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

Richard Dorsett† and Paolo Lucchino‡§

November 7, 2012

Abstract

This paper explores the school-to-work transition in the UK with the aim of achieving a richer understanding of individuals’ choices and activities in the five years after reaching school-leaving age. Through the technique of ‘optimal matching’, we assess the degree of similarity between individuals’ post-16 experiences in a way that captures the full detail of their five-year histories. We consider individuals reaching school-leaving age between 1991 and 2003 and, on the basis of the measures of similarity, 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 results confirm the predictive power of school attainment (grades), family background (parental qualifications, parental and sibling labour market status) and gender. 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.

†National Institute of Economic and Social Research
‡Corresponding author: National Institute of Economic and Social Research, 2 Dean Trench Street, Smith Square, London SW1P 3HE. Email: p.lucchino@niesr.ac.uk; Tel: +44 (0)20 7654 1939
§Acknowledgements: The authors would like to thank the Nuffield Foundation of its support through grant EDU/39082. The authors are grateful for comments received at seminars held at the ACEVO Commission on Youth Unemployment, the UK Social Policy Conference and at the National Institute of Economic and Social Research. Any errors are responsibility of the authors.
1 Introduction

While youth unemployment hit a record high in the wake of the recent recession, the UK youth labour market had in fact started to deteriorate as early as 2004 for reasons that are not well-understood (Goujard et al. 2011). Underlying the cyclical fluctuations, therefore, appears to be a deeper structural problem in the transition from school to work. Shifting social and economic conditions over the last three decades in Britain, and indeed globally (Bynner 2001, Lawy 2002), have diminished the centrality of the traditional route of early school-leaving and rapid entry into employment (Pollock 2007). Trajectories have become more individualised, with educational attainment gaining an increasing importance in shaping young people’s life chances and exposing the poorest and lowest-achieving young people to greater vulnerability (Pollock 2007, Bynner 2001). Indeed, a relatively large body of literature documents the social polarisation in the transition from school to work. (Mickiewicz 1989, Dickerson & Jones 2004, Rice 1999, Spielhofer 2009, Gregg et al. 1998) While the effects of disadvantage are similar across countries, these are found to be particularly marked in the UK (Ryan 2001). There is, furthermore, evidence of the existence of a hardcore group of youth who fall between the cracks in school-to-work transition institutions and spend a substantial amount of time NEET after the end of compulsory schooling. Research evidence points to specific personal circumstances and characteristics that go beyond mere family and social background, such as truancy, bullying, health, disability, pregnancy and caring responsibilities as well as attitudinal or behavioural factors such as low confidence or self esteem (Stone et al. 2000, Rennison et al. 2000, Sachdev et al. 2006, Gregg et al. 1998). Increasingly therefore, as confirmed in research by Fergusson et al. (2000), the experiences of many young people following compulsory education do not follow stable, linear and ‘traditional’ transition trajectoriest, but complex and often circular movements through multiple destinations.

This paper makes studies a sample of youth reaching the end of compulsory schooling between 1991 and 2003 in the UK, and traces their pathways over the following five years. It makes two contributions. Firstly, it uses an innovative statistical technique – optimal matching – to compare individuals’ experiences and uncover patterns shared among them. The appeal of this approach is that it captures the full richness of individuals’ experiences post-school leaving age. In doing so, it overcomes the limitations of commonly used statistics which generally 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) and, as a consequence, discard important information on labour market dynamics, such as the order in which events occur. Such statistics are likely to be less appropriate for the youth labour market, where the specificities of employment and joblessness problems (Rees 1986) and the importance of distinguishing between school-to-work transitions characterised by short or long-term dif-
ficulties (Ryan 2001) have been widely recognised. Optimal matching situates individuals’ experiences at a given point within their wider labour market histories and can distinguish, for example, between transitory ‘gap years’ and deep disconnect from the labour market. It shifts the emphasis from the analysis of a specific step to considering the general features of the transition process, including initial circumstances, the routes followed by individuals, and, importantly, final outcomes.

Drawing on this technique, we are able take a broad look at school-to-work transitions in the UK and identify a small number of distinct transition patterns. This builds on earlier research that used optimal matching to study youth employment histories and the school to work transition in the UK (Halpin & Chan 1998, Schoon et al. 2001, Anyadike-Danes & McVicar 2005, Martin et al. 2008) and in a comparative perspective (Scherer 2005, Brzinsky-Fay 2007, Quintini & 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 & Manfredi (2009) are exceptions to this, but use data on youth histories up to only 2000 and 2001 respectively. We therefore add to the existing literature by considering detailed monthly histories extending to 2008 and constructed from annually repeated 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 such technique is suitably justified by theory, as detailed in the relevant section below.

