NIESR Discussion Paper No. 414

Date: 18th October 2013

Name: Richard Dorsett and Paolo Lucchino

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

VISUALISING THE SCHOOL-TO-WORK TRANSITION: AN ANALYSIS USING OPTIMAL MATCHING

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.

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.

Visualising the school-to-work transition: an analysis using optimal matching

Richard Dorsett, National Institute of Economic and Social Research
Paolo Lucchino[1], National Institute of Economic and Social Research

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. (Micklewright, 1989; Dickerson and Jones, 2004; 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 et al., 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 et al. (2000), the experiences of many young people beyond compulsory education do not follow stable and 'traditional' trajectories, but complex ones across multiple states.

[1] Corresponding author: p.lucchino@niesr.ac.uk National Institute of Economic and Social Research
2 Dean Trench Street, Smith Square London, SW1P 3HE United Kingdom

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 an innovative 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 charts. 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 which used optimal matching to study the school to work transition in the UK (Halpin and Chan, 1998; Schoon et al., 2001; Anyadike-Danes and McVicar, 2005, 2010; Martin et al., 2008) and in a comparative perspective (Scherer, 2005; Brzinsky-Fay, 2007; Quintini and Manfredi, 2009). 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 and 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. The ability to know in advance who is at risk provides important clues as to the type of policy that might be effective and whom it should target. 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.

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 one 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. 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 (Sankoff and Kruskal, 1983; Abbot and Forrest, 1986). Secondly, we apply cluster analysis techniques to the derived measures of dissimilarity to group similar sequences together.

2.1 Using optimal matching to derive measures of dissimilarity

Optimal matching is a relatively novel technique in the social sciences. 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 operations allowed are insertion, deletion and substitution. Figure 1 gives examples of how the same two sequences could be reconciled in alternative ways.

In Panel A, Sequence B is transformed into Sequence A by using substitutions only. The approach, measuring what is known as the *Hamming distance*, retains the timing of events and measures dissimilarity as the number of elements that need to be substituted. Conversely Panel B shows how insertions and deletions can be used to reconcile the two sequences. In this case, the algorithm will try to align common subsequences. The resulting measure of dissimilarity will therefore be lower the more the sequences share common subsections. As such, this is known as the *longest common subsequence distance*. This measure emphasises the ordering of elements. However, the temporal dimension within a sequence may be altered as, when elements are deleted (inserted), neighbouring elements become temporally closer (more distant). This causes a 'warping' of time, which may not be suitable in certain research contexts. Falling between these two extremes, alternative measures of dissimilarity can be constructed using a combination of all types of operations. In this case, each operation is assigned a 'cost' it will add to the measure of dissimilarity. An arguably 'default' option is to set the cost of substitution to be equal to the cost of a deletion followed by an insertion, as these alternatives yield the same result.

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. A few considerations are relevant here.

Firstly, the relative importance of the timing of events compared to the order of events can be set through the cost assigned to insertion/deletion (*indel*) operations. As mentioned above, similarities within sequence subsections are emphasised by allowing *indels* to incur a low cost and may be appropriate when order is of most interest. This may be the case when studying, for example, the evolution of mental health or the structure of sentences. However, as *indels* cause a time-warp effect and break the contemporaneity between different sequences, they should incur a higher cost where timing is important. This will be the case when sequences are defined over a socio-economic 'calendar', which could be a very fixed temporal cycle (e.g. the working week), but also a somewhat looser institutional system (e.g. the higher education system).

Secondly, substitution costs can be assigned on the basis of the socio-economic proximity of different states. For example, depending on the research context, self-employment can be considered to be closer to employment than to inactivity (Anyadike-Danes and McVicar, 2005). In this case, substituting self-employment for employment may be viewed as incurring a lower cost than substituting self-employment for inactivity. Furthermore, the cost of a substitution may vary depending on where it occurs in the sequence. This variation can be set exogenously or be informed by the data. An example of the latter case could be setting the cost of a substitution to be inversely related to the conditional probability of a transition occurring at a given point in time. This distance measure is called the *dynamic Hamming distance* (Lesnard, 2006). More formally:

$$s_t(a,b) = \begin{cases} 4 - [P(X_t = a \mid X_{t-1} = b) + P(X_t = b \mid X_{t-1} = a) + P(X_{t+1} = a \mid X_t = b) + P(X_{t+1} = b \mid X_t = a)] & \text{if } a \neq b \\ 0 \text{ otherwise} \end{cases}$$

The cost of substituting *a* for *b*, or vice versa, at time *t* will be a declining function of the frequency of such a transition at that point in time. This is estimated from the data as the conditional probabilities of an *a* to *b* or *b* to *a* transition between the current and adjacent periods. This approach is used by Lesnard (2009) to analyse working patterns, as transitions between a 'not working' and 'working' tend to be clustered at key points during the day.

In analysing post-compulsory school histories, we 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 *indels*. 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.

2.3 Using cluster analysis to identify similar groups

Having derived measures of dissimilarity, cluster analysis techniques can be used to group similar sequences together. To do this, we follow the commonly used Ward minimum variance method, which groups sequences into a target number of clusters to minimise the variance within each. Deciding the number of groups requires careful consideration. We were in part guided by a comparison of statistical indices of fit for alternatives ranging from 2 to 20 groups, all of which favoured a number of groups towards the upper end of this range.[2] However, we opted for 14 groups as this captured the main patterns in the data while avoiding groups containing only a handful of observations.

