Longitudinal and Life Course Studies 2016 Volume 7 Issue 4 Pp 320 – 346 ISSN 1757-9597

Order or chaos? Understanding career mobility using sequence analysis and information-theoretic methods

Marc A. Scott New York University, US
marc.scott@nyu.edu
Matthew Zeidenberg Abt Associates, US

(Received June 2015 Revised March 2016) http://dx.doi.org/10.14301/llcs.v7i4.358

Abstract

We examine the careers of a nationally representative U.S. cohort of young adults using sequence analysis and information-theoretic techniques to describe these careers' structure and how this structure might inform differences in wage mobility. We operationalise the career as a sequence of industry-occupation pairs observed quarterly. We investigate how the content of these pairs and their organisation over time relate to future mobility. We perform the analysis across three different mobility groups, one of which is characterised by persistent low-wage work. Contrary to what one might expect, low-wage work is not typified by a lack of structure, even in many of the careers in which the worker is weakly attached to the labour market. Using clustering techniques customised to this problem, we build a typology of careers within three groups of workers defined by their wage mobility. We find significant variation within, as well as similarity across the three groups, enhancing our understanding of careers with different levels of mobility.

Keywords

Sequence analysis; information theory; optimal matching; categorical data clustering; intragenerational mobility; national longitudinal survey.

Introduction

The last quarter of the 20th century witnessed a tremendous growth in the share of the U.S. workforce that has been unable to increase their wages over a substantial portion of their work lives (Andersson, Holzer & Lane, 2004; Bernhardt, Morris, Handcock & Scott, 2001; Boushey, 2005; Strawn & Martinson, 2000). In absolute terms, the proportion stuck in low-wage careers in their late 30s is large, nearly 30% of the workforce.[i] Researchers have explored the theory that early careers are characterised by a somewhat chaotic period of job matching, followed by wage growth that results from a successful match and increased job-specific experience. While wage mobility is by definition limited for workers who are stuck in low-wage jobs, some researchers have examined whether job shopping (and job matching) or increased experience are central to wage growth in this job sector (French, Mazumder & Taber, 2007; Gardecki & Neumark, 1998; Topel, 1993). Other researchers have examined trends in job attachment itself, which has been in decline (Bernhardt, Morris, Handcock & Scott, 1999; Kalleberg, 2009; Neumark, Polsky & Hansen, 2000), and the increased lack of control over work schedules (Clawson & Gerstel, 2014). Some of these stylized characterisations of low-wage careers emphasise their chaotic nature and contend that they are not careers at all, as that term is commonly used. These characterisations stand somewhat in contrast to case-studies of low-wage work, such as Appelbaum, Bernhardt and Murnane (2006) and earlier work by Becker and Carper (1956), which suggest that low-wage careers are often found within well-defined sectors of the labour market and individuals identify their career with a particular sector.

This apparent lack of consensus in the literature characterising low-wage careers motivates our detailed examination and measurement of the extent of *order* or *disorder* in careers in three different segments of the labour market, with a focus on low-wage, or immobile careers. Specifically, we use a life course analysis approach to characterise low-wage careers and determine which, if any, features distinguish them from their more mobile counterparts. The central research questions we address are:

- Low-wage careers are sometimes characterised as chaotic. We evaluate this by operationalising and evaluating several related but distinct measures of structure on career paths.
- To what extent is the characterisation due to an inappropriate unit of analysis? What do we learn from individual versus aggregated trajectories?
- How do we effectively incorporate time into any assessment of sets of careers?

To address these questions, we utilise techniques developed in life course studies, incorporating them into a framework for comparing and contrasting longitudinal nominal sequences by employing information-theoretic measures, some of which are well established in the life course analysis literature, while others are introduced or customised by us.

In contrast to much of the previous work on low-wage labour (Dickens & Lang, 1985; Mincer, 1958; Mishel, Bernstein & Allegretto, 2007; Osterman, 1975), we examine the individuals' career trajectories as a whole. Specifically, we examine the sequence of industry and occupation pairs that form a career over a portion of the life course (we use a coding scheme that establishes several hundred unique pairs). We measure the coherence or similarity of these pairs to one another across time using information-theoretic and sequence-analytic measures. As suggested by prior research, we examine the role of attachment to the labour force, which we will show is much more *variable* for workers who experience low-wage careers than for other workers. We show that low-wage careers fall into two distinct categories, and we use clustering techniques to tease these out further. Our findings suggest that commonly held beliefs and generalisations about low-wage careers are too simplistic, ignoring substantial groups whose trajectories exhibit clear structure. Moreover, the generalisations obscure or neglect many of these careers' important features, which are revealed through the more nuanced approaches that we develop.

The organisation of this paper is as follows. In the second section, we review the literature on sequence analysis as it relates to the measurement of structure within sequences and between groups of sequences. In the third section, we introduce new measures, which complement or modify existing ones, and establish a time-lagged approach to evaluating structural differences in sequences. We discuss an approach to clustering within mobility groups to establish content-similar sequences and reveal heterogeneity. In the fourth section, we describe the data we will use and discuss the partitioning of subjects into mobility groups. The fifth section applies our techniques to the sequences of labour market entry, movement and exit, and then reassesses our sequence measures in the context of career clusters. The final section concludes with a discussion of the substantive research questions addressed and makes suggestions for future work.

Sequences and life course analysis

In sequence analysis, the outcome is the whole set of nominal states viewed as a discrete time series (Abbott, 1995; Blanchard, Bühlmann & Gauthier, 2014; Cornwell, 2015). Roughly speaking, there are three interrelated methods associated with sequences: visualisation, descriptive summary, and modelling, which we now introduce.

Visualisation of sequences is challenging, as the purpose of a good graphic is to convey just the right amount of information. Each sequence is an ordered list of tokens (nominal states), forming something akin to a sentence. In the usual situation in which there is no natural ordering to the nominal states, sequences cannot be plotted in the coordinate plane. A workaround is to aggregate the tokens across subjects at each time point, forming a moving window of token or state distributions, and these can be colour-coded to reflect content difference (see, for example, Fasang & Liao, 2013; Piccarreta, 2012). The R library TraMineR (Gabadinho, Ritschard, Müller & Studer, 2011; R Core Team, 2014), which we use extensively in our analysis, has excellent functionality for this purpose and allows for group-specific summaries. While informative, the typical state distribution plot masks one of the key features of a sequence: the dependence between states at different times. One way this could be visualised borrows techniques from network analysis, in which the states are nodes on a network, and edges are movements between states (see Bison, 2014; Hoff, Raftery & Handcock, 2002). A slightly different approach clusters the sequences to establish subgroups that are homogeneous with respect to content and dependence structure. Following this, the medoid or some other summary of each cluster may be compared and contrasted across clusters.

Sequence summaries are an active research topic, so we highlight only a subset of approaches and discuss the need for some extensions or modifications for our research problem. One summary that receives substantial attention is entropy, evaluated cross-sectionally for subgroups (see Billari, 2001; Pierce, 1980). We will make the definition formal in the next section, but the basic idea is that the more uniform the distribution of tokens for a given group or time, the more difficult it is to succinctly describe or make predictions of temporally nearby content. The more concentrated the distribution, the easier it is to predict or describe. Predictive capacity can be operationalised as within or between sequences, and the distinction is important. When we are characterising a mobility group, or a subgroup (cluster) within it, we are referring to the entropy of the ensemble of job types for that group or subgroup. In our example, if low-wage careers consisted entirely of the same industry and occupation pair (token IxO, hereafter) at any given time, they would be low entropy. Entropy could be a measure of structure, but as noted cross-sectional information ignores relationships *within* a sequence; pooled sequence content could be predictable as a whole whereas it could be highly variable within each sequence. To characterise the volatility within a sequence we turn to the notion of turbulence (Elzinga & Liefbroer, 2007; Elzinga, 2010). Turbulence is based on the number of distinct subsequences that may be derived from a given sequence, rescaled to reflect the potential maximal variation in token spell length within it. Intuitively, sequences that do not change tokens frequently tend to be low turbulence while those that have a variety of token spell lengths are higher turbulence. After measuring turbulence for each sequence, we can aggregate the measurements by subgroup. Prior research suggests that the timing of token progression is crucial to career development, thus we measure turbulence on subsequences reflecting different periods (e.g., early, middle and late career).

In the next section, we introduce one additional measure to directly assess the extent to which tokens co-vary within subgroups or sequences. It is a natural analogue to entropy known as mutual information (MI). An important distinguishing characteristic of MI is that it relates one distribution to another. We interpret this in the context of careers as a measure of aggregate similarity, predictability or dependence of one portion of an ensemble of sequences (e.g., all clerical workers) on a latter portion.

Statistical models for sequences are a somewhat controversial topic (for example, see Abbott & Tsay, 2000 and Wu, 2000). Important to this debate is whether sequences derived from the life course are to be viewed through the lens of event history modelling (i.e., time to event models) or processed somehow as a whole sequence, intact. When the state space is large, as it is in our labour market application, models must make simplifying assumptions in order to be tractable, and these assumptions are often questioned. However, event history models can sometimes be customised to address specific research questions. For example, Scott (2011) used a continuous time Markov model to handle a large number of states (tokens) via a Bayesian (random effects) approach, comparing subgroup behaviour and allowing for time heterogeneity. Non-model-based approaches often use techniques of optimal matching (Abbott, 1995; Abbott & Tsay, 2000; Durbin, Eddy, Krogh & Mitchison, 1998) to compute the distance between pairs of sequences, and then use the distances to construct clusters via hierarchical or agglomerative techniques (Everitt, Landau, Leese & Stahl, 2001; Kauffman & Rousseeuw, 1990). One challenge arises in the assignment of costs to the primitive operations of insertion, deletion and substitution in the OMA algorithm (see Vingron & Waterman, 1994, for example). Recent developments in sequence analysis (Elzinga & Studer, 2015; Halpin, 2014; Studer & Ritschard, 2015) have provided methods that address many of these concerns. Of course, clustering has inherent challenges that remain in the sequence analysis domain (see Hennig & Liao, 2013, for some discussion). We will utilise techniques of sequence analysis, OMA and clustering throughout our evaluation of mobility. As we introduce and use each technique, we will justify the methodological choices in the context of our problem domain.

Controlling heterogeneity

Most of the above measures involve pooled sequences, potentially compared over time. This can be misleading when the heterogeneity between subjects is large and the within-sequence link is obscured. We can ameliorate this substantially through the use of categorical data clustering. Clustering, broadly conceived, is a mechanism for partitioning subjects into homogeneous groups. In fact, goodness of fit measures of a clustering often compare the homogeneity within assigned groups to the heterogeneity between them (Everitt, et al., 2001). Clustering is sometimes used to discover a new typology within a subject, but this usually leaves the researcher open to criticism regarding the choices made in producing the clusters. Hennig and Liao (2013) contend that some subject matter knowledge should be brought to bear in any cluster analysis, even when the goal is to learn about latent subgroups in the population. For the purpose of this study, clustering allows us to re-assess our measures of structure within smaller, more homogeneous subgroups. Given that some of our measures are based on aggregated sequences or subsequences, this reassessment of smaller, more homogeneous groups can be illuminating. The literature suggests that careers are attempts by workers to optimise their match with employers (French, et al., 2007; Gardecki & Neumark, 1998; Topel, 1993), so to the extent that we can identify similar patterns in our career sequences, we will consider the clusters to define a career type.

Clustering sequences presents the researcher with this challenge: what is the distance between two career sequences? One way to address this is by first proposing the dissimilarity between two IxO tokens, or job-types. This is a central challenge, because once we have a proper dissimilarity between tokens, which translates into a substitution cost, methods of optimal matching (OMA) and recent extensions of the core methodology can produce the necessary sequence distances (Elzinga & Studer, 2015; Halpin, 2014; Studer & Ritschard, 2016). In our problem domain, we recognise the need for time-varying distances; the likelihood of moving between different job types changes over time (see Scott, 2011, for empirical support; see Halpin, 2014 and Studer & Ritschard, 2016, for related methodological developments to OMA). Equally important in this domain is the need for sensitivity analysis, which is a comparison of findings made under various choices of metric. The approach we take will be discussed in the next section.

