Journal of the
Royal Statistical Society
SERIES C
Applied
Statistics

Appl. Statist. (2011)
60, *Part* 3, *pp.* 417–436

Affinity models for career sequences

Marc A. Scott

New York University, New York City, USA

[Received September 2009. Final revision October 2010]

Summary. We develop an affinity model for longitudinal categorical data in which the number of categories is large and we apply the technique to 20 years of labour market data for a contemporary cohort of young adult workers in the USA. The method provides a representation of the underlying complexity of the labour market that can then be augmented to include covariate effects. These can be understood as effects net of transition patterns associated with labour market sorting. We include in our model pairwise affinities to relate nominal categories representing types of job. The affinities capture complex relationships between these types of job and how they change over time. We evaluate the role of gender and education to illustrate the different types of questions and answers that are addressed by this methodology.

Keywords: Categorical sequences; Longitudinal data; Markov model; National longitudinal survey

1. Introduction

In the late 20th century, wage inequality in the USA increased dramatically (Danziger and Gottschalk, 1993; Mishel *et al.*, 2001). This increase in variance was driven in part by a near doubling of the number of low wage workers in the period from the 1980s through the mid-1990s (Bernhardt *et al.*, 2001). Such workers now number between 30 million and 40 million, depending on the definition that is used. Increasing the educational attainment of the workforce is often suggested as a way to improve mobility prospects for workers through increased human capital (see Becker (1993) and Mincer (1974)). Such capital is also increased by long-term relationships with firms, or staying in the same job sector so that industry- and occupation-specific skills may be gained, sometimes through employer-sponsored training. In this type of career path, mobility is often attained through internal labour markets (Osterman, 1984; Neumark, 2000; Osterman and Burton, 2004).

The competing strategies of job changing *versus* long-term employment with the same firm has a complex relationship to mobility. Whereas it has been shown that some amount of early 'job shopping' is essential to wage mobility (Topel and Ward, 1992; Gardecki and Neumark, 1998), it can have both positive and negative consequences later in the career (Neumark, 2000). The interplay between education, industry- and occupation-specific skills and job turnover is less clear. We have evidence that workers eventually settle down and change jobs (more specifically, *type* of job) less frequently as they age, but is this true universally, or does it depend on education, gender or other characteristics? When a worker does change job type, what are the most common pathways of change and are these mediated by worker characteristics as well? Since education is such a strong predictor of wages, we learn more about how inequality

Address for correspondence: Marc A. Scott, 3rd Floor, Kimball Hall, New York University, 246 Greene Street, New York, NY 10003, USA.
E-mail: marc.scott@nyu.edu

 0035–9254/11/60417

plays out—in terms of these pathways—by examining workers with different initial levels of education.

For our analysis, we define the career pathway to be a progression through a series of categorical states, which we take as the industry and occupation of the job at any given point in time. Spells of non-work are included in the nominal categories because they are states through which a worker may pass, but they are of a different character, as they are not a type of work, and as such do not reflect a skill set *per se*. We require *models* for these transitions for several reasons. First, when comparing two groups, one of the ways that they may differ is simply in what types of jobs they typically obtain. Occupational segregation is fairly common in the USA—for example, some types of job are highly 'gendered', such as a registered nurse or elementary school teacher (Blau *et al.*, 2005; Rosenfeld and Spenner, 1992). Marginal differences between genders or between workers with different levels of education may be completely accounted for by the type of work that they are doing, which happens to have distinctly different rates of turnover. We would like to 'net-out' these types of differences before making comparisons.

We explore gender and education differences in the application section. Since 2-year colleges play a major role in the US post-secondary education system, enrolling nearly 35% of college students in 2006 (Provasnik and Planty, 2008), we shall divide our sample into workers who have at least this level of education and those who do not as one of our main comparisons. Second, we are genuinely interested in describing the labour market itself, and how this might change over time. Models are a good way to represent the patterns of turnover that are contained in career sequences; comparing model fits (e.g. by gender) will establish the extent to which different groups traverse these types of job in different ways. They may, of course, reveal that there is a convergence over time.