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 in this way provides important clues as to the type of policy that might be effective and who it should target. However, much of the relevant literature has, again, focused on the influence of characteristics on future outcomes defined at a specific point in time. As above, the prevailing approach does not distinguish by the transience of such outcomes. Building on Anyadike-Danes & McVicar (2005, 2010), we therefore use the groupings identified in the first part of the analysis to focus on what characteristics are associated with successful or unsuccessful outcomes as defined by the nature of one’s overall trajectory 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 or a formal placement. On the other hand, the remaining individuals exhibit a variety of histories that might warrant policy attention. We identify six key at risk groups: individuals who experience a (possibly planned) break from employment but then appear to struggle to return to work; individuals experiencing some employment but developing only limited labour market attachment; individuals exhibiting patterns of long-term worklessness straddling unemployment and inactivity; those in long-term inactivity from the age of 16;
those in long-term inactivity from the age of 18; and individuals who appear to withdraw from
the labour market following an apparently successful entry into employment. The subsequent
analysis confirms the importance of school attainment (grades), family background (parental
qualifications, parental and sibling labour market status), and gender as strong predictors
of future labour market trajectories, which, being highly correlated across individuals, gives
rise to a significant level of socio-economic polarisation.

The paper is structured as follows. Section 2 introduces the technique of optimal matching
and the methodological approach taken. A description of the data used is presented in Section
3. The results from optimal matching are presented in Section 4 and are followed by results
from the analysis of individual predictors of future labour market trajectories in Section 5.
Section 6 concludes.

2 Creating a typology of school-to-work transitions

2.1 Approach

We explore the ways the school-to-work transition unfolds in the UK 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 & Kruskal 1983, Abbot & Forrest 1986). Secondly, we apply
cluster analysis techniques to the derived measures of dissimilarity to group similar sequences
together.\(^1\)

Optimal matching is a relatively novel technique in the social sciences and deserves a more
thorough presentation. The optimal matching algorithm derives a measure of dissimilarity
between two sequences as a function of the number and type of operations on the elements of
these sequences that are necessary to transform one sequence into the other. Such operations
can be combinations of insertions/deletions (\textit{indels}) and substitutions of elements. 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 \textit{Hamming distance}, retains the timing of events
and measures dissimilarity as the number of elements that need to be altered. On the other
hand, 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 sub-sequences. The resulting measure of
dissimilarity will therefore be lower the more the sequences share common subsections. As
such, this is known as the \textit{longest common subsequence distance}. This measure emphasises

\(^1\)We do this using Ward’s method, which groups sequences in such a way as to minimise the variance
within each cluster.
the presence of common subsections and the ordering of elements. However, as elements are deleted or inserted, any relationship of contemporaneity across sequences may be broken. The temporal dimension within a sequence may also be altered as, when elements are deleted (inserted), neighbouring elements become temporally closer (more distant). This causes a warping or slowing down of time, which may not be suitable in certain contexts (more on this below). Falling between these two extremes, alternative measures of dissimilarity can be constructed using a combination of both types of operations by assigning each operation a specified ‘cost’. This cost will represent the amount such operation will add to the measure of dissimilarity. For example, an arguably ‘default’ option is to set the cost of substitution be equal to the cost of a deletion followed by an insertion, as these would yield the same change in the sequence.

### 2.2 Cost setting and measures of distance

The costs of each operation determine how dissimilarity is defined in the context under study and how sequences are matched. Specifying a menu of costs is therefore an important step in this process as it may influence the results that emerge. The literature does not set rigid rules on this. However, there is scope to parameterise the cost matrix to make it more consistent with theoretically-informed definitions of what constitutes similarity in the context under study. A few considerations may be 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 indel operations. As mentioned above, similarities within 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 sentence structure. However, as indels cause a time-warp effect and break the contemporaneity between different sequences, they should be limited where timing is important. One example of such context is the study of working patterns over the 24 hours of the day (Lesnard 2009). More generally, timing will be relevant when sequences are defined over a socio-economic ‘calendar’, which could be a very fixed temporal cycle (such as the working week), but also a somewhat looser institutional system (such as the higher education system) or even a natural calendar (such as the stages in early childhood development).