Measuring structure over time and clustering similar trajectories

Entropy

Entropy is a useful one-dimensional summary of the *distribution* of tokens used by a subpopulation (Elzinga, 2010; Pierce, 1980). Entropy measures the difficulty one would have guessing a worker's job type in the absence of additional information (given only the frequency distribution of job types for an ensemble of sequences). Formally, entropy is defined as $H(X) = -\sum_i P(X=i)\log_2 P(X=i)$, where *i* varies across the different job types, or IxO tokens, captured in variable *X*.[ii] For categorical data, entropy is zero when only a single type is present and maximised when the distribution is uniform across the set of types.

Choosing the unit of analysis or type of pooling is a key decision one must make when using entropy, as well as most other measures that we discuss. Given our research questions, we first stratify by mobility group. Within each group, aggregating across individuals, we assess entropy at each time point, providing a moving window of the measure, which might reveal change over time. We use entropy to measure the complexity or variety of job types for a group of workers, rather than complexity within a particular worker's sequence.[iii] For labour market data, there is some advantage to smoothing out short-term shifts in distributions, so while we compute entropy at a single time point (cross-sectionally), we smooth those measures using the running median approach of Tukey (1977).

Turbulence and mutual information

Entropy as we have operationalised it is based on pooled observations. The order or disorder in *individual* career paths cannot be assessed using entropy without significant modification, as the total number of distinct tokens in a single sequence may be several orders of magnitude smaller than the full token alphabet. Fortunately, turbulence is a measure that captures the complexity of single sequences. Turbulence is defined in terms of three functions of a sequence *x*, $\phi(x)$, $c_1(x)$ and $c_2(x)$, where $\phi(x)$ is the number of distinct subsequences, $c_1(x)$ is the maximal variance of spell lengths given the alphabet used and the length of the sequence and $c_2(x)$ is the actual variance of spell lengths. Given these terms, turbulence is given by $T(x) = log_2\left\{\phi(x)\frac{c_1(x)}{c_2(x)}\right\}$ (Elzinga, 2010). If turbulence describes the complexity of a sequence, then evaluating it over time, using moving windows, will reveal changes in career structure as it progresses. To understand how turbulence changes over time, we use an age-based moving window; for example, the first evaluation is based on the subsequence that covers ages 20 through 25; we then shift by one quarter to examine the window from age 20¼ to 25¼, and so forth.

Another way to assess structure within a career is by using a measure called mutual information (MI, hereafter). This is typically defined for two random variables, *X* and *Y*, as follows:

$$I(X,Y) = -\sum_i \sum_j P(X=i, Y=j)\log_2 \frac{P(X=i,Y=j)}{P(X=i)P(Y=j)}$$

(Pierce, 1980). In our analysis of career sequences, *X* and *Y* reflect job content at two points in the *same* career, such as the early and middle stages. While the formulas for MI and entropy are similar, the unit of analysis differs for each, with MI defined within sequences and entropy across them, yielding a negative relationship between these two measures. In our study, we take *P(X,Y)* to be the joint probability of observing job type *X* in the early career and job type *Y* later in the career *of the same individual*. If the likelihood of observing *X* and *Y* together in the same career greatly exceeds expectations based on chance alone (the marginal product in the denominator), MI will be large. Thus MI applied to two portions of a career measures the extent to which the content in an early portion of the career predicts the content in a later portion. We aggregate all of these within-sequence co-occurrence probabilities for a prespecified group of subjects, such as chronically low-wage or 'stuck' workers, and evaluate the MI of the joint distribution. We reiterate, however, that each co-occurrence is derived from an individual sequence.

Addressing the heterogeneity of careers

While the above measures should provide substantial insight into career structure over time, the necessary aggregation masks heterogeneity and homogeneity across the sequences. On the other hand, sequence level measures are hard to summarise due to the large number of individuals in the study, and the concomitant heterogeneity. To explore this, we cluster careers with similar job content together within mobility group, and then make comparisons between mobility groups, at the *cluster level*. We will examine entropy, turbulence and MI for these more homogeneous groupings and may obtain results that are different from the aggregate findings. When the basic trends still obtain, we have simply added a layer of robustness to the prior findings. In what follows, we describe the methodology used to obtain clusters.

Clustering typically begins with measuring the distance between objects, represented as multivariate vectors, but there is no obvious metric for comparing the components of two categorical sequences, as we have in our application (the simplest metric, exact matches, is too crude a measure). Another issue in comparing career sequences is the timing of the jobs – should the comparison be made at the exact same time point in each of the sequences, or do minor variations in the timing of jobs matter when defining a typical career pathway? It is clear we must consider timing; in the extreme, one could ignore the ordering of these sequences and just compare IxO frequencies across pairs. Without timing, career sequences have reduced meaning—they become mere collections of jobs.

The Optimal Matching approach

To meet the challenges of clustering these sequences, we divide the problem into two subproblems. First, we must decide how 'near' or 'far' each IxO is from the other to construct a pairwise token dissimilarity matrix. Second, we have to decide how to *align* two careers so that the token-by-token pairings may be compared less rigidly. The first subproblem is the topic of substantial research (e.g., Studer & Ritschard, 2016); optimal matching (OMA) is a fairly well established method, which deals effectively with the second subproblem (Abbott, 1995; Durbin et al., 1998; Lesnard, 2014; Sankoff & Kruskal, 1983).

To use OMA, we must specify the cost of each of the primitive operations, which are substitution, insertion, and deletion. Typically, the insertion and deletion operations (sometimes abbreviated 'indel') are given fixed costs (independent of which token is being inserted or deleted), while substitution depends on the token pair.[iv] We set the maximum cost of substitution to be no greater than the sum of the costs for insertion and deletion; otherwise, there would be no benefit to substitution. In our application, we use three different alternative substitution costs, capturing a range of assumptions and emphases. We take the insertion and deletion costs to be fixed at one. The first substitution cost is *fixed* at two, implying no two tokens are similar enough to warrant a substitution less costly than a deletion followed by an insertion. The next choice reduces the substitution cost of token 'a' for token 'b' by the probability of observing a *transition* between temporally adjacent tokens in either direction (we call this a transition based metric). Namely, substitution cost s(a,b)=2-P(a|b)-P(b|a), for arbitrary tokens 'a' and 'b'. This is implemented in the TraMineR function seqsubm by setting method='TRATE'. One concern with this approach is that most transition probabilities are small, as there are large runs of the same token within a sequence, which will set most substitution costs to nearly two.

We base the third of our substitution costs on a *ratio* commonly used in Biological Sequence Analysis (BSA; Durbin et al., 1998), which has the same functional form as the rightmost term in the mutual entropy calculation. We also allow for change over time by adjusting this cost metric as individuals age. We want to compute the similarity of two tokens, measured by how often they occur together. Consider the joint probability, $P(a,b)$, defined to be the probability of witnessing tokens 'a' and 'b' together in the same career sequence. To adjust for the fact that some tokens are extremely frequent while others are rare, we normalise the joint probability by the marginal product of the token component frequencies. That is, if the probability of seeing 'a' and 'b' together is large compared to the product of their marginal probabilities (under an independence assumption), then we have indication that 'a' and 'b' belong together (in MI parlance, one is predictive of the other). This metric may be written

$$s(a,b) \propto \log\left(\frac{P(a,b)}{P(a)P(b)}\right),$$

where s is the similarity and a and b are arbitrary tokens. Our substitution cost is a rescaled version of this similarity oriented so that the most similar pairs have the lowest substitution cost. The difference between this and the kernel of the mutual information measure resides in how we include time in the definition of all of the probabilities involved. We call this the *likelihood ratio* based metric. The use of this metric in the context of career sequences is novel.

We veer slightly from what is commonly done in BSA and base the joint probability on the empirical transition matrix, and to allow for heterogeneity, we compute these probabilities at each time in the analysis period.[v] Thus, $P_t(a,b)$ (substituted in the formula above) is the probability of observing subsequence {*ab*} beginning at t. The metric $s(a,b)$ spans the entire real line; to transform it to a substitution cost, we refer positive values to a χ_1^2 distribution (large values yield small p-values) and multiply by two so that the least similar substitutions (p=1.0) are as expensive as a single deletion followed by an insertion. Negative values suggest unusually infrequent co-occurrence, so these pairs also receive the maximal substitution cost of two.

Clustering dissimilarities – multiple approaches

Using OMA and the above cost matrices provides us with pairwise distances between sequences. To cluster the sequences, we consider two well-established techniques. The first is a hierarchical clustering approach in which Ward's criterion is used to choose the nearest cluster pair to join (Everitt et al., 2001; Kaufman & Rousseeuw, 1990). The second is a recursive-partitioning algorithm known as partitioning around medoids (PAM; Kaufman & Rousseeuw, 1990). PAM is sometimes called k-medoids clustering, in reference to the k-means algorithm (Hartigan & Wong, 1979).[vi] PAM is similar to k-means in that one pre-specifies the number of clusters (k), and the algorithm (non-hierarchically) finds the optimal partition with that number. The 'centre' of the cluster, the point closest to all others in that cluster, is called the *medoid* and serves as an exemplary 'representative' for the group. PAM algorithms often provide the useful diagnostic 'silhouette width' (SW; a measure between -1 and 1) to assess how well each item belongs in the cluster to which it is assigned (for details, see Kaufman & Rousseeuw, 1990). Roughly speaking, SW is zero or negative for a sequence in which the nearest cluster medoid is not that of the assigned cluster.[vii]

Number of clusters

Using both Ward's method and PAM, crossed with the three different measures of sequence dissimilarity for OMA discussed above (constant, transition based, likelihood ratio based), we form separate clusters within each mobility-based group to identify the common pathways in each. We considered two goodness-of-fit criteria: average silhouette width and the Calinski-Harabasz index. The latter was adapted for dissimilarity matrices by Hennig and Liao (2013). We maximise the fit criterion across different choices for the number of clusters. We view the final set of career clusters as a *typology*, or a way to label similar careers based on their component industry and occupation (along with time out of the labour force) trajectories. The clusters should be relatively *homogeneous within* and *heterogeneous between* each other.

Data and construction of mobility groups

Our data source is the National Longitudinal Survey of Youth (NLSY), which yields a representative sample of non-institutionalised[viii] men and women in the U.S. between the ages of 14 and 21 in 1979 (Bureau of Labor Statistics, 2000). This cohort was interviewed every year from 1979-1994, and then once every two years until the present. The last round in our analysis dataset ended in 2000, when the cohort was between 35 and 42 years old. Blacks, Hispanics and poor whites were oversampled in what are known as 'supplemental samples'. The supplemental sample of poor whites was dropped from this analysis because it was discontinued after 1990, truncating their career sequence prematurely. Poor whites are represented in the common, retained sample, and we have adjusted the weights to accurately reflect their proportionate contribution. A supplemental military sample was dropped from the sample used in our analysis as well. The original sample size, including all supplemental samples, is 12,686. This drops to 9,763 after the two supplemental samples are dropped. After careful evaluation of patterns of missed interviews, we decided to remove individuals who show a gap of more than four years between any two surveys.[ix]

As explained above in the first section, when we refer to the 'career', we mean a sequence of industries and occupations. We build this sequence from quarterly jobs spanning ages 20 to 36 using the individual's work history (the NLSY constructs a *weekly* job history from the questionnaire, and we define the quarterly job to be the job held in the first week of each quarter). We formulate 25 unique industry and 20 unique occupation codes that aggregate the three-digit 1970 Census Classification codes into reasonably homogeneous groups.[x] For example, all durable goods manufacturing industries are collapsed into a single code. In addition, unemployment, educational enrolment, and time spent out of the labour force are coded into the sequence. In the 16-year age span studied, approximately 450 unique industry and occupation pairings (IxOs) occur. These codes can be understood as ideal types (Weber, 2009), or coarse nominal groupings.[xi]

For our restricted sample, the key variables of industry and occupation for jobs recorded in the work history are missing at about 3%, which is minimal. About 100 individual cases have missing industry or occupation information for more than half of the work history, so these cases are dropped. The sample size becomes 7,816, or 80% of the maximum possible. We reweight the sample so that it is consistent with the demographics of the original baseline sample.