2.4 An assessment of the approach

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

[2] These indices included: the ratio of the average distance within clusters to the average distance between clusters, the average Silhouette width (Rousseeuw, 1987) and the Calinski-Harabasz index (Calinski and Harabasz, 1974)

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. By avoiding the simplification that arises from relying on summary statistics, optimal matching has proved to be a very flexible technique. While its origins are in the study of DNA sequences, it has increasingly found a wide variety of applications, ranging from comparing status biographies such as employment careers and mental health histories, to comparing English folk dances or birdsong patterns (see Martin and Wiggins, 2011, for a review).

This growing popularity has not been without criticism (Wu, 2000; Elzinga, 2003; Levine, 2000). In particular, critics argue that while operations such as insertions, deletions and substitutions relate to actual chemical processes in a DNA strand, their meaning in a socio-economic context is less clear, and that the same would therefore hold of any resulting measure of distance. Furthermore, the lack of formal rules for defining the cost matrix has attracted criticism, as results may be determined by arbitrary choices of the researcher (Wu, 2000). Finally, it is worth pointing out that cluster analysis is also not free from limitations, such as the existence of multiple solutions when the data contain ties (Morgan and Ray, 1995), the sensitivity of the results to different cluster algorithms (Everitt et al., 2011) and the element of judgement required in the selection of the number of clusters. Optimal matching analysts have tended to respond to these criticisms by stressing that optimal matching, and the operations on the sequence, do not intend to model the actual transformation of social reality, but are simply a way of constructing a synthetic measure of difference from sequences containing very complex information (Abbott, 2000; Lesnard, 2006).

While optimal matching may arguably be no more subjective or partial than many other descriptive techniques, it is nevertheless worth giving due consideration to how the algorithm should be applied and the limitations of the results it delivers. 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.[i] 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.

We restrict our attention to the 1,352 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. An immediate visualisation of the general labour market dynamics characterising each group can thus be obtained by stacking the series for all individuals in that group. We present our interpretation of these dynamics in the visualisations that follow. In all cases, the horizontal axis starts at the end of compulsory schooling (Y0) and covers the following five years (Y5).

Overall, we identified 14 groups, which can themselves be grouped into three high-level categories. The first of these is presented in Figure 2. It plots sequences for individuals falling within one of the five groups experiencing 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.

Three further groups describing predominantly educational trajectories are shown in Figure 3, and we jointly refer to these as the 'Accumulating human capital' category. The first group describes individuals who stay in education throughout, while individuals in the remaining two 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 Figure 4 and, together, they make up the 'Possible cause for concern' category. As discussed below, each of these groups is smaller than the previous ones, making their visualisations more difficult to interpret. To help with this, we present information on the characteristics of individuals in each category in Table 1, with each of the groups within the 'Possible cause for concern' category shown independently. Noting the small sample size of each group, we suggest the following characterisations. The first group represents individuals who experience a (possibly planned) break from employment but may struggle to return to work (we label this group: 'Planned interruption?'). Two out of five in this group are mothers by the time they are 21. The next two groups describe individuals experiencing some employment but developing only limited labour market attachment ('Partial recovery') or exhibiting patterns of long-term worklessness straddling unemployment and inactivity ('Long-term worklessness'). The final three groups consist of individuals in long-term inactivity from the age of 16 ('NEET from 16') or 18 ('NEET from 18') and 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.

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.

Table 2 presents the final typology we identified and the size of each group. The results suggest 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 final column provides an estimate of the number of 16-year-olds entering each trajectory each year, based on Office for National Statistics mid-2010 Population Estimates.

4.2 Predicting future labour market outcomes

The ability to identify in advance who is at risk of an unsuccessful transition into the labour market will be of most interest to policy-makers as this will inform the type of policy that might be effective and whom it should target. We use statistical techniques to understand whether there are any distinctive characteristics at age 16 which could help predict an individual’s future labour market trajectory. There is a rich body of literature on the issue of how early experiences and characteristics affect labour market outcomes, such as employment and wages, later in life. However, this has tended to describe outcomes as measured at a specific point in time rather than in a more holistic manner. Instead, building on our identified typology of trajectories, we can identify statistical correlations between characteristics at age 16 and the success of one’s *overall trajectory* over the five years after compulsory schooling.

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’. 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. 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 3. 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 23 percentage points higher than for an otherwise identical individual whose parents’ highest qualifications are at most GSCEs graded D-G.

School attainment, family background (parental qualifications and housing tenure) and gender emerge as the strongest predictors. 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 (Bynner, 2001; Rice, 1999; Dickerson and Jones, 2004; McVicar and Rice, 2000; Andrews and Bradley, 1997). Interestingly, family income emerges as lacking any predictive power when included alongside these background and environmental characteristics.

A number of additional age 16 characteristics are also found to be statistically associated with subsequent labour market pathways, though their impact is mixed. For example, non-white youth are found to be more likely to enter a human capital trajectory.[ii] Similarly to what is found in Crawford et al.(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. Individuals with life-limiting health conditions or disabilities are much less likely to be in a predominantly educational trajectory, but, perhaps surprisingly, appear more likely to make an 'Express' transition into work. Finally, 16-year-olds living in local authorities with high unemployment rates relative to other areas are found to be more likely to enter a 'Possible cause for concern' trajectory and less likely to invest in human capital. Wider evidence on this issue has been mixed, but increasingly indicates that local labour market conditions can influence the outcomes of young males with lower qualifications (Rice, 1999; Meschi et al., 2011).

The results hint at the inter-generational dimension of low labour market attachment, as unsuccessful trajectories are found to be less common where the household head is in employment. However, 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 and Lewis, 1973). Where significant, our results are generally consistent with this prediction and other empirical evidence on the issue (Hanushek, 1992; Björklund et al., 2004). 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 et al., 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.