Secondly, it is possible to define the socioeconomic proximity of different states through the costs assigned to different pairs of substitutions. For example, in some contexts, self-employment can be considered to be closer to employment than to inactivity (Anyadike-Danes & McVicar 2005) and a substitution into the former might therefore be given a lower substitution cost. 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 the inverse of the unconditional probability of a transition occurring at a given point in time as calculated from the data. This approach is used in Lesnard (2009) on working time and implies, for example, considering the substitution ‘not working’ for ‘working’ at the 9am element of the sequence to be substantially cheaper than the same substitution at 1am, as many more individuals exhibit such a transition at 9am than at 1am.

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 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 (e.g. the A-level structure implies there are likely to be transition points at the end of the first two years, while less so at other points in the academic calendar). To address this we use time-varying substitution costs defined as the inverse of the unconditional transition probability at the specific point in the sequence,

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

This states that the substitution cost of substituting a for b or vice versa will be a declining function of the frequency of such transition at the given point in time, as measured in the data by the conditional probabilities of an a to b or b to a transition between the current and adjacent periods.
as described above. The distance measure obtained when imposing these two conditions is called the *dynamic hamming distance* (Lesnard 2006).

### 2.3 Advantages and limitations

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. For this reason, it overcomes limitations of other commonly used statistics, which, as already mentioned, generally summarise outcomes at a point in time or over a specified period and therefore discard 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 then 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 found in the study of DNA sequences, it has increasingly been used in the social sciences. For example, it has been used to compare status biographies (such as employment careers, partnership histories, mental health ‘careers’ of service use etc), the content of college textbooks, English folk dances, birdsong patterns, local dialects, lynching patterns and more (see Martin & 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 equate actual chemical processes in a DNA strand, their meaning in a socio-economic context is less clear, and that the same therefore would hold of any resulting measure of distance. Furthermore, the lack of formal rules to the define the cost matrix has also 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 & 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 respond by stressing that optimal matching, and the operations on the sequence, are not intended to be a model of reality or an exercise in social engineering, but rather a way of constructing a synthetic measure of difference from sequences containing very complex information (Abbott 2000, Lesnard 2006). Simulation exercises aimed at understanding the sensitivity of resulting distance measures to cost-setting assumptions have only recently emerged (Halpin 2010) and this relationship is not yet fully...
understood (Martin & Wiggins 2011). Some have commented that optimal matching results are often robust to variation in costs (Abbott & Tsay 2000), which can be viewed as evidence in favour of the existence of objective patterns in the data. On the other hand, this would also imply that the different parameterisations are therefore not falsifiable (Levine 2000).

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 as we have motivated above. The intention here is not to mine the data for patterns, but to define how the algorithm should conceive of similarity. Nevertheless, while a given parameterisation of optimal matching followed by cluster analysis will yield a typology that satisfies 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 relatively 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

The analysis used 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 full (adult) interviews once they reach age 16, and are interviewed annually thereafter. Our attention is focused on such children.

We constructed a month-by-month history for each young person, building on the careful studies of other researchers (Paull 2002, Maré 2006) into how the BHPS can best be used to generate consistent series. 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 to 18 months. Depending on the month of interview, the recall period will overlap with information given at the previous interview wave. Constructing consistent work-life histories requires reconciling any inconsistencies present in this overlap. We do so following the reconciliation techniques provided in Maré

---

3 Booster samples were added in 1999 and 2001.

4 Wave X interviews occur between September X and April X+1. Individuals are asked to recall their activity back to September X-1, thereby covering a period of 12 to 18 months, depending on the month of
An advantage of focusing on those observed to turn 16 is that we observe their full labour market histories without having to rely on long-term respondent recall. Since interviews take place annually, respondents are not required to recollect experiences from more than (roughly) one year before.\(^5\) 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 the greatest difficulty accurately recalling their prior experiences. She also finds that young people tend to overestimate the amount of time spent in employment as the recall period lengthens. The granularity of information within a labour market trajectory is lost when using long-term retrospective data, so relying on recall periods of about one year keeps this potential bias to an absolute minimum. Secondly, as we are interested in status biographies covering periods of employment as well as non-employment, we follow the ‘main activity’ definition of status as in Paull (2002). This state is defined according to the individuals’ own identification of their main activity from a list of 10 available choices.\(^6\) 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 to defining an individual’s labour market status has 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 allow for the possibility that individuals may be engaged in more than one activity at the same time, 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.\(^7\) For these reasons, our definitions will not be consistent with official labour market measures such as the ILO unemployment rate. Indeed, the histories tend to overestimate educational participation, underestimate official youth employment rates but closely track Department for Education NEET rates. Finally, we do not have information on part-time education. Overall, however, the data provide a rich description of the history of youth in the sample and can therefore provide important insights into their labour market experience.