Mobility group definitions

For each respondent, longitudinal wage profiles were constructed using inflation-adjusted, logged hourly wages associated with the CPS job[xii] at the time of each annual interview. Taken as a whole these wages form a profile of growth or stagnation over time. The inflation adjustment is made using the Consumer Price Index research series (CPI-U-RS; see Bureau of Labor Statistics, 2015).

The wage profiles were cleaned of short-term wage fluctuations by substituting, for the original observations, best linear unbiased predictions (BLUPs; see Robinson, 1991) from a longitudinal mixed-effects model. Each predicted profile could be understood as the 'permanent' wage level over a broad time span. Gottschalk and Moffitt (1994) discuss the theory behind such permanent and transient wage decomposition and provide a methodology for their identification. We use a slightly different method to generate permanent wage trajectories, following Bernhardt et al. (2001). Using age as the underlying timeline, we found that a quintic mean structure, quadratic random effects and year-specific variances represented the data best, given the variation observed. The Bayesian Information Criterion (BIC; Kass & Raftery, 1995; Schwarz, 1978) was used to select these model components, which were also guided by and were similar to those used by Murphy and Welch (1990).

We classify each permanent wage as either *low* or *not low* using a poverty line threshold. Wages below 1.25 times the poverty line for a family of four (converted from annual income to an hourly wage) in a given year are considered low.[xiii] At the end of the period examined, 2000, this was about \$11.00/hour; for reference, in 2015 this is about \$14.50/hour. An additional twenty-five individuals were dropped from this analysis due to inconsistencies in their wage profiles, including severely outlying wages.

We divide workers into three mobility groups based on their permanent wages at age 24 and 38.[xiv] The first age reflects a point in the life course at which most individuals have entered the labour force, and the latter is a point at which most family formation, if it is to occur, has begun. Intragenerational mobility occurs over the life course, within a worker's career; intergenerational mobility occurs from parent to child. Each individual was classified as belonging to one of the following three intragenerational mobility groups:

- *Stuck*: wages are low at age 24 and are still low at age 38
- *Mobile*: wages are low at age 24 but are no longer low at age 38
- *Never low*: wages are never low, at either age.

Seventy-nine workers were found to be downwardly mobile, but this category was dropped due to small sample size. The final sample size is 7,712.

Under our poverty-based definition, fully 28% of the sample is stuck in low-wage jobs over the career. Another 33% begin working in low-wage jobs, but then escape them by mid-career. Fully 39% manage effectively to avoid low-wage jobs altogether, throughout the career. We find these numbers striking, showing both a significant amount of immobility and a significant amount of mobility out of low-wage jobs.

Basic descriptive findings

We have examined many demographic and education measures across career and worker dimensions, and not surprisingly, these differ across the three mobility groups in expected ways. For example, stuck workers are more often female, less-educated, and African American. Never low workers are more often male, white, and college-educated.

This is consistent with the literature on low-wage workers in the U.S. (Andersson et al., 2004; Appelbaum et al., 2006; Boushey, 2005; Danziger, Blank & Schoeni, 2008; Mishel et al., 2007).

We turn to markers of career structure and find that the situation is more complex than one might think. In aggregate, both time spent out of the labour force (OLF) and unemployment spells are a very strong marker of low-wage careers (see also Bernstein & Hartmann, 2000 and Topel, 1993). Stuck workers are OLF (and not enrolled in school) 36.2% of the 16 years examined, while mobile and never low workers average 18.7% and 9.7%, respectively. However, such figures mask tremendous variability. Stuck workers, in particular, have great variation in their attachment to the labour force: at the 75th percentile, nearly 60% of a stuck worker's career is spent OLF, while at the 25th percentile, that figure drops to about 15%, which is close to the average experience of mobile and never low workers.

Another important feature of these careers that distinguishes groups from each other are the job types themselves. While attachment to the labour force, on the whole, is weaker for low-wage workers, the remaining (non-OLF) portions of the career sequences differ as well. We evaluate the similarity of IxO distributions at three different stages of the career using Pearson's χ^2 test of association, pooling the tokens across subjects but within mobility groups. The three stages are early, middle and late career, and are defined as six-year intervals. We do not include all 457 token types; rather, we remove the non-working tokens, and limit our analysis to the top *K* most frequent tokens, with *K*=10 or 50. The latter, being most common, cover the majority of all jobs and the former considers overlap in a smaller subset of common jobs. When all three mobility groups are included, there is clear association between the groups and IxO patterns at each stage (p<0.001; simulation-based, using Hope, 1968). When we restrict this to the stuck and mobile groups, it still holds.

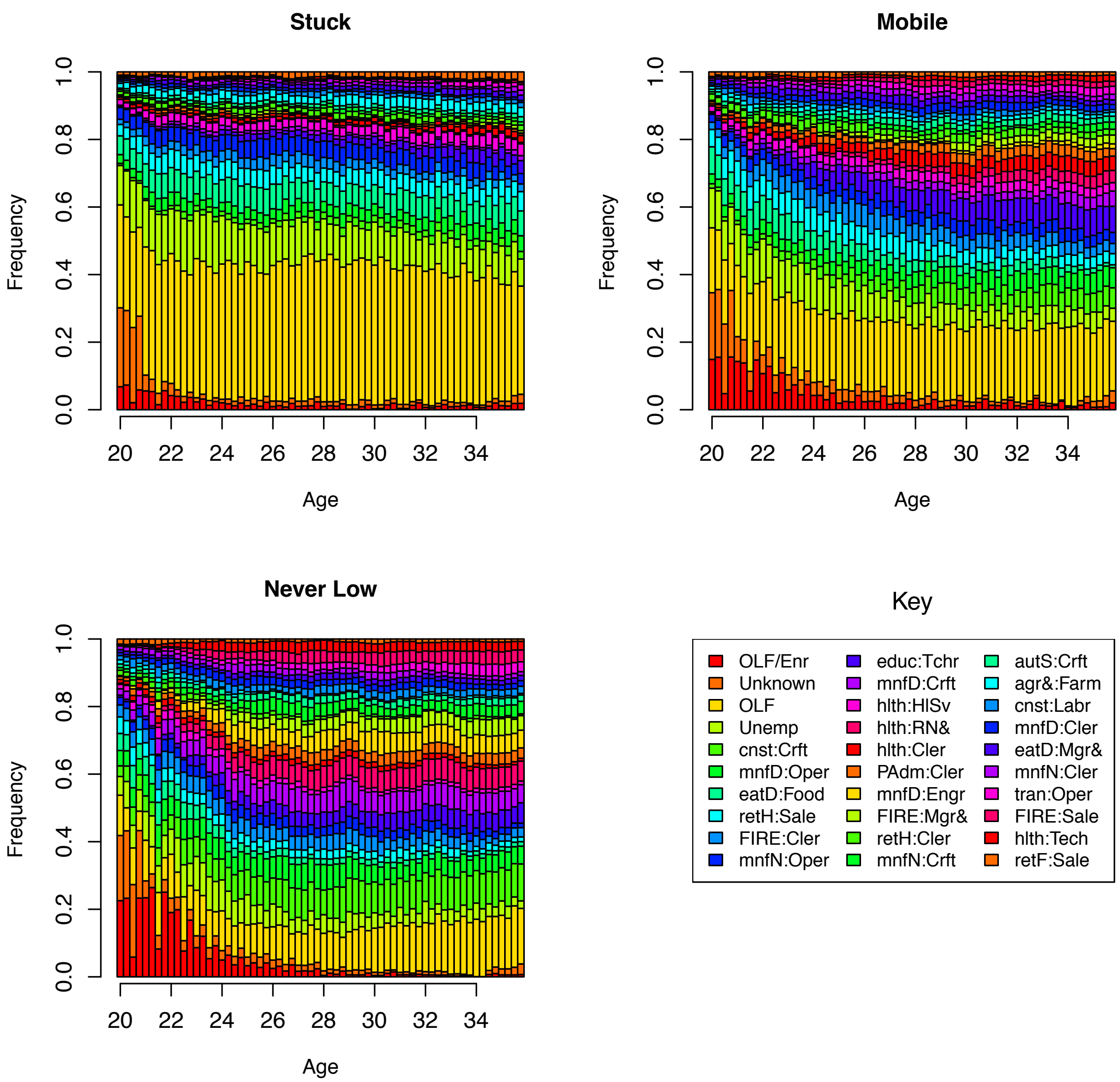
While there are clear differences between these groups, many traditional measures that could be used to describe structure are strikingly similar in all three mobility groups. For example, in aggregate, all three have about the same number of industries, occupations, and IxO pairs across ages 20 to 36. That is, individuals accumulate about the same number of unique job types over this portion of their careers. Equally noteworthy, stuck and mobile workers accumulate about the same number of employers, in aggregate. To understand these findings more completely, we need to disaggregate these workers into meaningful subgroups and consider change over time. As an example of the latter point, the aggregate number of employers masks an important change over time: stuck workers begin their careers with fewer employers but end with slightly more than their more mobile counterparts. However, this is only part of the story, because there is great variation in the number of employers among these workers, suggesting that meaningful subgroups of more and less attached low-wage worker careers evolve over time.

Results

Visualisation

Visualising the content of a large number of sequences relies, commonly, on aggregation, but this can still be revealing. In figure 1, we colour-coded the top 30[xv] of 457 IxO pairs (including labour force status) using fourteen colours (equispaced spectrum-wise, using R's rainbow function), repeated to cycle through those colours, and display the *content* of careers in three mobility groups over time following the approach of Piccarreta (2012) and Fasang and Liao (2013). Any token outside of this prespecified range is removed, so the state distributions can be understood as conditional on work being in the smaller set of tokens. With this smaller token space, we can utilise a legend, which we provide in the bottom right panel (codes are described in Appendix B). Note that three of the first four colours (these are always on the *bottom* of the state distribution plots) represent time spent outside of the labour force (OLF), and are thus non-working periods, with yellow specifically representing OLF periods in which the subject is not in school nor looking for work. As alluded to previously, OLF periods dominate the content of low-wage careers. We re-examined these state distributions after removing non-working tokens (not shown), and we still see differences in content as depicted by colour distributions and patterns. Even in this reduced set of 30 tokens, it can be seen that the spectrum of colours repeats more fully, with a greater proportion of the lower frequency IxOs present for never low workers, suggesting that they (conditionally) utilise a larger spectrum of job types. This suggests that there is more variety in the types of careers that are obtained. These top 30 tokens account for between 54% and 71% of the full token space, depending on mobility, with the smallest percentage associated with never low workers. This is consistent with our characterisation that these workers explore a wider variety of job types (requiring more than 30 tokens to fully represent them) over their life course.