We also examine the relationship between self-confidence, motivation problems 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 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. As a robustness check, we also test an alternative specification which uses factor analysis to construct variables capturing the pattern of variation in responses to these questions. In both cases, results appear to confirm that self-confidence and motivation problems show lasting associations with future outcomes, increasing the probability of 'Accumulating human capital' or being in the 'Possible cause for concern' category but reducing the probability of being in the 'Express' category. Our model also considers 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 the effect is most likely not due to a direct impact of smoking per se, one might speculate that this signals traits such as rebelliousness or low self-confidence. The inclusion of these variables may also contribute to capturing the impact on labour market choices of personal traits which 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. The importance of the latter sheds light on the extent to which structural inequalities may 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 3. 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 three young males obtaining no GCSEs at 16 and living with unemployed parents holding low qualifications.

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 4. 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 3. This suggests that the observable characteristics of the subsample 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. Most of the additional characteristics do not appear to improve the predictive power of the model, although this could be due to the small sample size. 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 all of the five-year observation period in full-time education, albeit 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. There are, of course, concerns about high levels of graduate unemployment. but this is often relatively short-term in nature. For some years now, the rate of unemployment among 21-24 year olds with a degree has been somewhat lower than that amongst those without a degree.

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 et al. 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 and many in the

‘Planned interruption?’ group 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).

These relatively well-defined factors provide important clues for policymakers. Raising the participation age is a significant change, with potentially far-reaching consequences. The extent to which this helps young people will depend on whether it materially increases attainment (and skill) levels. Accompanying this, careful thought must be given to the structure of performance targets faced by schools. These can give rise to perverse incentives, whereby individuals unlikely to reach required attainment levels are left behind. Equally, our results have shown the significant role of truancy. Appropriate attendance targets may go some way towards alleviating this, though the fundamental causes of truancy will also need to be tackled.

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 et al., 2004). This possibly highlights how effective career advice and job-search assistance programmes might be used to facilitate successful employee-employer matches.

The fact that young mothers feature so prominently amongst those individuals who have not either entered work or remained in education suggests that there may also be a need for more general support with ‘life-planning’. While the choice to become a mother is a personal matter, the role of policy should be to help young women reach an informed view as to whether this is right for them.

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.

---

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

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

Tables

Table 1 : Share of individuals in each group exhibiting given characteristics

	Total	1	2	3	3a	3b	3c	3d	3e	3f
Ethnic minority	6%	4%	9%	4%	0%	4%	2%	0%	15%	0%
Female	48%	45%	47%	65%	58%	29%	64%	91%	81%	100%
Has children at 21	9%	7%	1%	45%	39%	4%	18%	83%	81%	100%
Health limits daily activities	5%	6%	2%	6%	0%	0%	12%	8%	12%	4%
5+ AC GCSE	50%	38%	81%	16%	20%	16%	10%	19%	29%	0%
1-4 AC GCSE	18%	26%	5%	16%	27%	27%	6%	16%	0%	29%
D-G GCSE	22%	26%	10%	37%	46%	34%	43%	22%	45%	34%
No qualifications	10%	10%	5%	31%	8%	24%	41%	43%	26%	38%
Receipt Educational Grant	5%	5%	4%	7%	4%	13%	12%	0%	3%	1%
Parental qualifications high	16%	10%	31%	3%	0%	10%	0%	0%	0%	0%
Parental qualifications medium	58%	61%	59%	38%	49%	55%	16%	33%	48%	27%
Parental qualifications low	25%	28%	10%	59%	51%	35%	84%	67%	52%	73%
Household head employed	79%	81%	84%	52%	68%	67%	41%	37%	51%	44%
Owned housing	78%	76%	92%	49%	91%	71%	19%	45%	57%	17%
Social rented	18%	20%	7%	48%	9%	25%	80%	55%	36%	75%
Private rented	3%	5%	1%	3%	0%	4%	1%	0%	7%	8%
No sibling	59%	57%	63%	57%	49%	46%	69%	52%	64%	61%
Employed sibling	22%	27%	13%	26%	43%	28%	17%	36%	12%	29%
NEET sibling	4%	4%	1%	13%	0%	26%	14%	7%	5%	10%
Sibling FT student	15%	12%	22%	4%	8%	1%	0%	5%	18%	0%
Observations	1297	743	419	135	15	28	30	23	21	18

(1) Express; (2) Accumulating human capital; (3) Possible cause for concern

(3a) Planned interruption?; (3b) Partial recovery; (3c) Long-term worklessness; (3d) NEET from 16; (3e) NEET from 18; (3f) Withdrawals from the labour market

Table 2: Trajectory groups and relative size in percentage of 16-year-old cohort

Description of trajectory	Express	Accumulating human capital	Possible cause for concern	Estimated number each year ('000s)
FTE throughout		24.4%		190
FTE with employment spell		7.8%		60
Express	56.4%			430
Planned interruption?			1.0%	10
Partial recovery			2.9%	20
Long-term worklessness			2.6%	20
NEET from 16			2.1%	20
NEET from 18			1.3%	10
Withdrawals from the labour market			1.3%	10
Total	**56.4%**	**32.3%**	**11.3%**	**760**