We restrict our attention to the (roughly) 1,400 individuals observed for five consecutive years starting from the month they could legally leave school. This yields a 60-element sequence for each of the individuals in our sample, where each element can take only one

\(^5\)97% of the months in the reconstructed life-work histories rely on recall of 14 months or less.

\(^6\)There were: self-employed; employed; unemployed; retired; maternity leave; family care; full-time student; long-term sick/disabled; Government training scheme; and other.

\(^7\)Concurrent employment and main activities other than employment can be identified at the time of interview but not in the months between interview waves where only one’s main activity is recorded. For this reason, only individuals’ main activities could be derived consistently across the observation period.
of the four above-mentioned labour market states. As discussed in the previous section, 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 use this to adjust each young person’s BHPS cross-sectional weight taken at the point where they can legally leave school.

4 Trajectory types

The trajectories typifying each of the groups identified through the two-step procedure described above can be visualised through graphs showing the full histories of each group member. Recalling how 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 by a horizontal series of colour-coded dots. Stacking such plots for all individuals in a given group gives an immediate picture of the general labour market dynamics characterising that group.

Overall, we identified 14 groups. These can themselves be grouped into three high-level categories. Figure 2 plots the sequences from school leaving age (Y0) to five years later (Y5) for individuals falling within one of the five groups experiencing smooth transitions from education to work, and only differing in the number of additional years of education before the transitions occurred. Note that the size of each plot is not an indication of the group’s size (which is discussed at the end of this section). Following Brzinsky-Fay (2007) and Quintini & Manfredi (2009), we called these ‘Express’ education to work transitions.

A further four groups describing predominantly educational trajectories are shown in Figure 3. The first group describes individuals who stay in education throughout, while individuals in the remaining two groups also spend a substantial time in education interrupted by one (or in fewer cases two) academic year(s) in employment. We called these ‘Full-time education’ and ‘Full-time education with an employment spell’ respectively. The latter title is likely to be inadequate for a minority of individuals in these groups who experienced a longer interruption from education and might be best termed ‘returners to education’. However, given their relatively small number we decided against splitting this group further.

The remaining individuals exhibit a variety of histories that might warrant policy attention. Their trajectories are depicted in Figure 4 and include individuals who experience a (possibly planned) break from employment but may struggle to return to work; individuals experiencing some employment but developing only limited labour market attachment; patterns of long-term worklessness straddling unemployment and inactivity; long-term inactivity from the age of 16 or from age 18; and individuals who appear to withdraw from the
Figure 2: Express transitions from school-to-work
Figure 3: Accumulating human capital
labour market following an apparently successful entry into employment.

Table 1 presents the final typology we identified and an estimate of the size of each type. The results suggest that 9 out of 10 young people experience generally successful labour market trajectories, 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.

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

Table 1: Trajectory groups and relative size

The types of trajectories identified above show a broad agreement with previous research using similar approaches. Indeed, both Brzinsky-Fay (2007) and Quintini & Manfredi (2009) identify typologies that closely resemble some of the histories grouped above. Both find a large ‘Express’ group; a group returning to education after a period in work or inactivity (their ‘Link’, ‘Return’, ‘Break’ and ‘Gap year’ groups); a group finding work after a relatively long period of unemployment (‘Detour’); groups experiencing long periods out
Figure 4: Possible cause for concern
of work ('Failure' and 'Discouraged'); and groups in sustained inactivity ('Disconnect' and 'Dropout'). 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' pathways is by far the most common. While they 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 as an integral part of the unravelling of one’s transition from the world of education to that of work.\textsuperscript{8} It is possible, therefore, that individuals who we find as exhibiting a stable trajectory within the defining structure of the educational system over the five years we consider may later move onto less fortunate trajectories once they leave the education system.