The career sequence is a marked point process; the marks are types of job and the durations are lengths of spell. We are interested in the whole US labour market, so the number of types of job is large, with perhaps 100 or more unique categories. We assign job types on the basis of a government classification of industry and occupation (hereafter, $I \times O$), resulting in 25 unique industries, 20 unique occupations and about 450 unique pairs of the two. About 100 of the most frequent $I \times O$-pairs account for 90% of all jobs, so we base our analysis on these 100 most frequent $I \times O$-pairs and three types of non-work: unemployment (actively looking for work), enrolled in school (and not working) and out of the labour force (not employed and not looking for work). We are looking for common pathways, so, although we are losing a small amount of information by discarding the least frequent types of job, this information is not central to our understanding. In related work with these data, it was found that very infrequent types of job have the potential to obscure relationships in the same way as an outlier does in regression.

One way that data such as these may be modelled is by using event history or duration models, such as those developed initially in Cox (1972). The multiple end points (job types) and repeated outcomes pose many challenges due to the number of source and destination states and the relationships between them. We address these challenges by using a random-effects formulation, which offers a convenient way to capture the very large number of implicit relationships that exist in careers, including the common pathways and covariate effects. A continuous time Markov process is the kernel of our model for data of this type, as it captures transitions between nominal states. Waiting times can be approximated by exponential distributions, with rates that potentially depend on covariates of interest, as well as the origin and destination state. We shall extend this model to allow for individual-specific heterogeneity, covariates and change over time; the time component because the chain may be non-homogeneous over moderate length intervals.

This paper is organized as follows. In Section 2, we discuss in further detail some alternative approaches for modelling sequence data and motivate our choice of modelling class. In Section

3, we describe the affinity model that captures relationships between career sequence components and handles covariates elegantly. In Section 4, we discuss estimation, inference and model selection. We then apply these methods to our labour market data in Section 5. In Section 6, we summarize the contribution of these models and discuss related directions of research.

2. Modelling sequences

We now mention several methods that are commonly brought to bear on data of this nature and discuss them in the context of our application and research questions. To make these methods as comparable as possible, we first consider the discrete representation of these careers as a quarterly sequence. This interval is sufficiently small to capture jobs of short duration. The nominal types of job observed at each point in the sequence are often referred to as tokens.

One approach for sequence analysis treats the trajectory as a string and uses matching algorithms to establish pairwise distances between each trajectory. In optimal matching analysis (see Abbott (1995)), the user defines costs for substitution, insertion and deletion, and the algorithm finds the set of these operations that transforms one sequence into another with the lowest cost, which is then taken as the pair's distance. The method is used quite heavily in biological sequence analysis (see Durbin *et al.* (1998)). One natural use of these distances is to form clusters, and we have found that this may work well in practice, particularly when care is given to the choice of substitution cost. We note that analysis based on optimal matching is quite sophisticated: other methods for clustering categorical sequences use a Hamming distance, in which every token mismatch across sequences has equal weight (see Huang (1998) and Guha *et al.* (2000)). Le and Ho (2005) are a notable exception, basing the distance between tokens on the distributional properties of their neighbouring tokens; it is unclear whether this approach can be scaled up to larger token spaces or longer sequences or whether covariates could be included. It is clear that optimal matching, although quite effective at generating distances and clusters, does not incorporate a direct way to capture covariate effects, in part because no single 'outcome' is being modelled. We have no real sense of the stochastic process that gave rise to the data.

Hidden Markov models (Juang and Rabiner, 1991) and mixed Markov models (van de Pol and Langeheine, 1990) offer useful alternatives to classical Markov chain models. Both allow for a form of dependence that is plausible without an explosion in the number of additional parameters. Hidden Markov models use a hidden state to track this dependence, and observations are conditionally multinomial, given that state, whereas mixed Markov models use a latent state that is akin to a hidden cluster label. Conditional on the latent state (cluster), observations follow a simple, usually first-order, Markov process. This process may differ from latent state to latent state, providing ample diversity in outcomes. Both of these approaches, along with extensions of generalized linear models in the longitudinal setting (see Molenberghs and Verbeke (2005)), are likely to face challenges when the number of outcomes is large. Hidden Markov models are the only set of models that are employed in the context of large state spaces, but the speech and genomic data that have been used with these have the necessarily higher observation-to-parameter ratio, which is less common in the longitudinal setting.