Table 3 : 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	
Sex (ref: males)						
Female	-0.038		-0.016		0.054	***
	[0.026]		[0.023]		[0.015]	
Ethnicity (ref: white)						
Non-white	-0.123	*	0.146	**	-0.022	
	[0.059]		[0.054]		[0.031]	
Parental qualifications (ref: Low)						
High (degree)	-0.154	**	0.229	***	-0.075	*
	[0.049]		[0.043]		[0.032]	
Medium (>GCSE A-C)	-0.062		0.114	***	-0.052	**
	[0.034]		[0.031]		[0.019]	
Housing tenure (ref: owned)						
Social rented	0.063		-0.095	**	0.031	
	[0.038]		[0.036]		[0.020]	
Private rented	0.234	***	-0.197	***	-0.037	
	[0.064]		[0.059]		[0.028]	
Household income	0		0		0	
	[0.000]		[0.000]		[0.000]	
Year of birth (time trend)	0		-0.003		0.003	
	[0.003]		[0.003]		[0.002]	
Month of birth (ref: May-Aug)						
Jan-Apr	0.067	*	-0.064	*	-0.003	
	[0.031]		[0.027]		[0.018]	
Sept-Dec	0.09	**	-0.063	*	-0.028	
	[0.031]		[0.028]		[0.018]	
Health (ref: no limitations)						
Health limits daily activities	0.173	**	-0.138	**	-0.035	
	[0.057]		[0.053]		[0.024]	
Smoker (ref: Non-smoker)						
Smoker	0.028		-0.145	***	0.117	***
	[0.036]		[0.032]		[0.023]	
School attainment (ref: 5+ GCSE A*-C)						
1-4 GCSE A-C	0.325	***	-0.353	***	0.029	
	[0.034]		[0.029]		[0.021]	
GCSE D-G	0.202	***	-0.257	***	0.054	**
	[0.037]		[0.034]		[0.020]	
No qualifications	0.142	**	-0.266	***	0.124	***
	[0.050]		[0.044]		[0.033]	
Educational grant (ref: none)						
In receipt	0.012		-0.025		0.014	
	[0.061]		[0.056]		[0.031]	
Local area claimant count rate dev (16-24)	-0.001		-0.006		0.007	
	[0.007]		[0.007]		[0.004]	

Table 3 : 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	
Employment of household head (ref: not employed)						
In employment	0.079	*	-0.024		-0.055	**
	[0.035]		[0.032]		[0.020]	
Sibling labour force status (ref: no siblings)						
Employed	0.067	*	-0.08	**	0.013	
	[0.033]		[0.030]		[0.019]	
NEET	0.057		-0.114		0.057	
	[0.073]		[0.068]		[0.037]	
In FTE	0.01		0.023		-0.032	
	[0.038]		[0.033]		[0.023]	
Count of 'negative' GHQ responses	-0.02	***	0.015	**	0.005	*
	[0.005]		[0.005]		[0.003]	
Number of observations			1282			

*Standard errors in brackets. Significant at the: * 5%; ** 1%;*** 0.1% level*

Table 4 : 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)	
Sex (ref: males)												
Female	0.022		0.011		-0.071		-0.07		0.049		0.059	*
	[0.041]		[0.044]		[0.036]		[0.039]		[0.025]		[0.027]	
Ethnicity (ref: white)												
Non-white	-0.24	**	-0.212	*	0.214	**	0.207	*	0.026		0.005	
	[0.082]		[0.083]		[0.081]		[0.080]		[0.056]		[0.053]	
Parental qualifications (ref: Low)												
High (degree)	-0.096		-0.096		0.145	*	0.137	*	-0.049		-0.041	
	[0.074]		[0.076]		[0.066]		[0.067]		[0.046]		[0.050]	
Medium (>GCSE A-C)	-0.091		-0.079		0.098		0.086		-0.006		-0.007	
	[0.057]		[0.058]		[0.052]		[0.054]		[0.030]		[0.031]	
Housing tenure (ref: owned)												
Social rented	0.073		0.059		-0.098		-0.095		0.025		0.036	
	[0.062]		[0.062]		[0.058]		[0.058]		[0.032]		[0.034]	
Private rented	0.319	**	0.284	*	-0.29	***	-0.279	**	-0.03		-0.005	
	[0.103]		[0.120]		[0.086]		[0.100]		[0.055]		[0.071]	
Household income	0		0		0		0		0		0	
	[0.000]		[0.000]		[0.000]		[0.000]		[0.000]		[0.000]	
Year of birth (time trend)	-0.003		-0.001		-0.001		-0.003		0.004		0.004	
	[0.013]		[0.013]		[0.012]		[0.011]		[0.008]		[0.008]	
Month of birth (ref: May-Aug)												
Jan-Apr	0.084		0.084		-0.081		-0.076		-0.003		-0.007	
	[0.048]		[0.048]		[0.043]		[0.043]		[0.029]		[0.030]	
Sept-Dec	0.144	**	0.155	**	-0.098	*	-0.093	*	-0.046		-0.062	*
	[0.052]		[0.051]		[0.047]		[0.047]		[0.029]		[0.029]	
Health (ref: no limitations)												
Health limits daily activities	0.159		0.16		-0.149	*	-0.134		-0.009		-0.026	
	[0.088]		[0.088]		[0.076]		[0.079]		[0.051]		[0.048]	
Smoker (ref: Non-smoker)												
Smoker	-0.108		-0.137	*	-0.087		-0.058		0.195	***	0.195	***
	[0.062]		[0.061]		[0.057]		[0.059]		[0.048]		[0.047]	
School attainment (ref: 5+ GCSE A*-C)												
1-4 GCSE A-C	0.343	***	0.337	***	-0.356	***	-0.34	***	0.013		0.003	
	[0.051]		[0.051]		[0.044]		[0.045]		[0.032]		[0.032]	
GCSE D-G	0.23	***	0.207	**	-0.287	***	-0.262	***	0.057		0.055	
	[0.061]		[0.063]		[0.056]		[0.059]		[0.036]		[0.037]	
No qualifications	0.135		0.154	*	-0.24	***	-0.237	***	0.105		0.083	
	[0.075]		[0.074]		[0.066]		[0.066]		[0.054]		[0.054]	
Educational grant (ref: none)												
In receipt	0.009		0.005		-0.042		-0.022		0.033		0.017	
	[0.097]		[0.097]		[0.089]		[0.091]		[0.053]		[0.050]	