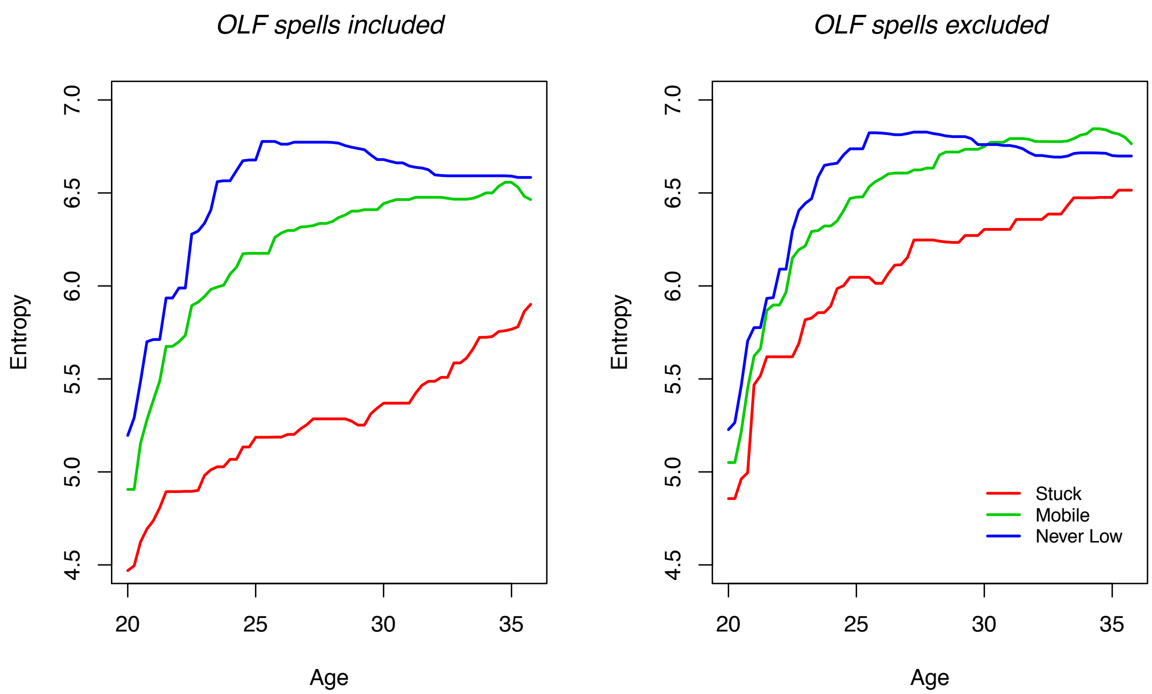
Figure 1. State distributions over time by mobility group



Entropy

The entropy measure is often applied to sequence data cross-sectionally, at every time point. Essentially, we pool job types across workers within mobility group, yielding an empirical distribution of IxOs *for each group* and take the entropy of that distribution. With job IxOs, short-term changes in this distribution can lead to irregular fluctuation in the entropy measure, viewed as a time series, so we smooth it slightly with a running median approach (Tukey, 1977). We measure entropy a second time, after removing quarters of non-work, and plot these in a separate panel.

Figure 2. Entropy over time by mobility group

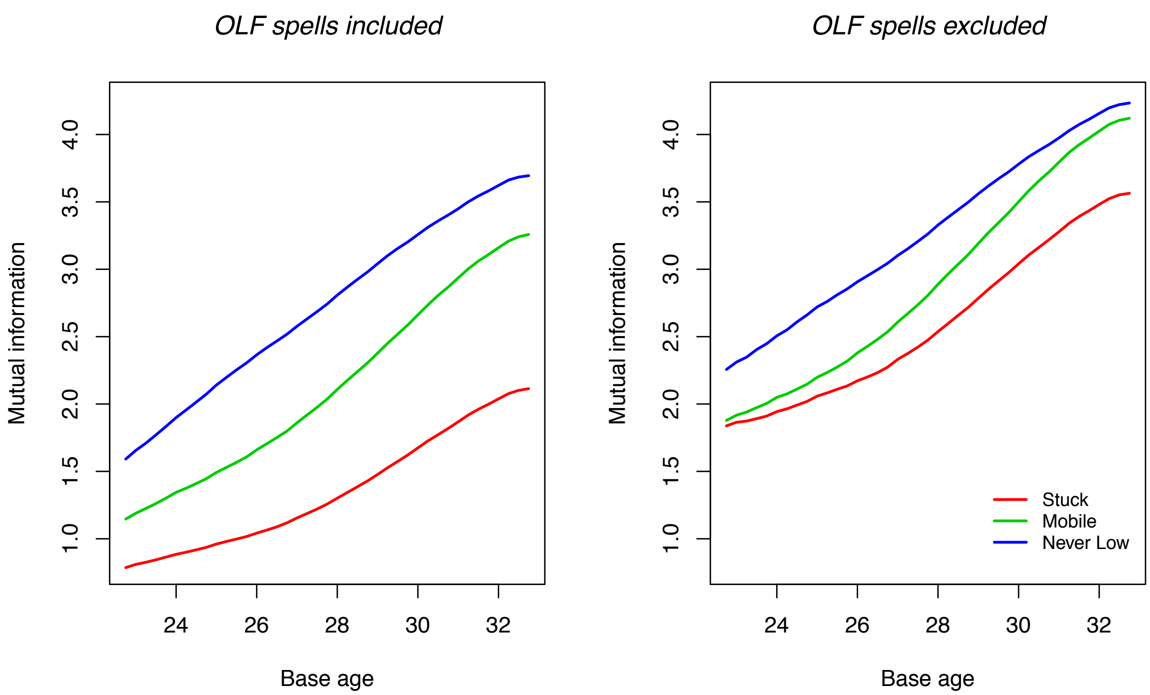


The results are given in figure 2, left and right panels. Surprising, at first, is the *lower* entropy for stuck workers. If lower entropy implies less complexity, then stuck workers appear to have the simplest trajectories. This is consistent with stuck careers being fairly stagnant; the workers in these careers could be failing to 'explore' many of the potential job types in the labour market, and this may be linked to their lack of mobility (note that the sensitivity of our findings to the job classification system is discussed in Appendix A). How the timing and amount of job 'shopping' affects wage growth has been examined in Bernhardt et al. (2001), Murphy and Welch (1990), Topel and Ward (1992) and Gardecki and Neumark (1998), among others. Removing OLF periods of non-work (right panel) increases the stuck career entropy, but relative to more mobile workers, the pattern persists: stuck workers apparently have somewhat less complex careers, in aggregate.

Mutual Information

For the MI analysis, we use a six-year moving window, for a generic subsequence labelled *X*, and we fix the 'target' window for *Y* as the last six years of the career, so *X* gradually moves toward *Y*. Thus, the first evaluation of MI is based on the subsequence from age 20-25 (*X*) compared to 31-36 (*Y*). We then shift by one quarter to compare the windows 20¼-25¼ to 31-36, and so forth, holding the target *Y* fixed. In figure 3, left panel, we show MI as an estimate of the predictability of careers as one gets closer in time, by mobility group, assigning the MI to the midpoint of each window for *X*, so our plot begins at age 22½. What is immediately striking is the similarity between these results and those for entropy, but the implications are different. Recall that MI is evaluated using joint distributions built up from within-sequence information. Stuck workers have the lowest levels of MI across time. The overall trend of increasing MI over time is consistent with our understanding that early careers consist of more job-shopping[xvi] than later careers, which are more settled (Topel & Ward, 1992). Since stuck workers are least predictable (given a subsequence from the past), well into their 30s, we might conclude that as a group, their careers are more chaotic. Even when we restrict the analysis to time spent working (right panel), the basic differences between mobility groups hold, although they are attenuated.

Figure 3. Mutual information over time by mobility group

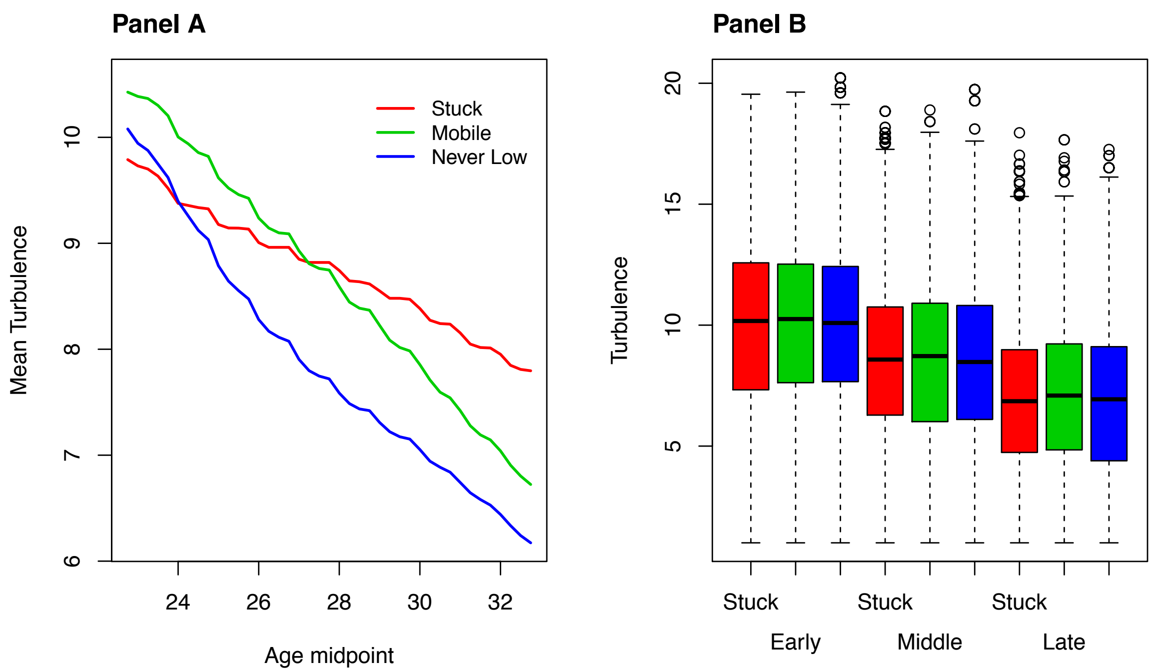


Combining the visualisation with entropy and mutual information measures of structure, we contend that in aggregate, stuck careers explore a smaller spectrum of job types, and that this partially explains their lower entropy, yet this does not translate into predictability. The lack of predictability is consistent with a more chaotic career for stuck workers, partially driven by time in and out of the labour force. Even though the alphabet of IxOs is smaller for these workers, movement back and forth between job types is not forming a consistent pattern or career at this level of analysis. We will learn more from the cluster-based reassessment.

Turbulence

Turbulence is a property of a single sequence, yet we wish to understand differences in turbulence over time and between mobility groups. The latter may be achieved through mean and standard deviation statistics, while the former requires a moving window approach. For this analysis, we use the same longer window of six years so that we witness sufficient sequence complexity.[xvii] We move the window one quarter at a time, generating 41 time points from early to late career, and we median smooth these. For this measure, we do not remove non-working periods for several reasons. First, we wish to consider movement in and out of work and spells of unemployment as part of the structure (or lack thereof) in the career. Second, the introduction of missing tokens creates completely missing subsequences, and turbulence is not defined on these.[xviii] This could result in stuck careers registering relatively low amounts of turbulence (e.g., if the bulk of the career were spent OLF), and we view this as an empirical question.

Figure 4. Mean and distribution of turbulence over time by mobility group



Judging from figure 4, panel A, mean turbulence is much higher early career for all mobility groups, then it drops, presumably as workers establish the core job types that form their careers. Stuck workers may start out with slightly less turbulent careers, on average, but they do not become less turbulent at the same rate as their more mobile counterparts. We do not conduct significance tests on these comparisons, but any gap of 0.25 units may be considered significant at traditional levels, and such a gap exists for most age and mobility group comparisons. Moreover, we wish to document the heterogeneity of sequence turbulence. Thus, in boxplots given in figure 4, panel B, we see that the interquartile range of turbulence within mobility group is fairly consistent over time (it seems to drop slightly for stuck workers). So while turbulence itself is dropping by age 30, the within-group variance persists. By mid-career, stuck workers' careers appear to be more turbulent judging by the mean trend although not by the more robust median. Furthermore, this characterisation masks tremendous heterogeneity present in each mobility group. This is part of what motivates our cluster analysis, to follow.