Traditional regression methods are likely to lack the complexity that is required to capture the dynamics of the labour market, as we have operationalized them. Recall that one question that we may wish to evaluate is whether women have shorter or longer job tenures (here, defined as the time spent continuously in a single type of job) net of any occupational (and industrial) sorting. It is one thing for a data analysis to show that women have longer job tenures than men, and quite another to learn that women and men *doing the same type of work and with the same education* have different tenures. We might further wish to know whether the movement through the labour market is similar for men and women (in terms of what types of job typically follow one another). We might try to evaluate some of these questions by using a Cox proportional hazards survival model (Cox, 1972) and its extensions, in particular frailty models that capture heterogeneity (Aalen, 1988). To net out job characteristics, we must incorporate as many features of the jobs as possible. Indicators for the origin $I \times O$ could be included in such a model, so that starting on an $I \times O$ that happened to be typically of long duration is 'netted out' (captured by the model). Ignoring for the moment the large parameter space that is implied by this approach, what if the durations vary dramatically depending on the *destination* state? This could be modelled by using a competing risks framework (Crowder, 2001; Scott and Kennedy, 2005) in which risk of changing $I \times O$ varies for each destination state. With a little over 100 states, this requires about 100×100 parameters in the model. We essentially do not have enough information in our panel data to estimate so many parameters by using classical approaches.

A contribution of this paper is to implement a version of this type of competing risks model that treats the relationships between source and destination $I \times O$s as random effects. This draws on methods that were introduced in Hu *et al.* (2009), who took a Bayesian approach to competing risks to handle patterned missingness. In contrast, we use random effects and Bayesian methods to handle the large number of origin- and destination-specific differences. Dahinden *et al.* (2007) took a penalized likelihood approach to the sparsity problem that arises when there are so many nominal states. Our approach combines these ideas to make possible a large state space and incorporates a type of frailty as well. Given the above classical methods and extensions, our method can be characterized as a Bayesian extension of competing risks event history modelling with frailty. These are essentially continuous time Markov process models (with frailty), which have been in existence for a considerable time, but only with the advent of modern computing power, a random-effects formulation and Markov chain Monte Carlo (MCMC) techniques has it been possible to consider, as we do, applications with large state spaces.

The underlying transition affinities in our labour market data may change over time, and thus a non-time-homogeneous continuous time Markov process may be more realistic. We describe one approach to capturing this time inhomogeneity in Section 4.

3. Modelling framework

Sequence data consist of a series of origin states, durations in those states and transitions into new states. These data may be highly discrete—e.g. annual career data—or nearly continuous (as in weekly data). The underlying process is nearly continuous, so we build our model for duration times rather than for single-step transitions. We conceptualize the duration process as one of competing risks, in which there is an underlying hazard function that is associated with every origin–destination state combination. We observe only the first (shortest) realization of these competing duration processes (see Andersen *et al.* (1993) for further discussion). Under the assumption that durations are exponentially distributed, this is a continuous time Markov process.

We now describe a continuous time Markov process in terms of its latent competing duration times. We assume that the rate parameter for the duration process is given by $\lambda_{kq}(t)$, which is the rate from origin k to destination q at time t. For each origin state k, there are n_k records; i is used as an index for these records, from 1 to n_k. Note that we have ignored subject identifiers in this formulation; individual-specific effects will be included shortly. The total number of observations in the data set is $n = \Sigma_k n_k$. The observations may be organized into an $n \times 4$ matrix Y;

each row is a vector $(k, q_{ki}, t_{ki}, u_{ki})$. Here, q_{ki} is the destination token that is associated with the ith record with origin k; we set it to 0 if censoring occurs. t_{ki} is the time that this spell began, and u_{ki} is the spell duration (or time of censoring). We order the data so that the first d_k entries in each origin-specific group are uncensored.