5 Predicting future labour market outcomes

Understandably, individuals in the six 'possible cause for concern' groups will be of greatest interest to policy-makers. In particular, the ability to identify in advance who is at risk of an unsuccessful transition into the labour market is clearly important to inform the type of policy that might be effective and who it should target.

Table 2 shows the distribution of selected characteristics among individuals in each group.\textsuperscript{9} Note, however, that these should be considered indicative due to the small sample size of each group. Relative to those who remain in education or make a successful transition from education to work, those in the 'possible cause for concern' groups generally exhibit lower educational attainment at age 16; are more likely to live in social rented accommodation; and have parents with lower educational qualifications. Importantly, there is also a high degree of gender variation among these groups. In particular, while unsuccessful trajectories predominantly consisting of time in 'NEET - unemployed' (groups 4 and 5) show a balanced gender profile overall, the overwhelming majority of those entering predominantly 'NEET - inactive' trajectories (groups 6-8) are female and in almost all 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, the groups ‘NEET from 16’, ‘NEET from 18’ and ‘Withdrawals from the labour market’ only emerge when considering females only. This point reinforces the importance of qualifying the description of these groups as giving rise to a possible cause for concern.

\textsuperscript{8}As mentioned in Section 3, this also has the advantage of placing sequences within a common socio-economic calendar.

\textsuperscript{9}The groups ‘full-time education throughout’ and ‘full-time education with an employment spell’ have been combined under the broader heading ‘accumulating human capital’.
Table 2: Share of individuals in each group exhibiting given characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethnic minority</td>
<td>9%</td>
<td>4%</td>
<td>0%</td>
<td>4%</td>
<td>2%</td>
<td>0%</td>
<td>15%</td>
<td>0%</td>
<td>6%</td>
</tr>
<tr>
<td>Female</td>
<td>47%</td>
<td>45%</td>
<td>58%</td>
<td>30%</td>
<td>64%</td>
<td>91%</td>
<td>81%</td>
<td>100%</td>
<td>48%</td>
</tr>
<tr>
<td>Has children at 21</td>
<td>1%</td>
<td>7%</td>
<td>39%</td>
<td>4%</td>
<td>18%</td>
<td>83%</td>
<td>82%</td>
<td>100%</td>
<td>9%</td>
</tr>
<tr>
<td>Health limits daily activities</td>
<td>2%</td>
<td>6%</td>
<td>0%</td>
<td>0%</td>
<td>12%</td>
<td>8%</td>
<td>12%</td>
<td>4%</td>
<td>5%</td>
</tr>
<tr>
<td>GCSE A-C</td>
<td>92%</td>
<td>73%</td>
<td>55%</td>
<td>51%</td>
<td>19%</td>
<td>43%</td>
<td>37%</td>
<td>37%</td>
<td>76%</td>
</tr>
<tr>
<td>GCSE D-G</td>
<td>3%</td>
<td>17%</td>
<td>37%</td>
<td>25%</td>
<td>40%</td>
<td>14%</td>
<td>37%</td>
<td>25%</td>
<td>14%</td>
</tr>
<tr>
<td>No qualifications</td>
<td>5%</td>
<td>10%</td>
<td>8%</td>
<td>24%</td>
<td>41%</td>
<td>43%</td>
<td>25%</td>
<td>38%</td>
<td>10%</td>
</tr>
<tr>
<td>Receipt Educational Grant</td>
<td>3%</td>
<td>5%</td>
<td>4%</td>
<td>13%</td>
<td>12%</td>
<td>0%</td>
<td>3%</td>
<td>1%</td>
<td>5%</td>
</tr>
<tr>
<td>Parental qualifications high</td>
<td>31%</td>
<td>10%</td>
<td>0%</td>
<td>10%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>16%</td>
</tr>
<tr>
<td>Parental qualifications medium</td>
<td>59%</td>
<td>61%</td>
<td>49%</td>
<td>55%</td>
<td>16%</td>
<td>33%</td>
<td>48%</td>
<td>27%</td>
<td>58%</td>
</tr>
<tr>
<td>Parental qualifications low</td>
<td>10%</td>
<td>28%</td>
<td>51%</td>
<td>35%</td>
<td>84%</td>
<td>67%</td>
<td>52%</td>
<td>73%</td>
<td>25%</td>
</tr>
<tr>
<td>Owned housing</td>
<td>92%</td>
<td>76%</td>
<td>91%</td>
<td>71%</td>
<td>44%</td>
<td>57%</td>
<td>17%</td>
<td>78%</td>
<td></td>
</tr>
<tr>
<td>Social rented</td>
<td>7%</td>
<td>20%</td>
<td>9%</td>
<td>25%</td>
<td>80%</td>
<td>56%</td>
<td>36%</td>
<td>75%</td>
<td>18%</td>
</tr>
<tr>
<td>Private rented</td>
<td>1%</td>
<td>5%</td>
<td>0%</td>
<td>4%</td>
<td>1%</td>
<td>0%</td>
<td>7%</td>
<td>9%</td>
<td>3%</td>
</tr>
<tr>
<td>No sibling</td>
<td>63%</td>
<td>57%</td>
<td>49%</td>
<td>45%</td>
<td>69%</td>
<td>52%</td>
<td>65%</td>
<td>61%</td>
<td>59%</td>
</tr>
<tr>
<td>Employed sibling</td>
<td>12%</td>
<td>26%</td>
<td>42%</td>
<td>28%</td>
<td>14%</td>
<td>27%</td>
<td>12%</td>
<td>29%</td>
<td>21%</td>
</tr>
<tr>
<td>NEET sibling</td>
<td>1%</td>
<td>4%</td>
<td>0%</td>
<td>27%</td>
<td>17%</td>
<td>7%</td>
<td>5%</td>
<td>10%</td>
<td>4%</td>
</tr>
<tr>
<td>Sibling FT student</td>
<td>24%</td>
<td>13%</td>
<td>8%</td>
<td>1%</td>
<td>0%</td>
<td>14%</td>
<td>18%</td>
<td>0%</td>
<td>16%</td>
</tr>
</tbody>
</table>