We have amassed some evidence that stuck workers utilise a more concentrated subset of job IxOs. They explore a limited set of career types, and spend considerable time outside the labour force, in aggregate. It is harder to predict a stuck worker's career from prior content than to do so for a worker in another group. We must be cautious: aggregate measures have the potential to obscure more subtle differences between and within groups. In what follows, we group careers based on their industry and occupation content to determine whether there are more and less coherent careers in the low-wage sector. If stuck careers do not all contain large spells of OLF, then perhaps there is an underclass with low attachment to the labour force and more chaotic careers, while a solid group of stuck workers' histories are quite similar to those of mobile and never low workers. This is an empirical question that can be explored using cluster analysis.

An important finding based on our evaluation of marginal OLF distributions (not shown) is that stuck careers are bifurcated into two distinct types (for a general discussion of dual or segmented labour markets, see: Dickens & Lang, 1985; Hudson, 2007; Osterman, 1975; Piore, 1983). In the first third of the career, the distribution of time spent OLF or unemployed exhibits clear bimodality, with a smaller mode near 100% – these are nearly completely unattached, 20-25 year old workers. This suggests that we may gain new insight by first separating careers into those with similar attachment *and* job content. A deeper study, perhaps both qualitative and quantitative, would form matched career types to examine whether or not stuck careers *in retail*, for example, are different from their mobile counterparts *in retail*. We will use our cluster analysis to highlight potential matched types with similar content across mobility groups to highlight similarities and differences that may yield insight into the structure of careers.

Career sequence clustering

We wish to place careers with similar IxO pairs and structure in the same cluster. To review, the career sequences consist of industry-occupation pairs (such as retail hard goods/sales clerk) for the 64 quarters spanning ages 20-36. Quarters spent unemployed, enrolled in school, or out of the labour force are included with unique codes, since they form an important part of the career structure.

As discussed in the third section of this paper, we evaluated three dissimilarity measures and their corresponding substitution costs, two different clustering techniques and two criteria for choosing the number of clusters. In terms of measures, we compare substitution costs of constant (twice the insertion or deletion cost, which are the same), transition rate based, and likelihood ratio based. We found that the raw transition rates were generally low for these data (most transitions are self-loops, indicated by large diagonal entries in the transition matrix), so we included a fourth variant in which rates were conditional on a non-self-loop transition. In other words, we removed the diagonal and renormalised the transition rates before computing substitution costs for the transition matrix based dissimilarity measure. This correction provided the strongest competitor (and most similarity) to the likelihood ratio based approach, which adjusts for the probability of a transition under independence. Within mobility group, we examined a range of 2 to 64 clusters, and used the Calinski-Harabasz index and average silhouette width to identify the most appropriate clustering (Everitt, et al., 2001; Kaufman & Rousseeuw, 1990). While we always identify an optimal number of clusters for each scenario, most approaches selected either very small cluster sizes, or had great disparity in their sizes. For example, a clustering with 2 clusters for stuck workers and 64 for never low workers, even if 'optimal', was deemed a poor fit on *substantive grounds* – it reveals little to nothing about either group.[xix] With the likelihood ratio based metric, the clustering results had much better properties. Whether we used Calinski-Harabasz index or silhouette width to evaluate, and to some extent whether we used PAM or Ward's method to cluster, the number of clusters and the size of the clusters were comparable across clusterings using likelihood ratio based dissimilarity. We chose the PAM, likelihood ratio based, average silhouette width evaluated clustering as our optimal choice, which yielded 12, 13 and 17 clusters for the stuck, mobile and never low mobility groups, respectively. Comparing this choice to our second best choice using the transition-based measure, we find unadjusted Rand Index (Rand, 1971) similarities of 0.73, 0.88, and 0.93, respectively for stuck, mobile and never low workers. This is evidence of content agreement by cluster, but adjusted for chance indices are much lower (10-30%), so there are real differences as well.

Clustering is often viewed as a subjective exploratory approach (Hennig & Liao, 2013), and while there is always some subjective input provided by the researcher in the form of dissimilarity measures and cluster generation, we use available techniques for evaluating the goodness of fit, and note that visualising the clusters often provides additional face validity. We compared graphs of state distributions over time by cluster and mobility group, and found that most had several clearly dominant, or important, IxOs that characterised the cluster. Given the number of clusters and IxO states, we summarise the clustering in table 1, which provides the size (proportion of workers, by mobility group) and the top four IxOs that are dominant in the state distributions. The four-letter codes for the IxO pairs are given in Appendix B. To facilitate comparison, we name the clusters with a mobility code and number, so that S1 is the first (most frequent) stuck cluster, M3 is the third mobile, N5 is the fifth never low, and so forth.

Table 1. Characterisation of clusters by top four IxO state distributions.

Cluster	*Tot. Freq.*	*Top IxO*	*%*	*2nd IxO*	*%*	*3rd IxO*	*%*	*4th IxO*	*%*
S1	18.6	OLF	31.4	agr&:Farm	7.1	Unemp	3.3	Unknown	1.6
S2	17.4	OLF	21.7	Unemp	8.3	eatD:Food	5.3	OLF/Enr	3.5
S3	17.2	Unemp	9.6	OLF	9.5	Wkg:Msg_IxO	1.9	mnfD:Oper	1.9
S4	13.7	OLF	24.9	Unemp	6.2	eatD:Food	2.3	retH:Sale	2.1
S5	7.5	retH:Sale	18.0	OLF	6.9	Unemp	6.4	retH:Cler	4.9
S6	5.5	eatD:Food	15.5	OLF	13.4	eatD:Mgr&	9.2	Unemp	2.8
S7	4.6	cnst:Crft	25.6	Unemp	4.3	OLF	3.7	cnst:Labr	1.9
S8	4.3	mnfD:Oper	12.6	OLF	7.7	Unemp	3.1	mnfN:Oper	3.0
S9	3.7	mnfN:Oper	23.1	OLF	10.7	Unemp	5.2	mnfN:Crft	2.8
S10	2.7	FIRE:Cler	9.7	OLF	9.6	FIRE:Mgr&	9.1	FIRE:Sale	4.5
S11	2.6	prsl:PrSv	29.9	educ:Tchr	4.8	OLF	3.5	Unemp	3.2
S12	2.3	hlth:HlSv	27.1	OLF	6.4	Unemp	4.6	hlth:RN&	2.1
M1	23.2	OLF	22.0	Unemp	7.3	autS:Crft	4.5	eatD:Food	4.3
M2	14.3	OLF	14.8	OLF/Enr	4.5	Unemp	4.2	Unknown	3.9
M3	13.9	Unemp	22.8	OLF	4.2	tran:Oper	1.5	educ:Cler	1.4
M4	7.0	retH:Sale	5.6	retH:Cler	3.5	OLF	3.0	retH:Mgr&	2.3
M5	6.6	educ:Tchr	20.5	OLF	7.4	OLF/Enr	3.7	prsl:PrSv	3.2
M6	5.7	hlth:HlSv	17.2	hlth:RN&	14.1	OLF	3.5	hlth:Tech	2.1
M7	5.5	FIRE:Cler	12.8	FIRE:Mgr&	10.5	OLF	6.3	FIRE:Sale	5.2
M8	5.4	mnfN:Oper	6.5	Unemp	6.1	mnfN:Crft	3.2	OLF	2.2
M9	4.9	cnst:Crft	14.0	OLF	10.9	cnst:Labr	3.6	cnst:Oper	2.8
M10	4.8	mnfD:Oper	19.7	mnfD:Crft	4.7	Unemp	4.7	OLF	4.6
M11	3.8	hlth:Cler	18.1	PAdm:Cler	3.9	OLF	3.6	Unemp	3.3
M12	2.9	eatD:Food	20.4	eatD:Mgr&	11.8	OLF	4.5	Unemp	2.5
M13	2.0	PAdm:Prot	23.1	Unknown	6.6	Unemp	4.7	OLF	2.8
N1	11.4	OLF	12.6	prfS:Cler	2.9	Unemp	2.8	hlth:Cler	2.4
N2	10.1	mnfD:Oper	18.8	mnfD:Crft	4.9	mnfD:Mgr&	4.4	Unemp	2.6
N3	8.9	cnst:Crft	14.7	cnst:Labr	12.5	Unemp	4.3	cnst:Oper	3.1
N4	8.6	mnfN:Mgr&	3.5	comm:Cler	2.8	OLF/Enr	2.1	Wkg:Msg_IxO	1.9
N5	7.5	FIRE:Cler	15.0	FIRE:Mgr&	10.1	FIRE:Sale	8.1	FIRE:Prof	2.9
N6	7.4	tran:Oper	8.7	Unemp	6.4	OLF	3.8	mnfD:Cler	3.1
N7	7.1	hlth:RN&	20.8	hlth:Tech	6.4	hlth:HlSv	3.4	OLF	2.3
N8	5.5	retH:Sale	3.5	retH:Mgr&	3.1	OLF	2.2	mnfD:Sale	2.1
N9	5.4	educ:Tchr	20.4	OLF/Enr	8.0	OLF	6.7	recr:WrAr	4.4
N10	5.2	mnfD:Engr	7.2	bzSv:Engr	6.5	OLF/Enr	6.1	prfS:Engr	3.5
N11	5.2	mnfN:Oper	22.6	mnfN:Crft	5.5	mnfN:Engr	4.2	OLF	3.8
N12	3.8	PAdm:Cler	19.4	tran:Oper	7.1	util:Cler	3.3	OLF	2.6
N13	3.6	prfS:Prof	13.8	OLF/Enr	9.9	prfS:Cler	6.3	OLF	2.9
N14	3.3	OLF	18.1	retF:Sale	5.0	Unknown	4.7	retF:Mgr&	4.5
N15	3.0	eatD:Food	32.2	eatD:Mgr&	2.2	OLF	2.1	autS:Crft	1.8
N16	2.1	PAdm:Prot	13.4	bldS:Prot	11.3	PAdm:SocW	3.3	Wkg:Msg_IxO	2.9
N17	1.9	util:Crft	12.7	PAdm:Engr	8.6	PAdm:Crft	4.0	OLF/Enr	3.9

In table 1 (first third), we notice that stuck workers have four clusters, S1-S4, in which time spent OLF is paramount. In cluster S3, the most frequent job type is time spent unemployed, but time spent OLF is a very close second. These represent highly unattached workers, but their trajectories seem to differ in subtle ways. For example, the day labourer (e.g., farm worker) job appears somewhat regularly in S1, while food service (e.g., waiter/waitress) jobs appear in S2 and S4. S3 has some connection to durable manufacturing, operative work (e.g., fork lift operator). The remaining eight clusters, S5-S12, accounting for about one-third of the workers, are identified with very specific job types (through the modal IxO): e.g., retail sales, food service, construction, operative work in manufacturing, clerical work, household service (e.g., childcare), health services, respectively. Many of these job types have higher wage equivalents, as we shall see, but they are usually indicated by a different IxO code (e.g., a more professional occupation in the same industry).

Mobile workers' career clusters (second third of table 1) are not as dominated by total time spent OLF, although a solid 51% of careers, as captured by clusters M1-M3, have OLF or unemployment as their most-prevalent token. It is not immediately clear how these three clusters differ from S1-S4 for stuck workers. Visualisation would reveal that these are fairly turbulent careers, with less attachment but still some semblance of IxO content and thus characterisation outside of simply being unattached. Of potential importance is the presence of enrolment in college (code: OLF/Enr), which could provide skills to secure a more stable job in the future. Perhaps more noticeable is the solid dominance of one or two IxO job types in each of the remaining clusters. These clusters are well-described by their modal IxO, which now includes teachers (cluster M5) and medical office managers (cluster M11), in addition to many of the career types in the stuck worker clusters. As an example of how ostensibly similar career types (as determined by dominant IxOs) can differ across mobility groups, we note that in cluster M6, entry-level health service work is eventually replaced, over time, by nursing, technician and administrator jobs, all in healthcare (confirmed via state distributions over time). Such a change is noticeably absent from the comparable stuck worker in healthcare, cluster S12, within which entry level work (hospital attendants, nurses' aides) remains dominant over time.