Table 4 : 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)
Local area claimant count rate dev (16-24)	0.025	0.025	-0.025	-0.027	0	0.002
	[0.016]	[0.016]	[0.014]	[0.014]	[0.009]	[0.009]
Employment of household head (ref: not employed)						
In employment	0.062	0.062	-0.007	-0.008	-0.055	-0.053
	[0.056]	[0.056]	[0.053]	[0.053]	[0.031]	[0.031]
Sibling labour force status (ref: no siblings)						
Employed	0.08	0.067	-0.097 *	-0.095 *	0.016	0.028
	[0.053]	[0.053]	[0.047]	[0.047]	[0.031]	[0.032]
NEET	-0.035	-0.038	-0.065	-0.065	0.101	0.103
	[0.113]	[0.109]	[0.110]	[0.110]	[0.065]	[0.063]
In FTE	0.038	0.035	-0.013	-0.017	-0.025	-0.018
	[0.060]	[0.059]	[0.052]	[0.052]	[0.037]	[0.039]
Count of 'negative' GHQ responses	-0.014	-0.011	0.015	0.013	-0.001	-0.003
	[0.009]	[0.009]	[0.008]	[0.009]	[0.005]	[0.005]
Truancy (ref: none)						
Plays truant		0.169 ***		-0.15 **		-0.018
		[0.051]		[0.047]		[0.027]
Discipline (ref: none)						
Expelled/suspended or has vandalised property		0.075		-0.011		-0.064
		[0.051]		[0.045]		[0.037]
Index of agreement with traditional gender roles		0.001		0.009		-0.011
		[0.025]		[0.023]		[0.014]
Means a lot to do well at school (ref: agrees)						
Does not agree		0.019		-0.004		-0.015
		[0.107]		[0.106]		[0.040]
Bullying (ref: does not worry about it)						
Is worried about it		-0.104 *		0.032		0.072 *
		[0.047]		[0.042]		[0.030]
Number of observations			530			

*Standard errors in brackets. Significant at the: * 5%; ** 1%;*** 0.1% level*
Columns labelled 'subsample' re-estimate the model in Table 3 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'