(1) Accumulating human capital; (2) Express; (3) Planned interruption?; (4) Partial recovery; (5) Long-term worklessness; (6) NEET from 16; (7) NEET from 18; (8) Withdrawals from the labour market.

In many cases, those ‘NEET - inactive’ trajectories may be so through voluntary choice, despite their detrimental effects on labour market progression. However, to the extent any such choice is constrained, there may still be a legitimate role for policy.

We use statistical techniques to explore whether there are any distinctive characteristics at age 16 which could help predict an individual’s future labour market trajectory type. A rich body of literature has explored 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 explore statistical associations between characteristics at age 16 with outcomes associated with
one’s *overall trajectory* over the five years after compulsory schooling. We run multinomial logit estimations to predict whether an individual’s history falls under the three high-level outcomes mentioned above: accumulating human capital, an express transition into work or an outcome of possible concern. Due to the small sample size of those in each of the ‘concern’ pathways, we had to treat these as a single group rather than analyse each sub-group individually. It is worth pointing out that the ambition is to establish the presence of strong correlation rather than causality. The former is in fact a sufficient condition to identify in advance who is at risk of an unsuccessful transition into the labour market and is what will therefore be of interest to policy-making.

<table>
<thead>
<tr>
<th></th>
<th>Human capital</th>
<th>Express</th>
<th>Possible cause for concern</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (ref: males)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>-0.005</td>
<td>-0.057</td>
<td>* 0.062 ***</td>
</tr>
<tr>
<td></td>
<td>[0.024]</td>
<td>[0.027]</td>
<td>[0.016]</td>
</tr>
<tr>
<td>Ethnicity (ref: white)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-white</td>
<td>0.21 **</td>
<td>-0.157 **</td>
<td>-0.053 *</td>
</tr>
<tr>
<td></td>
<td>[0.060]</td>
<td>[0.061]</td>
<td>[0.023]</td>
</tr>
<tr>
<td>Parental qualifications (ref: Low)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High (degree)</td>
<td>0.292 ***</td>
<td>-0.184 ***</td>
<td>-0.108 ***</td>
</tr>
<tr>
<td></td>
<td>[0.044]</td>
<td>[0.048]</td>
<td>[0.028]</td>
</tr>
<tr>
<td>Medium (GCSE A-C)</td>
<td>0.111 ***</td>
<td>-0.04</td>
<td>-0.071 ***</td>
</tr>
<tr>
<td></td>
<td>[0.031]</td>
<td>[0.035]</td>
<td>[0.021]</td>
</tr>
<tr>
<td>Housing tenure (ref: owned)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social rented</td>
<td>-0.133 ***</td>
<td>0.08</td>
<td>* 0.052 *</td>
</tr>
<tr>
<td></td>
<td>[0.035]</td>
<td>[0.039]</td>
<td>[0.022]</td>
</tr>
<tr>
<td>Private rented</td>
<td>-0.225 ***</td>
<td>0.243 ***</td>
<td>-0.019</td>
</tr>
<tr>
<td></td>
<td>[0.054]</td>
<td>[0.062]</td>
<td>[0.035]</td>
</tr>
<tr>
<td>Year of birth (time trend)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.002</td>
<td>-0.001</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>[0.003]</td>
<td>[0.004]</td>
<td>[0.002]</td>
</tr>
<tr>
<td>Month of birth (ref: May-Aug)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jan-Apr</td>
<td>-0.043</td>
<td>0.056</td>
<td>-0.012</td>
</tr>
<tr>
<td></td>
<td>[0.029]</td>
<td>[0.032]</td>
<td>[0.019]</td>
</tr>
<tr>
<td>Sept-Dec</td>
<td>-0.045</td>
<td>0.079</td>
<td>* -0.033</td>
</tr>
<tr>
<td></td>
<td>[0.030]</td>
<td>[0.033]</td>
<td>[0.018]</td>
</tr>
<tr>
<td>Health (ref: no limitations)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health limits daily activities</td>
<td>-0.145 **</td>
<td>0.175 **</td>
<td>-0.03</td>
</tr>
</tbody>
</table>
Estimated average marginal effects are presented in Table 3. The numbers shown indicate the estimated percentage point change in the probability of a given individual entering the named trajectory when possessing a given characteristic as opposed to the reference value. 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 just under 28 percentage points higher than for an otherwise identical individual whose parents’ highest qualifications are at most GSCEs graded D-G. Standard errors are shown in brackets.