We shall write $\lambda_{kq_{ki}}$ for $\lambda_{kq_{ki}}(t_{ki})$ and note that there is a λ corresponding to every origin state k, time t_{ki} and destination q_{ki}. The likelihood for the records that are associated with origin state k (all assumed independent) is

$$\mathcal{L}_k = \prod_{i=1}^{d_k} \left\{ \lambda_{kq_{ki}} \exp(-\lambda_{kq_{ki}} u_{ki}) \exp\left(-\sum_{j \neq q_{ki}} \lambda_{kj} u_{ki} \right) \right\} \prod_{i=d_k+1}^{n_k} \exp\left(-\sum_j \lambda_{kj} u_{ki} \right), \tag{1}$$

where it is understood that $\lambda_{kk} = 0$ or is just left out of any summation. The first product term captures the probability that a transition to q_{ki} occurs with duration u_{ki} and that all other latent durations are longer. We shall combine the components of this term before taking logarithms, but we leave them in this form for expository purposes as they represent the observed transitions. The second product term captures the censored observations: we use the time of censoring and compute the probability that all latent durations are longer than this. We note that censoring may be defined as when the observation is truncated because of the survey instrument, or when we lose information about the token state. In the latter case, we may make a transition to a state in which we know that the subject is working, but not which type of work, so a form of censoring occurs.

All the exponential terms in likelihood (1) combine readily, so, when we take logarithms and iterate over all observations, we have the overall log-likelihood

$$l(\Theta|Y) = \sum_k \left\{ \sum_{i=1}^{d_k} \log(\lambda_{kq_{ki}}) - \sum_{i=1}^{n_k} \lambda_k u_{ki} \right\}, \tag{2}$$

where $\lambda_k = \Sigma_j \lambda_{kj}$, and Θ is a vector containing all model parameters. Any within-subject correlation is handled by adding another component to the observation vector and including individual-specific effects in the model. This variant of a frailty model would associate effect α_r with subject $r = 1, \ldots, R$, assuming that these come from some mean 0 distribution known up to a scaling parameter. A simple modification of the rate parameter reflects this effect. Set $\lambda_{kq_{ki}r} = \lambda_{kq_{ki}} \exp(\alpha_r)$ and augment the rows of Y to include the subject identifier, so that it becomes $(k, q_{ki}, t_{ki}, u_{ki}, r)$. Observations are assumed conditionally independent, given α_r.

4. Affinity models

4.1. Model formulation

In the previous discussion, the latent transition rates $\lambda_{kq}(t)$ were assumed given, but we now build a log-linear model for them. The rates govern the relationship between an origin and destination token, and as such we include main effects for the origin and destination, along with effects for the interaction between the two. The latter can be understood as the affinity between the origin and destination token, net of everything else that is modelled. We include individual-specific and covariate effects in the model as well.

The latent transition rate from origin k to potential destination q for subject r at a single point in time is governed by

$$\log(\lambda_{kqr}) = \nu_k + \eta_q + d_{kq} + X_r\beta + \alpha_r, \qquad \text{for } k \neq q, \tag{3}$$

where ν_k is a (main) source effect, η_q is a (main) destination effect, d_{kq} their affinity (interaction) and $k, q \in \{1, \ldots, K\}$. P-dimensional column vector β captures subject level covariate effects that are represented by row vector X_r and α_r is a subject-specific effect. Job characteristics can be modelled with the addition of terms that are indexed by $I \times O$ rather than subject.