For never low workers (last third of table 1), perhaps surprisingly, clusters N1 and N14 contain a large proportion of time spent OLF. The variety of types of work that remain (administrative, sales and managerial) seems to characterise this group, despite its weaker attachment. Upon closer inspection using a time-based state distribution plot (not shown), the proportion of OLF periods begins fairly small and grows somewhat over the life course. Given the wage mobility of this group, these are likely intentional exits, perhaps related to family formation (e.g., some of these clusters have a large proportion of females). Also noteworthy is a simple difference in how one might characterise the remaining never low clusters; namely, many have a single industry followed by several occupations within it. For example, N3 are construction workers, N5 work in Finance, Insurance and Real Estate (FIRE), while N2 and N11 are manufacturing clusters. In many of the clusters, occupations shift from sales to manager, with the latter clearly representing a more stable and better-compensated position within a firm.

Alluded to in the above description of the top IxOs for each cluster within mobility group, these clusters inform our understanding of *stuck* careers through the differences in the *variety* of content over time. For example, while the more structured stuck career clusters contain a dominant token describing a job type, there is rarely more than one with any meaningful presence in the distribution. Mobile workers tend to have two or more (working) IxOs with substantial presence. Never low workers often have three or more solid (working) IxOs in a cluster, perhaps most clearly indicating dynamic changes in a career. For example, we witness sales work leading to managerial work in mobile and never low workers, but this progression is not typically observed in a stuck worker's career.

The above highlights several industries and common occupations within them. We note that exploratory analyses, in which clusters are matched – as an amalgam – across mobility groups suggests that nearly all clusters have at least one 'companion' cluster that is more or less mobile. For example, we find stuck, mobile and never low health care clusters, but each is characterised by different mixtures of occupations, and attachment. The stability and other features of the careers will be addressed next. A complete analysis comparing and contrasting the full set of typical careers, focussing on the timing of job changing and the extent to which later stuck careers begin to resemble portions of their more mobile counterparts is left to future work.

Cluster-based reassessment of sequence characteristics

The cluster analysis reveals that mobility groups can be sorted reasonably well based on the tokens representing job types and their dependence relationships. They reveal increasing complexity, in the sense that career paths share a greater variety of content within a cluster, as we move from stuck to mobile to never low workers. The measures of structure that we have evaluated in the aggregate, entropy, MI and turbulence, can be reassessed through the lens of these more homogeneous clusters to complete the characterisation of career mobility groups.

Figure 5 is our entropy measure evaluated separately for each cluster within mobility group (cluster codes are given near the right margin). In this analysis, OLF spells are included, although some mention of what we may learn from their exclusion will be noted. The thickness of each line is proportional to the size of the cluster. Side by side panels allow for same-scale comparison. Note that in aggregate, never low workers have the highest level of entropy (recall figure 2), and this is true as they age. In contrast to this, within cluster, the 17 clusters in the never low group do not seem terribly different from the 13 clusters in the mobile group, in terms of entropy trajectories over time.[xx] There is roughly a bifurcation into higher and lower entropy paths. The higher entropy clusters are the least attached M1, M2 and M3 (thicker lines, representing a larger portion of the workforce) for mobile workers, whereas only one less-attached cluster, N1, is among those with higher entropy for the never low workers. The remaining high entropy clusters are N4 and N6, in manufacturing and transportation, for which MI and turbulence (to be presented in figures 6 and 7) are also quite high, which is suggestive of predictable, but complex pathways. The stuck workers, however, tend to remain lower entropy as an ensemble, even when their homogeneous clusters are the unit of analysis. Note that there is a single stuck cluster that stands out with larger entropy than most others, consistently over time. It is cluster S3, which has substantial levels of unemployment along with time spent OLF. We know that these workers were placed in the same cluster because of this non-working content. This could imply that the remaining content of these careers lacks structure, at least in the aggregate. This suggests that such careers consist of movement across a wide range of IxOs as opposed to concentrating in a specific sector. In contrast and equally noteworthy, many stuck career clusters are in the same basic entropy range as their more mobile counterparts; perhaps these are comparably structured as well? If so, this stands in stark contrast to a common characterisation of low-wage careers as being chaotic; this cluster analysis has revealed structure in many low-wage sectors of employment. Lastly, when the OLF spells are removed (not shown), the less-attached clusters, M1, M2 and M4, show higher levels of entropy later in the career, confirming that OLF spells suppressed the complexity inherent in the content of the working periods in these careers.

Figure 5. Entropy over time by cluster within mobility group

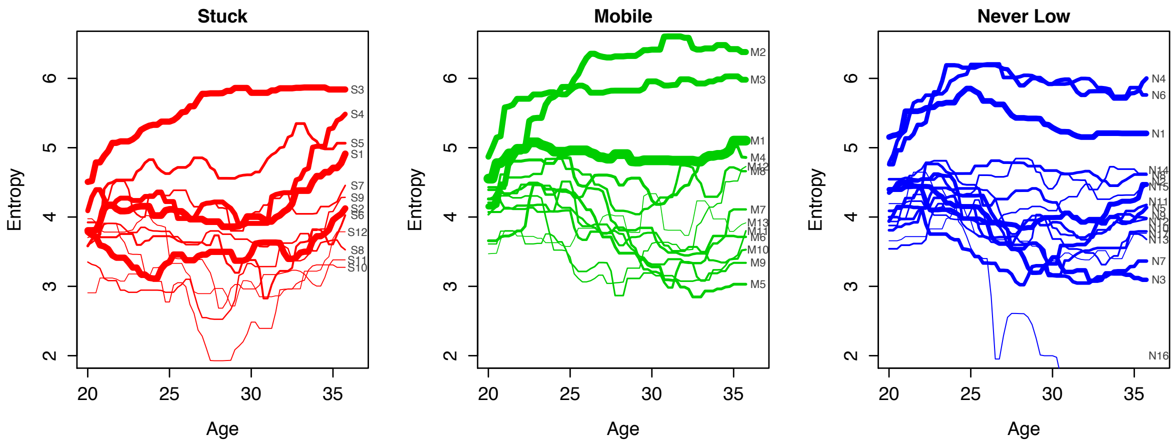


Figure 6 repeats this analysis using the MI measure. Again, we refer primarily to the top panel, in which the content analysis includes OLF spells, and the line width remains proportional to cluster size. Recall that the measure is based on a moving window; when MI increases over time, this implies increasingly similar and predictable content. These curves are much smoother than those for entropy due to the large window used in computing the measure. The most noticeable feature is that while the mean level and trends mimic the aggregate analysis (recall figure 3), with stuck workers at generally lower levels of MI, heterogeneity has emerged with respect to this measure. In particular, never low workers exhibit a wide range of MI, suggesting that some of this group's career types involve much more IxO pair changing than others (closer inspection would reveal some 'steady' career types, such as police officers, are lower MI). An explanation consistent with higher MI clusters is that increased job changing for some never low clusters involves a form of promotion, making their sequence of job types a bit harder to predict throughout the career. It is important to note that clusters are already sorted in a manner that should reflect predictability within elements in the same cluster, so differences in MI reflect differences in predictability, *post-clustering*. In the bottom panel, after removing OLF spells, MI is now much more heterogeneous for the stuck workers. In fact, the mobility groups appear quite similar, when exclusively working periods are considered.

Figure 6. Mutual information over time by cluster within mobility group

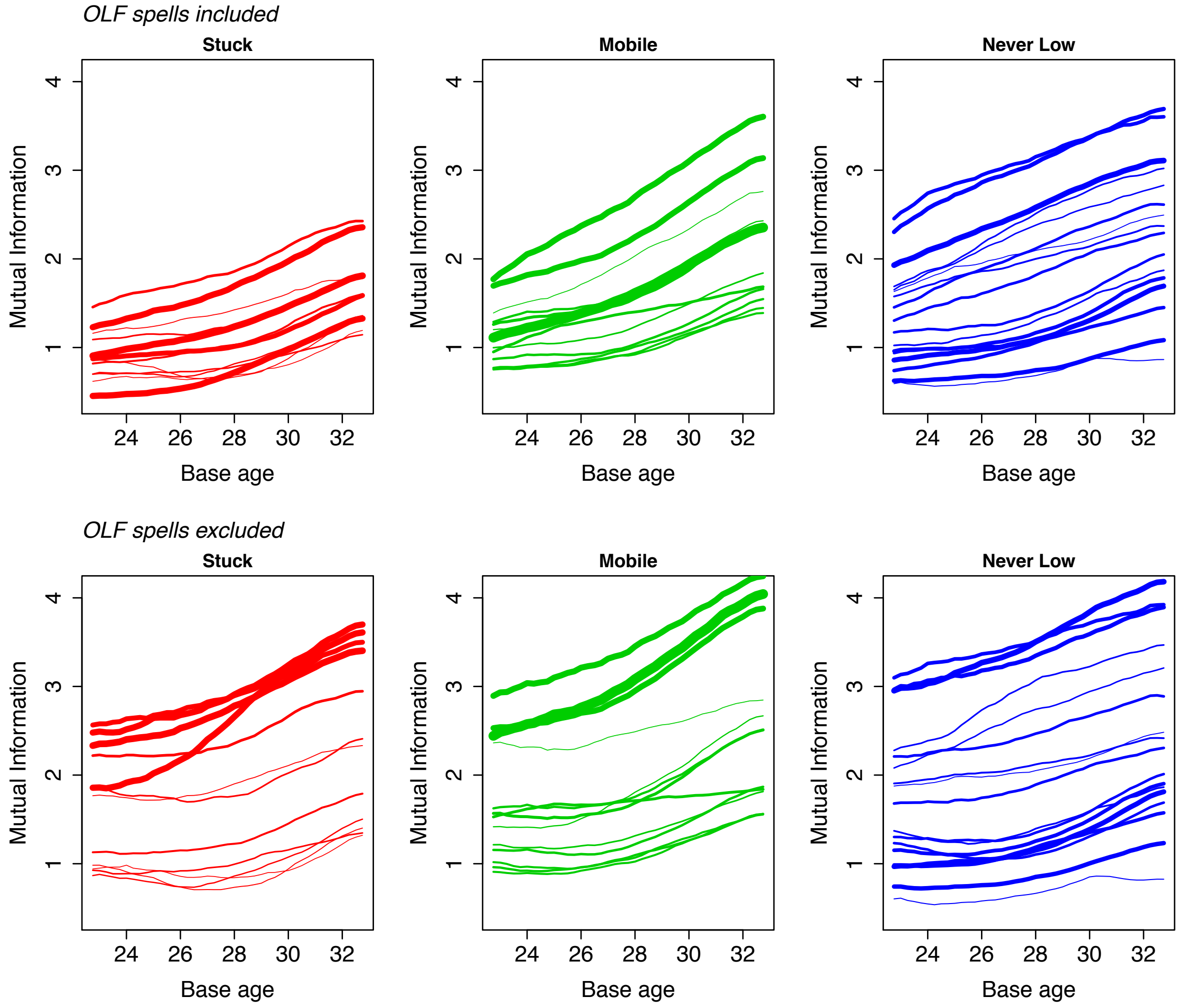
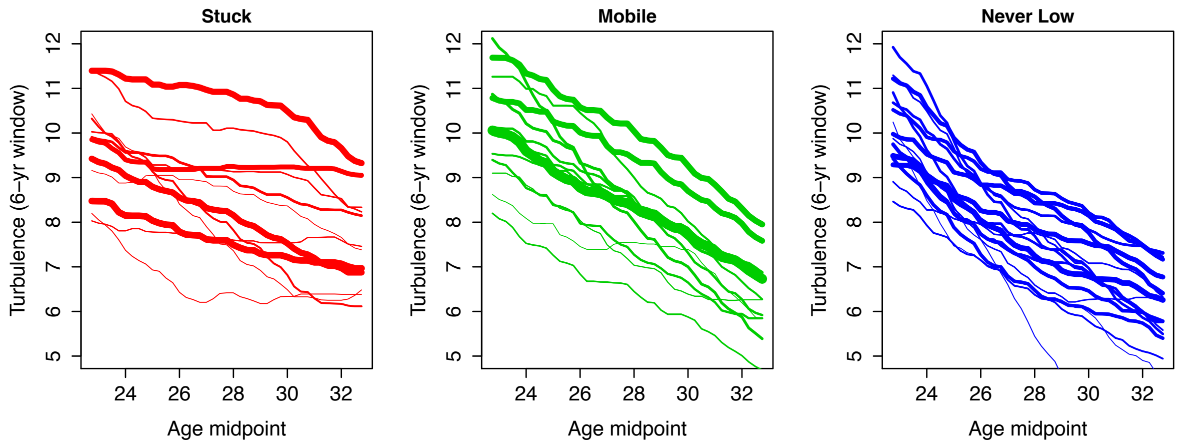


