

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis*. Wiley.

Fergusson, R., Pye, D., Esland, G., McLaughlin, E., & Muncie, J. (2000). Normalized dislocation and new subjectivities in post-16 markets for education and work. *Critical Social Policy, 20*(3), 283–305.

Furlong, A., & Cartmel, F. (2004). *Vulnerable young men in fragile labour markets: Employment, unemployment and the search for long-term security*. Joseph Rowntree Foundation.

Gabadinho, A., Ritschard, G., Studer, M., & Müller, N. S. (2009). *Mining sequence data in R with the TraMineR package: A user's guide*. University of Geneva, http://mephisto.unige.ch/traminer

Goujard, A., Petrongolo, B., & Van Reenen, J. (2011). The labour market for young people. In P. Gregg & J. Wadsworth (Eds.), *The labour market in Winter: The state of working Britain*. Oxford University Press.

Halpin, B., & Chan, T. W. (1998). Class careers as sequences: An optimal matching analysis of work-life-histories. *European Sociological Review, 14*, 111.

Hanushek, E. A. (1992). The trade-off between child quantity and quality. *Journal of Political Economy, 100*(1), 84–117.

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis* (99th ed.). New York: Wiley-Blackwell.

Lesnard, L. (2010). Setting cost in optimal matching to uncover contemporaneous socio-temporal patterns. *Sociological Methods & Research, 38*(3), 389–419. http://dx.doi.org/10.1177/0049124110362526

Levine, J. H. (2000). But what have you done for us lately? Commentary on Abbott and Tsay: Sequence analysis. *Sociological Methods & Research, 29*(1), 34–40.

Maré, D. C. (2006). Constructing consistent work-life histories: A guide for users of the British Household Panel Survey. *ISER Working Paper Series*.

Martin, P., & Wiggins, R. D. (2011). *Optimal matching analysis*. Sage Publications Ltd.

Martin, P., Schoon, I., & Ross, A. (2008). Beyond transitions: Applying optimal matching analysis to life course research. *International Journal of Social Research Methodology, 11*(3), 179–199.

McVicar, D., & Rice, P. (2000). *Participation in further education in England and Wales: An analysis of post-war trends*. Economics Division, School of Social Sciences, University of Southampton.

Meschi, E., Swaffield, J. K., & Vignoles, A. (2011). *The relative importance of local labour market conditions and pupil attainment on post-compulsory schooling decisions*. Institute for the Study of Labor (IZA).

Micklewright, J. (1989). Choice at sixteen. *Economica, 56*(221), 25–39.

Morgan, B. J. T., & Ray, A. P. G. (1995). Non-uniqueness and inversions in cluster analysis. *Applied Statistics,* 117–134.

Paull, G. (2002). *Biases in the reporting of labour market dynamics*.

Piccarreta, R. (2012). Graphical and smoothing techniques for sequence analysis. *Sociological Methods & Research, 41*(2), 362–380. http://dx.doi.org/10.1177/0049124112452394

Pollock, G. (2007). Holistic trajectories: A study of combined employment, housing and family careers using multiple sequence analysis. *Journal of the Royal Statistical Society Series A, 170*, 167.

Quintini, G., & Manfredi, T. (2009). *Going separate ways? School to work transitions in the United States and Europe* OECD Publishing.

R Development Core Team (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, http://www.R-project.org

Rice, P. (1999). The impact of local labour markets on investment in further education: Evidence from the England and Wales youth cohort studies. *Journal of Population Economics, 12*(2), 287–312.

Rousseeuw, J. P. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53–65.

Ryan, P. (2001). The school to work transition: A cross-national perspective. *Journal of Economic Literature, 39*(1), 34–92.

Sankoff, D., & Kruskal, J. B. (1983). *Time warps, string edits, and macromolecules: The theory and practice of sequence comparison*. Addison-Wesley Pub Co Advanced Book Program.

Scherer, S. (2005). Patterns of labour market entry: Long wait or career instability? An empirical comparison of Italy, Great Britain and West Germany. *European Sociological Review, 21*(5), 427–440.

Schoon, I., McCulloch, A., Joshi, H., Wiggins, R. D., & Bynner, J. (2001). Transitions from school to work in a changing social context. *Young, 9*, 4.

Spielhofer, T. (2009). *Increasing participation: Understanding young people who do not participate in education or training at 16 and 17*. Department for Children, Schools and Families.

Tyas, S. L., & Pederson, L. L. (1998). Psychosocial factors related to adolescent smoking: A critical review of the literature. *Tobacco Control, 7*(4), 409–420. http://dx.doi.org/10.1136/tc.7.4.409

Waddell, G. R. (2006). Labor-market consequences of poor attitude and low self-esteem in youth. *Economic Inquiry, 44*(1), 69–97.

Wu, L. L. (2000). ''Some Comments on'' Sequence analysis and optimal matching methods in sociology: Review and prospect. *Sociological Methods & Research, 29*(1), 41–64.