Figures

Figure 1

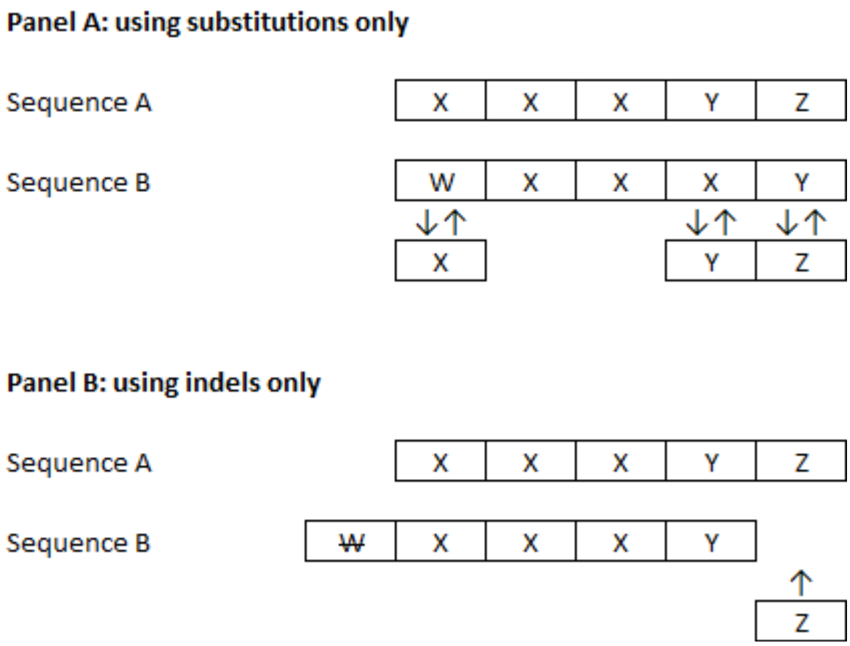


Figure 2

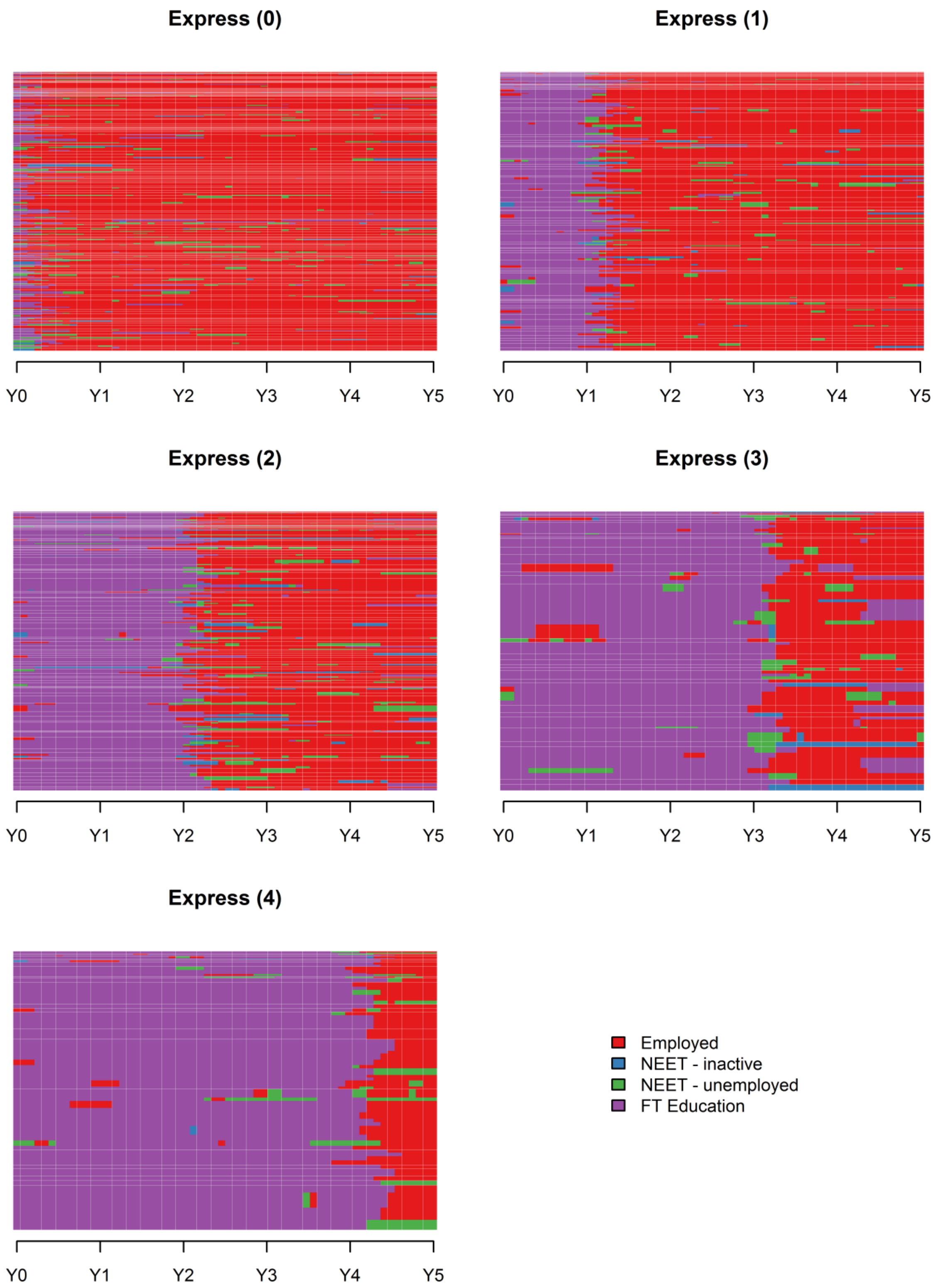

Express (0)
Express (1)
Express (2)
Express (3)
Express (4)
Y0
Y1
Y2
Y3
Y4
Y5
Employed
NEET - inactive
NEET - unemployed
FT Education