Figure 7 summarises turbulence by cluster within mobility group. The following analysis includes OLF spells.[xxi] While the overall decreasing trends resemble the aggregated analysis given in the third section of this paper (recall figure 4), the variation is again important to note, and it appears to be largest in the stuck group.[xxii] Several larger stuck clusters exhibit very little decline in turbulence over time, suggesting a form of complexity – these are the less attached clusters. At this point, there is mounting evidence that this *subgroup* within stuck careers is more chaotic, but movement in and out of the labour force appears to generate this. For some clusters within the stuck group, there may also be an absence of settling on a more focused path. However, the entropy analysis suggests that there are limitations to the types of jobs in stuck workers' careers – they do not traverse a wide range of job types, which is consistent with dual labour market theory (see Dickens & Lang, 1985; Hudson, 2007; Osterman, 1975; Piore, 1983). This suggests that 'chaotic' is too simplistic a label; less-attached, with reasonably limited exploration, net of this, seems closer to what is observed through this analysis.

Figure 7. Turbulence over time by cluster within mobility group



Discussion

Our characterisation of mobility-based labour markets using sequence-analytic and information-theoretic measures reveals that multiple levels of analysis are important to our understanding of the structure of stuck (and more mobile) careers. We evaluated several research questions regarding the mobility prospects of U.S. workers during a period of increasing wage inequality. One of the key questions was whether the careers of stuck workers were more or less chaotic than those of their more mobile counterparts. Our analysis suggests that a substantial subgroup has weak attachment to the labour force, more movement to and from job types *including types of non-work*, and that these careers are less predictable within any given person. Yet cluster analysis also reveals a sizable set of career types that are highly attached, less chaotic, and which have natural, content-driven ties to career paths associated with more mobile workers, consistent with Appelbaum, et al. (2006). Notably, the stuck group utilises a smaller 'alphabet' of job IxOs, making the type of work more predictable in aggregate. This highlights an important methodological contribution of this research; namely, that the level of analysis, choice of descriptive measure, and the temporal aspect of these must be customised to address the research questions of interest.

Examples of customisation include the following. Two very good existing measures of sequence structure, entropy and turbulence, initially provided very different characterisations, and turbulence was much more informative once we evaluated it using a running window approach (this revealed the natural trend of decreasing turbulence over the career). Entropy as a measure of complexity is too coarse an instrument, but when disaggregated to career clusters, it corresponds more closely to the notion of 'career lines' (see Mouw & Kalleberg, 2006; Spilerman, 1977) or movement to related forms of work, particularly for the more mobile workers. We also introduced an information-theoretic measure, MI, and evaluated it within-sequence for a moving time window. This provided a measure of predictability of future IxOs from the current subsequence; MI appears to be underutilised in the literature. Through the lens of MI, under different levels of granularity (aggregate and cluster-specific), our findings suggest that stuck workers' careers are harder to predict at short and long intervals, while this is only the case in some more mobile career clusters.

This last point highlights the often-neglected feature of careers specifically, and sequences more generally; namely, that substantial heterogeneity exists within them. By employing an OMA-based categorical clustering method to each mobility group, we managed to isolate fairly homogeneous clusters within them. The clusters had face validity, in that industries and occupations paired in a familiar manner (e.g., defining police officers, or nurses, implicitly) in most clusters. Applying the same measures (particularly MI) to the cluster-based subgroups revealed that the never low group has substantially more variability in its (within-cluster) predictability, which was quite unexpected.

By generating clusters, grouping together careers with similar IxO patterns, we have refined our conception of these different labour markets. We have demonstrated many ways in which these mobility groups differ with respect to their IxO progression. We have also found significant similarities between the groups on sequence-analytic measures. This is due in part to the wide variation uncovered through clustering. The clusters, in turn, form many matched trajectory types across mobility groups, providing an avenue for future comparative analysis. We note that we have mentioned demographic differences between mobility groups but not between the career *clusters*. For example, the gender of the worker is likely to be important as we examine mobility differences across matched career types. Further examination of the similarities and differences between these matched trajectory types should lead to deeper understanding of the determinants of mobility in the modern U.S. labour market.

Acknowledgements

The authors wish to thank Annette Bernhardt, Laura Dresser and Joel Rogers for their support and intellectual contribution to this work, as well as the Annie E. Casey Foundation, which provided grant support. The authors also wish to thank the editors and anonymous reviewers for their insights and suggestions.

References

Abbott, A. (1995). Sequence Analysis: New Methods for Old Ideas. *Annual Review of Sociology*, 21, 93-113. http://dx.doi.org/10.1146/annurev.so.21.080195.000521

Abbott, A. & Tsay, A. (2000). Sequence analysis and optimal matching methods in sociology review and prospect. *Sociological Methods & Research*, *29*(1), 3-33. http://dx.doi.org/10.1177/0049124100029001001

Andersson, F., Holzer, H.J. & Lane, J.I. (2004). *Moving Up or Moving On: Who Advances in the Low-Wage Labor Market?* New York: Russell Sage Foundation.

Appelbaum, E., Bernhardt, A.D. & Murnane, R.J. (2006). *Low-Wage America: How Employers Are Reshaping Opportunity in the Workplace*. New York: Russell Sage Foundation.

Becker, H. S. & Carper, J. (1956). The elements of identification with an occupation. *American Sociological Review*, 341-348. http://dx.doi.org/10.2307/2089290

Bernhardt, A., Morris, M., Handcock, M. S. & Scott, M. A. (1999). Trends in Job Instability and Wages for Young Adult Men. *Journal of Labor Economics*, 17(4), S65–S90. http://dx.doi.org/10.1086/209943

Bernhardt, A., Morris, M., Handcock, M. & Scott, M. (2001). *Divergent Paths: Economic Mobility in the New American Labor Market*. New York: Russell Sage Foundation.

Bernstein, J. & Hartmann, H. (2000). Defining and Characterizing the Low-Wage Labor Market. In K. Kaye & D. Smith Nightingale (Eds.), *The Low-Wage Labor Market: Challenges and Opportunities for Economic Self-Sufficiency* (pp. 15-40). Washington, DC: Urban Institute Press

Billari, F. C. (2001). The analysis of early life courses: Complex descriptions of the transition to adulthood. *Journal of Population Research*, 18(2), 119–142. http://dx.doi.org/10.1007/BF03031885

Bison, I. (2014). Sequence as network: An attempt to apply network analysis to sequence analysis. In P. Blanchard, F. Bühlmann & J.A. Gauthier (Eds.), *Advances in Sequence Analysis: Theory, Method, Applications*. (pp. 75-103). New York: Springer. http://dx.doi.org/10.1007/978-3-319-04969-4_12

Blanchard, P., Bühlmann, F. & Gauthier, J.-A. (Eds.). (2014). *Advances in Sequence Analysis: Theory, Method, Applications*. New York: Springer.

Boushey, H. (2005). No Way Out: How Prime-Age Workers Get Trapped in Minimum Wage Jobs. *Briefing Paper.* Washington, DC: Center for Economic and Policy Research.

Bureau of Labor Statistics, U.S. Department of Labor. (2000). National Longitudinal Survey of Youth 1979 cohort, 1979-2000 (rounds 1-19) [computer file]. Produced and distributed by the Center for Human Resource Research, The Ohio State University. Columbus, OH.

Bureau of Labor Statistics, U.S. Dept. of Labor. (2015). Accessed June 18, 2015 from http://www.bls.gov/cpi/cpirsdc.htm.

Clawson, D. & Gerstel, N. (2014). *Unequal Time: Gender, Class, and Family in Employment Schedules*. New York: Russell Sage Foundation.

Cornwell, B. (2015). *Social Sequence Analysis: Methods and Applications.* New York: Cambridge University Press.

Danziger, S.H., Blank, R.M. & Schoeni, R.F. (2008). *Working and Poor: How Economic and Policy Changes Are Affecting Low-Wage Workers*. New York: Russell Sage Foundation. http://dx.doi.org/10.1017/CBO9781316212530

Dickens, W.T. & Lang, K. (1985). A Test of Dual Labor Market Theory. *American Economic Review*, 75(4), 792-805.

Durbin, R., Eddy, S., Krogh, A. & Mitchison, G. (1998). *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. New York: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511790492

Elzinga, C. H. (2010). Complexity of Categorical Time Series. *Sociological Methods & Research*, 38(3), 463–481. http://dx.doi.org/10.1177/0049124109357535

Elzinga, C. H. & Liefbroer, A. C. (2007). De-standardization of Family-Life Trajectories of Young Adults: A Cross-National Comparison Using Sequence Analysis. *European Journal of Population*, 23(3-4), 225–250. http://dx.doi.org/10.1007/s10680-007-9133-7

Elzinga, C. H. & Studer, M. (2015). Spell Sequences, State Proximities, and Distance Metrics. *Sociological Methods & Research*, 44(1), 3–47. http://dx.doi.org/10.1177/0049124114540707

Everitt, B. S., Landau, S., Leese, M. & Stahl, D. (2001). *Cluster Analysis (4th Ed.),* London: Arnold. http://dx.doi.org/10.1201/9781420057492.ch10

Fasang, A. E. & Liao, T. F. (2013). Visualizing Sequences in the Social Sciences Relative Frequency Sequence Plots. *Sociological Methods & Research*, 43(4).

French, E., Mazumder, B. & Taber, C. (2007). The changing pattern of wage growth for low skilled workers. In R.M. Blank, S. Danziger, & R.F. Schoeni (Eds.), *Working and poor: How economic and policy changes are affecting low-wage workers* (pp. 141-72). New York: Russell Sage Foundation.

Gabadinho, A., Ritschard, G., Müller, N. S., & Studer, M. (2011). Analyzing and Visualizing State Sequences in R with TraMineR. *Journal of Statistical Software*, 40(4), 1-37. http://dx.doi.org/10.18637/jss.v040.i04

Gardecki, R., & Neumark, D. (1998). Order from Chaos? The Effects of Early Labor Market Experiences on Adult Labor Market Outcomes. *Industrial and Labor Relations Review*, 51 (2), 299-322. http://dx.doi.org/10.1177/001979399805100209

Gottschalk, P., & Moffitt, R. (1994). The Growth in Earnings Instability in the U.S. Labor Market. *Brookings Papers on Economic Activity,* 2: 217-54. http://dx.doi.org/10.2307/2534657

Halpin, B. (2014). Three narratives of sequence analysis. In P. Blanchard, F. Bühlmann & J.A. Gauthier (Eds.), *Advances in Sequence Analysis: Theory, Method, Applications*. (pp. 75-103). New York: Springer. http://dx.doi.org/10.1007/978-3-319-04969-4_5

Hartigan, J.A. & Wong, M.A. (1979). A K-Means Clustering Algorithm. *Applied Statistics,* 28(1): 100–108. http://dx.doi.org/10.2307/2346830

Hennig, C. & Liao, T. F. (2013). How to find an appropriate clustering for mixed-type variables with application to socio-economic stratification. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 62(3), 309–369. http://dx.doi.org/10.1111/j.1467-9876.2012.01066.x

Hoff, P., Raftery, A.E. & Handcock, M. (2002). Latent Space Approaches to Social Network Analysis. *Journal of the American Statistical Association*, 97, 1090-1098. http://dx.doi.org/10.1198/016214502388618906

Hope, A. C. A. (1968) A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society, Series B,* 30, 582–598.