School attainment (grades), family background (parental qualifications, housing tenure) and gender emerge as the strongest predictors of labour market outcomes. These results, together with the large and relatively homogenous groups found to invest in education or enter work straight from education, are consistent with evidence indicating that high-achieving and

A number of additional 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. This is consistent with higher staying on rates among certain minority ethnic youth (Middleton et al. 2006). Similarly to what is found in (Crawford et al. 2010), youth born between September and December, who are therefore the oldest in their classroom, 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, individuals living in local authorities with high unemployment relative to other areas are found to be more likely to enter a possible cause for concern trajectory and less likely to be investing 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 effect of older sibling labour market status 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), Becker & Tomes (1976), Willis (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 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). The breaking down of the postulated quantity-quality trade-off may be explained by seeing human capital acting as a public good within the family and by negative resource effects being counterbalanced by socialisation advantages. Furthermore, our results differentiate according to the labour market status of the older sibling and hint at a correlation across sibling status, possibly evoking a role model effect.

We also explore the relationship between self-confidence and motivation problems and labour market trajectories. A growing body of literature has explored the effect of such non-cognitive skills on wages, and more recently on 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

---

10Unfortunately, it was not possible to analyse sub-groups within the non-white population due to the small sample size.
indicating the number of ‘negative’ responses to the eight questions in the survey (Waddell 2006). 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 seem to confirm that self-confidence and motivation problems have a lasting effect on future outcomes, increasing the probability of accumulating human capital, increasing the probability of being in the ‘possible cause for concern’ group but reducing the probability of being in the Express group. 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 therefore interesting to see that, even when including a measure of 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, which the individual has at least some capacity to influence, and others, such as family background and gender which are predetermined. The importance of the latter factors sheds light on the extent to which structural inequalities may play a role in determining individual’s pathways, and therefore the reproduction of the inequalities themselves. Indeed, there will be cases where individuals exhibit more than one of the above ‘risk factors’, thereby giving rise to a greater polarisation than can be inferred from Table 2. For example, educational achievement is likely to be correlated across generations. Indeed, in our sample, around 90% of youth with highly educated parents obtain GCSEs at grades A-C at age 16, while this figure is only around 54% for those with parents having only grades D-G at GSCE or equivalent. Similarly, it is well known that qualification levels are closely related to employment stability and labour market attachment. Highly educated parents are therefore also more likely to be employed parents. As a consequence, it is likely that the effects of school attainment and parental education and employment will combine and reinforce each other. Indeed, using the model results, we estimate that while virtually no young males, with grades A-C at GCSE at age 16 and living with highly educated and employed parents will enter a ‘possible 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.