Figure 3

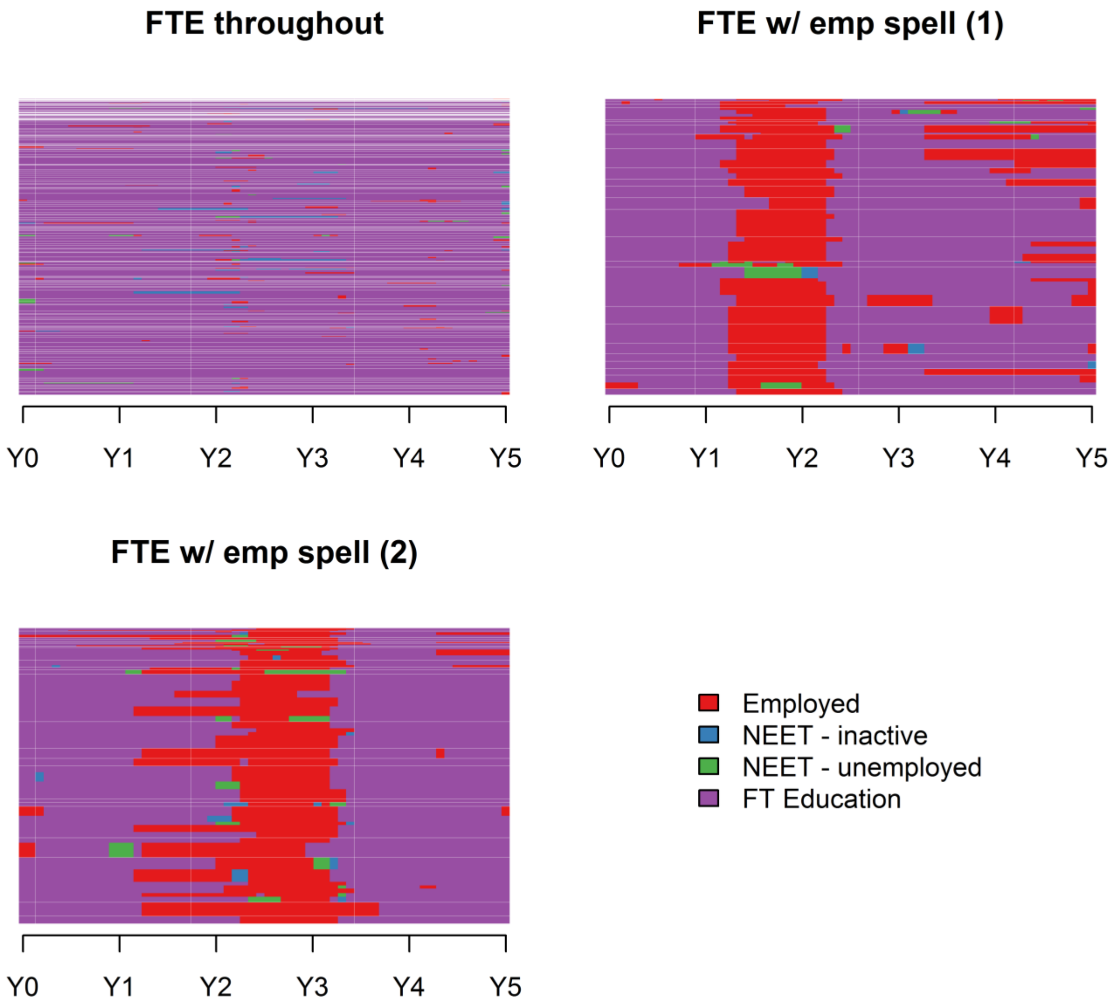
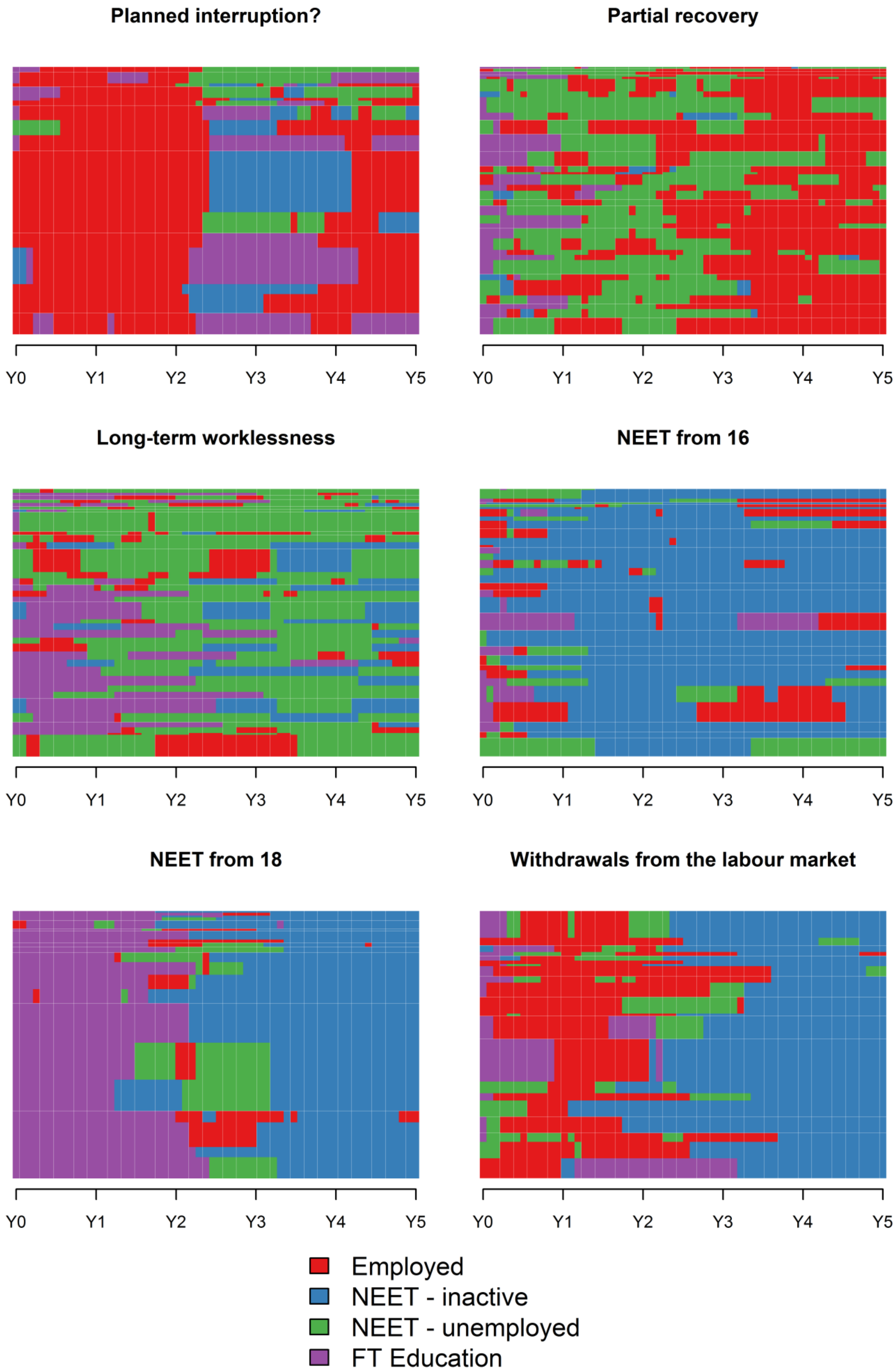

FTE throughout
FTE w/ emp spell (1)
FTE w/ emp spell (2)
Y0
Y1
Y2
Y3
Y4
Y5
Employed
NEET - inactive
NEET - unemployed
FT Education

Figure 4



Planned interruption?
Partial recovery
Long-term worklessness
NEET from 16
NEET from 18
Withdrawals from the labour market
Y0
Y1
Y2
Y3
Y4
Y5
Employed
NEET - inactive
NEET - unemployed
FT Education

References

Abbot, A. and Forrest, J., "Optimal Matching Methods for Historical Sequences", Journal of Interdisciplinary History 16, 3 (1986), pp. 471–494.