Hudson, K. (2007). The new labor market segmentation: Labor market dualism in the new economy. *Social Science Research*, 36, 286–312. http://dx.doi.org/10.1016/j.ssresearch.2005.11.005

Kalleberg, A. L. (2009). Precarious work, insecure workers: Employment relations in transition. *American Sociological Review*, 74(1), 1-22. http://dx.doi.org/10.1177/000312240907400101

Kass, R.E., & Raftery, A.E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90(430), 773-795. http://dx.doi.org/10.1080/01621459.1995.10476572

Kaufman, L. & Rousseeuw, P.J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: Wiley. http://dx.doi.org/10.1002/9780470316801

Lesnard, L. (2014). Using Optimal Matching Analysis in Sociology: Cost Setting and Sociology of Time. In P. Blanchard, F. Bühlmann, & J.A. Gauthier (Eds.), *Advances in Sequence Analysis: Theory, Method, Applications*. (pp. 39-50). New York: Springer. http://dx.doi.org/10.1007/978-3-319-04969-4_3

Mincer, J. (1958). Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy*, 66(4), 281-302. http://dx.doi.org/10.1086/258055

Mishel, L., Bernstein, J., & Allegretto, S. (2007). *The State of Working America 2006/2007*. Ithaca, NY: ILR Press.

Mouw, T. & Kalleberg, A. L. (2006). Intragenerational Mobility, Careers, and Task-Specific Human Capital. Paper presented at the annual meeting of the American Sociological Association, Montreal Convention Center, Montreal, Quebec, Canada. Accessed online 29 December 2015, from http://citation.allacademic.com//meta/p_mla_apa_research_citation/1/0/4/7/3/pages104734/p104734-1.php

Murphy, K.M. & Welch F. (1990). Empirical Age-Earnings Profiles. *Journal of Labor Economics* 8(2): 202-229. http://dx.doi.org/10.1086/298220

Neumark, D., Polsky, D., & Hansen, D. (2000). Has Job Stability Declined Yet?: New Evidence for the 1990s. In D. Neumark, ed., *On the Job: Is Long-Term Employment a Thing of the Past?* (pp. 70–110). Russell Sage Foundation.

Osterman, P. (1975). An empirical study of labor market segmentation. *Industrial and Labor Relations Review*, 28(4), 508-523. http://dx.doi.org/10.1177/001979397502800402

Piccarreta, R. (2012). Graphical and Smoothing Techniques for Sequence Analysis. *Sociological Methods & Research*, 41(2). http://dx.doi.org/10.1177/0049124112452394

Pierce, J. R. (1980). *An Introduction to Information Theory: Symbols, Signals and Noise*. New York: Dover Books.

Piore, M.J. (1983). Labor Market Segmentation: To What Paradigm Does It Belong?
The American Economic Review, 73(2), 249-253.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association,* 66 (336): 846–850. http://dx.doi.org/10.1080/01621459.1971.10482356

Robinson, G.K. (1991). That BLUP is a Good Thing: The Estimation of Random Effects. *Statistical Science*, 6(1), 15-32. http://dx.doi.org/10.1214/ss/1177011926

Sankoff, D., & Kruskal, J.B. (1983). Time *Warps, String Edits, and Macromolecules*. Reading MA: Addison Wesley.

Sawhill, E., & McMurrer, D. (1998). *Getting Ahead: Economic and Social Mobility in America.* Washington, DC: Urban Institute Press.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2), 461-464. http://dx.doi.org/10.1214/aos/1176344136

Scott, M. A. (2011). Affinity models for career sequences. *Journal of the Royal Statistical Society: Series C (Applied Statistics),* 60(3), 417–436. http://dx.doi.org/10.1111/j.1467-9876.2010.00752.x

Spilerman S. (1977). Careers, Labor Market Structure, and Socio-Economic Achievement. *American Journal of Sociology,* 83(3), 551-593. http://dx.doi.org/10.1086/226595

Strawn, J., & Martinson, K. (2000). Promoting Access to Better Jobs: Lessons for Job Advancement from Welfare Reform. In Kazis, R. & Miller, M. S. (eds.), *Low-Wage Workers in the New Economy*, Washington, D.C.: The Urban Institute Press.

Studer, M. & Ritschard, G. (2016). What matters in differences between life trajectories: a comparative review of sequence dissimilarity measures. *Journal of the Royal Statistical Society: Series A.* Published online: 14 JUL 2015. http://dx.doi.org/10.1111/rssa.12125

Topel, R.H. (1993). What Have We Learned from Empirical Studies of Unemployment and Turnover? *The American Economic Review*, 83 (2), Papers and Proceedings of the 105th Annual Meeting of the American Economic Association, 110-115.

Topel, R.H. & Ward, M.P. (1992). Job Mobility and the Careers of Young Men. *The Quarterly Journal of Economics*, vol. 107(2), 439-79. http://dx.doi.org/10.2307/2118478

Tukey, J. W. (1977). *Exploratory Data Analysis*, Reading Massachusetts: Addison-Wesley.

Vingron, M., & Waterman, M.S. (1994). Sequence Alignment and Penalty Choice. *Journal of Molecular Biology*, 235:1-12. http://dx.doi.org/10.1016/S0022-2836(05)80006-3

Weber, M. (2009). *The theory of social and economic organization*. New York: Simon and Schuster.

Wu, L. L. (2000). Some comments on 'Sequence analysis and optimal matching methods in sociology: Review and prospect.' *Sociological Methods & Research*, *29*(1), 41-64. http://dx.doi.org/10.1177/0049124100029001003

Appendix A. Sensitivity to industry and occupation codes

Entropy and any other metrics, as they are defined, depend on the industry and occupation coding scheme. For example, if a single clerical occupation could be further refined into two clerical sub-occupations, this would change entropy, and the changes could vary by mobility group. We might be concerned that our finding that stuck workers have lower entropy over time is driven, somehow, by the coding scheme itself. In developing our coding scheme, we took care to collapse relatively homogeneous occupations and industries. We used census-based codes, which attempt to organise the market by what is being produced and the skill set involved in the production (see Weber, 2009).

Lower-skilled work tends to be, at least occupationally, less differentiated. In part, this is because there is simply less complexity in the jobs themselves. However, we have direct evidence that this, in and of itself, is not driving our findings of lower IxO entropy for stuck workers. We re-analysed entropy using only our industry codes and then did this again using occupation codes. For stuck workers, entropy quickly grows over time so that by their mid-20s, stuck workers have the second highest levels of industry-based entropy and the highest levels of occupation-based entropy. Rather than being an artefact of the coding scheme, entropy for stuck workers shows great variety, depending on the job characteristic being evaluated.

Appendix B. Abbreviations used in the figures and tables

Labour Force Status	
Abbreviation	Description
OLF/Enr	Enrolled in school
Unknown	Unknown
OLF	Out of the labour force
Unemp	Unemployed
Wkg_Msg_IxO	Working (IxO not available)*

* Excluded from figure 1 state distribution plot

Industries	
Abbreviation	Description
agr&	Agriculture, forestry, fishing, mining
cnst	Construction
mnfD	Manufacturing durable
mnfN	Manufacturing non-durable
tran	Transportation
comm	Communications
util	Utilities, sanitary services
whol	Wholesale trade
retH	Retail hard goods (except automotive)
retF	Retail food
retA	Retail automotive
eatD	Eating & drinking
FIRE	Finance, insurance, real estate
bzSv	High end business services
bldS	Building services
tmps	Temporary agencies
autS	Automotive & repair services
hotl	Hotels & laundry
prsl	Personal services
recr	Entertainment, recreation services
hlth	Healthcare
educ	Education
nonp	Non-profit
prfS	Professional services
PAdm	Public administration

Occupations	
Abbreviation	Description
Prof	Professionals (Drs., lawyers, etc.)
Engr	Engineers, scientists, engineering technicians, non-health technicians
Farm	Farm labourers, farm foremen
RN&	Nurses, dieticians, therapists
Tech	Health technologists, technicians
SocW	Religious, social scientists, social workers
Tchr	Teachers
WrAr	Writers, artists, entertainers
Mgr&	Managers, administrators
Sale	Sales workers
Cler	Clerical, unskilled workers
Crft	Craft workers and mechanics
Oper	Operatives
Labr	Labourers except farm
Clng	Cleaning service workers
Food	Food service workers
HlSv	Health service workers
PrSv	Personal service workers
Prot	Protective service workers
HHWk	Private household workers

Endnotes

[i] By this stage of the career, *two-thirds* of wage growth has occurred for the bulk of workers (Topel & Ward, 1992).
[ii] When estimating entropy from a sample, we drop tokens that never occur.
[iii] Within-sequence entropy could be computed and then summarised at various levels of grouping – this would more closely resemble the turbulence measure that we use, which is sequence-based.
[iv]Some OMA implementations allow for minor variations on these assignments; we use the version implemented in the R package TraMineR.
[v] This mirrors the transition based approach implemented in TraMineR (in fact, we 'smooth' at the endpoints, as they do), allowing the implied substitution costs to vary with time.
[vi] We cannot use the k-means algorithm, as it relies on a multivariate continuous feature set, while we only have a distance matrix as input.
[vii]The cluster assignments are globally optimal, meaning that moving any sequence out of that assignment reduces the overall goodness of fit. Locally, however, a sequence might appear to 'belong' more to a nearby cluster even though moving it there would reduce the fit.
[viii] These are subjects who are free to seek work. Patients restricted to psychiatric hospitals, incarcerated prisoners and military personnel are considered 'institutionalized' in the NLSY.
[ix] The reasoning behind this is based in part on our unit of analysis, which is the sequence of industries and occupations held by each worker. The NLSY carefully reconstructs the work history for any subjects who miss

interviews, but the reconstruction is limited to the five most recent employers. Five employers are sufficient when the time between current and last interval is short. Across wide gaps, however, there is a concern that the occupation reported will mask prior occupations with the same employer. Thus, we adopt a moderately strict criterion for inclusion in the study.

[x] 1970 rather than 1980 Census codes were chosen because these are the only codes collected consistently over the survey span.

[xi] We discuss the sensitivity of the findings to the choices made in building the token alphabet in Appendix A.

[xii] The CPS job is defined as the current or most recent job. For multiple job holders, their CPS job is the job with the employer at which they usually work the most hours. It is so-called because the wording of the question in the NLSY is similar to that used in the Current Population Survey.

[xiii] The poverty line is a need-based assessment adjusted for household size and ages of children/dependents.

[xiv] This approach is well established in mobility analyses (e.g., Andersson et al., 2004; Sawhill & McMurrer, 1998).

[xv] More precisely, we include four labour force status tokens (OLF, unemployed, enrolled, and unknown) regardless of their frequency, and the remaining 26 are the most frequent job types.

[xvi] The term, job *shopping*, is used by labor economists to refer to the worker's search for a well-suited (skill-matched) employer. The related term, job *hopping*, refers to frequently changing employers, with unclear rationale.

[xvii] One or two years would yield relatively few job transitions, and this is a sequence-based measure.

[xviii] For the cluster-based re-assessment, we restrict the subjects to those who spend less than half of their time OLF, and this makes the computation feasible.

[xix] One could view this as a lack of structure in some mobility groups, but we view it more as a function of the methods used.

[xx] Note that the cluster with entropy that goes below two after age 25 represents police officers, who simply do not explore other careers over this time interval.

[xxi] Removal of OLF spells was non-trivial for the MI measure, in part because it created subsequences that were entirely missing. After resolving some of these technical concerns, the findings including and excluding OLF spells were found to be quite similar.

[xxii] The analysis excluding OLF spells revealed slightly greater heterogeneity in all groups, but it was most pronounced in the stuck group.