6 Conclusion and policy considerations

The school-to-work transition has increased in complexity over the years. In light of this, this paper has made two contributions to the evidence base on the issue. Firstly, it used optimal matching to examine youth school-to-work experience in a holistic way. Drawing on this technique, we were able take a broad look at school-to-work transitions in the UK
and identify small number of distinct transition patterns. Secondly, we explored which characteristics at age 16 can act as early predictors of unsuccessful trajectories in the labour market. By using the groupings identified in the first part of the analysis, we could focus on what characteristics are associated with successful or unsuccessful outcomes as defined by the nature of one’s overall trajectory post compulsory schooling.

Our results suggest that 9 out of 10 young people experience generally successful labour market trajectories between ages 16 and 21. However, the remaining individuals exhibit a variety of histories that might warrant policy attention. We identify six key groups, which we have called ‘Planned interruption’, ‘Partial recovery’, ‘Long-term worklessness’, ‘NEET from 16’, ‘NEET from 18’ and ‘withdrawals from the labour market’. The subsequent analysis confirmed the importance of school attainment (grades), family background (parental qualifications, parental and sibling labour market status), and gender 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 of the transition from school-to-work.

The ability to identify in advance who is at risk of an unsuccessful transition into the labour market provides important clues as to the type of policy that might be effective and who it should target. Importantly, 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. 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 may be well suited to facilitate successful employee-employer matches. More structured systems combining class-based and work-based training could also contribute to shaping school-to-work transition institutions which provide individuals with clear and structured options to chose from when deciding each next step.

References


URL: http://ideas.repec.org/a/bla/econom/v64y1997i255p387-413.html

URL: http://ideas.repec.org/a/ucp/jlabeco/v28y2010i4p773-824.html

URL: http://ideas.repec.org/a/eee/labeco/v12y2005i4p511-530.html

URL: http://ideas.repec.org/a/sae/somer/v38y2010i3p482-512.html

URL: http://ideas.repec.org/a/ucp/jpolec/v81y1973i2ps279-88.html

URL: http://www.jstor.org/stable/1831106

URL: http://books.google.co.uk/books?id=fQC7Pdr5FLMC

URL: http://esr.oxfordjournals.org/content/23/4/409.abstract


URL: [http://ideas.repec.org/p/ifs/ifsewp/10-06.html](http://ideas.repec.org/p/ifs/ifsewp/10-06.html)


URL: [http://smr.sagepub.com/content/32/1/3.abstract](http://smr.sagepub.com/content/32/1/3.abstract)


URL: [http://books.google.co.uk/books?id=w3bE1kqd-48C](http://books.google.co.uk/books?id=w3bE1kqd-48C)


URL: [http://csp.sagepub.com/content/20/3/283.abstract](http://csp.sagepub.com/content/20/3/283.abstract)


URL: [http://books.google.co.uk/books?id=KA3PPQAACA AJ](http://books.google.co.uk/books?id=KA3PPQAACA AJ)


URL: [http://books.google.co.uk/books?id=qy8FHAAACAAJ](http://books.google.co.uk/books?id=qy8FHAAACAAJ)


URL: http://www.jstor.org/stable/2138807

URL: http://www.ingentaconnect.com/content/routledg/cjys/2002/00000005/00000004/art00004


URL: http://www.tandfonline.com/doi/abs/10.1080/13645570701622025


URL: http://ideas.repec.org/p/stn/sotoec/0014.html

URL: http://ideas.repec.org/p/iza/izadps/dp6143.html


URL: http://www.ifs.org.uk/wps/wp0210.pdf


URL: http://ideas.repec.org/p/oecd/elsaab/90-en.html


URL: http://www.dfes.gov.uk/research/data/uploadfiles/RR628.pdf

URL: http://ideas.repec.org/a/spr/jopoec/v12y1999i2p287-312.html

URL: http://www.jstor.org/stable/2698454


URL: [http://esr.oxfordjournals.org/content/21/5/427.abstract](http://esr.oxfordjournals.org/content/21/5/427.abstract)


URL: [http://dera.ioe.ac.uk/11328/](http://dera.ioe.ac.uk/11328/)


URL: [http://dx.doi.org/10.1093/ei/cbj005](http://dx.doi.org/10.1093/ei/cbj005)


URL: [http://ideas.repec.org/a/ucp/jpolec/v81y1973i2ps14-64.html](http://ideas.repec.org/a/ucp/jpolec/v81y1973i2ps14-64.html)