Abbott, Andrew, "Reply to Levine and Wu", Sociological Methods & Research 29, 1 (2000), pp. 65--76.

Andrews, Martyn and Bradley, Steve, "Modelling the Transition from School and the Demand for Training in the United Kingdom", Economica 64, 255 (1997), pp. 387--413.

Angrist, Joshua and Lavy, Victor and Schlosser, Analia, "Multiple Experiments for the Causal Link between the Quantity and Quality of Children", Journal of Labor Economics 28, 4 (2010), pp. 773--824.

Anyadike-Danes, Michael and McVicar, Duncan, "You'll never walk alone: Childhood influences and male career path clusters", Labour Economics 12, 4 (2005), pp. 511--530.

Anyadike-Danes, Michael and McVicar, Duncan, "My Brilliant Career: Characterizing the Early Labor Market Trajectories of British Women From Generation X", Sociological Methods & Research 38, 3 (2010), pp. 482--512.

Barnes, M., Green, R. and Ross, A.,. "Understanding Vulnerable Young People: Analysis from the Longitudinal Study of Young People in England". 118. Research Report. Department for Education. (2011)

Becker, Gary S and Lewis, H Gregg, "On the Interaction between the Quantity and Quality of Children", Journal of Political Economy 81, 2 (1973), pp. S279--88.

Björklund, A. and Eriksson, T. and Jäntti, M. and Oddbjørn, R. and Österbacka, E., "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 Corak, M., ed., Generational Income Mobility in North America and Europe (Cambridge University Press, 2004).

Brzinsky-Fay, Christian, "Lost in Transition? Labour Market Entry Sequences of School Leavers in Europe", European Sociological Review 23, 4 (2007), pp. 409--422.

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

Calinski, T., and Harabasz. J., "A Dendrite Method for Cluster Analysis." *Communications in Statistics - Theory and Methods* 3, 1 (1974), pp.1–27.

Coles, B., Youth And Social Policy: Youth Citizenship And Young Careers (UCL Press, 1995).

Crawford, Claire and Dearden, Lorraine and Meghir, Costas, "When you are born matters: the impact of date of birth on educational outcomes in England", Institute for Fiscal Studies (2010).

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

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

Elzinga, Cees H., "Sequence Similarity", Sociological Methods & Research 32, 1 (2003), pp. 3--29.

Everitt, B.S. and Landau, S. and Leese, M. and Stahl, D., Cluster Analysis (Wiley, 2011).

Fergusson, Ross and Pye, David and Esland, Geoff and McLaughlin, Eugene and Muncie, John, "Normalized dislocation and new subjectivities in post-16 markets for education and work", Critical Social Policy 20, 3 (2000), pp. 283--305.

Furlong, A. and Cartmel, F. and Foundation, Joseph Rowntree, Vulnerable Young Men in Fragile Labour Markets: Employment, Unemployment and the Search for Long-term Security (Joseph Rowntree Foundation, 2004).

Goujard, Antoine and Petrongolo, Barbara and Van Reenen, John, "The Labour Market for Young People", in Gregg, Paul and Wadsworth, Jonathan, ed., The Labour Market in Winter: The State of Working Britain (Oxford University Press, 2011).

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

Hanushek, Eric A., "The Trade-off between Child Quantity and Quality", Journal of Political Economy 100, 1 (1992), pp. pp. 84--117.

Lesnard, L., Optimal matching and social sciences ((Paris, France): INSEE, 2006).

Lesnard, Laurent, "Cost Setting in Optimal Matching to Uncover Contemporaneous Socio-Temporal Patterns", SciencesPo. Observatoire sociologique du changement (2009).

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

Martin, P. and Wiggins, R.D., Optimal Matching Analysis (Sage Publications Ltd, 2011).

Martin, Peter and Schoon, Ingrid and Ross, Andy, "Beyond Transitions: Applying Optimal Matching Analysis to Life Course Research", International Journal of Social Research Methodology 11, 3 (2008), pp. 179--199.

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

McVicar, Duncan and Rice, Patricia, "Participation in further education in England and Wales: an analysis of post-war trends", Economics Division, School of Social Sciences, University of Southampton (2000).

Meschi, Elena and Swaffield, Joanna K. and Vignoles, Anna, "The Relative Importance of Local Labour Market Conditions and Pupil Attainment on Post-Compulsory Schooling Decisions", Institute for the Study of Labor (IZA) (2011).

Micklewright, John, "Choice at Sixteen", Economica 56, 221 (1989), pp. 25--39.

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

Paull, Gillian, "Biases in the reporting of labour market dynamics" (2002).

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

Quintini, Glenda and Manfredi, Thomas, "Going Separate Ways? School to work Transitions in the United States and Europe", OECD Publishing (2009).

Rice, Patricia, "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 (1999), pp. 287--312.

Rousseeuw, J. P.,. "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis." Journal of Computational and Applied Mathematics 20, 0: (1987), pp.53 – 65.

Ryan, Paul, "The School to work Transition: A Cross-National Perspective", Journal of Economic Literature 39, 1 (2001), pp. pp. 34--92.

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

Scherer, Stefani, "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 (2005), pp. 427--440.

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

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

Waddell, Glen R., "Labor-Market Consequences Of Poor Attitude And Low Self-Esteem In Youth", Economic Inquiry 44, 1 (2006), pp. 69–97.

Wu, L.L., "Some Comments on" Sequence Analysis and Optimal Matching Methods in Sociology: Review and Prospect"", Sociological methods and research 29, 1 (2000), pp. 41–64.