The rates may not be static. Overall time trends can be captured by including time in the covariate design matrix X_r. A more elaborate framework for handling change over time is to index d_{kq} as

$$d_{kqt} = \delta_{0kq} + \delta_{1kq} t, \tag{4}$$

so the (net) affinity between k and q at time t drifts linearly with time (over the observed time horizon; other functional forms are certainly possible). This simple linear component (which is unconstrained) doubles the number of affinity parameters. Alternatively, only K additional effects are needed if δ_{1kq} does not depend on q, and a similar reduction occurs when it does not depend on k (these are source- and destination-specific drifts respectively). Expression (4) can be expanded in a regression-like fashion to include additional covariates or their interactions.

4.2. Random effects

The information that we have on the transitions between different types of job, or tokens, varies greatly and is a function of the labour market itself. We represent these as affinity terms δ_{0kq} and δ_{1kq} specified in expression (4) and, to the extent that they can be determined, we wish to control for them in our models. Modelling them as random effects allows us to include them as controls and to estimate their range of impact accurately through variance components. Moreover, we can estimate specific affinities and their corresponding precision under a likelihood framework. Formally, we assume, as is common in random-effects modelling, that these affinities are drawn from a normal distribution: $\delta_{0kq} \sim N(0, \sigma^2_{\delta_0})$ and $\delta_{1kq} \sim N(0, \sigma^2_{\delta_1})$. Subject-specific effects are similarly treated as random: $\alpha_r \sim N(0, \sigma^2_{\alpha})$. The remaining parameters in the model are treated as 'fixed effects'.

4.3. Estimation and inference

Models such as those described above may be fitted by using a variety of techniques. We chose to use a Bayesian hierarchical modelling framework for estimation and inference. Our prior distributions are used primarily to establish a reasonable range for the parameters, whereas the hierarchical structure reflects our varying degree of uncertainty about them.

For our baseline model, we estimate P covariate effects and the two sets of K main effects ν_k and η_q by using reasonably flat priors, interpreting these as 'fixed' effects. The R individual subject effects are modelled as random effects as described above, and the variance component governing these effects is also modelled. We shall then estimate two additional models by adding increasingly complex affinity terms, adding $K(K-1)$ parameters that govern the affinities and another $K(K-1)$ to capture their drift over time. In what follows, we describe the affinity drift model as it is the largest model fit.

The $2K$ main effects ν_k and η_q, as well as covariate effects β, are assumed to be independent identically distributed normal, mean 0 and variance σ^2_g; we set the variance sufficiently large that the prior is approximately locally uniform. As above, $\alpha_r \sim N(0, \sigma^2_{\alpha})$, but now σ^2_{α} is modelled as a scaled Inv-χ^2 distribution with scale $\sigma^2_{\alpha_0}$ and ν_{α} degrees of freedom. Baseline affinities and drifts δ_{0kq} and δ_{1kq} are as above, but now variances $\sigma^2_{\delta_0}$ and $\sigma^2_{\delta_1}$ are modelled as scaled Inv-χ^2 distributions with scales $\sigma^2_{\delta_{00}}$ and $\sigma^2_{\delta_{10}}$, and ν_{δ_0} and ν_{δ_1} degrees of freedom respectively. The values of the hyperparameters $(\sigma^2_g, \sigma^2_{\alpha_0}, \nu_{\alpha}, \sigma^2_{\delta_{00}}, \sigma^2_{\delta_{10}}, \nu_{\delta_0}, \nu_{\delta_1})$ are set by the researcher and represent our prior beliefs about the variance of the individual effects and the affinity components.

Summarizing the above, we have

$$
\begin{aligned}
\nu_k &\sim N(0,\sigma_g^2), && k=1,\dots,K,\\
\eta_q &\sim N(0,\sigma_g^2), && q=1,\dots,K,\\
\beta_p &\sim N(0,\sigma_g^2), && p=1,\dots,P,\\
\alpha_r &\sim N(0,\sigma_\alpha^2), && r=1,\dots,R,
\end{aligned}
$$

$$\sigma_\alpha^2 \sim \sigma_{\alpha_0}^2 \times \text{Inv-}\chi^2_{\nu_\alpha},$$

$$
\begin{aligned}
\delta_{0kq} &\sim N(0,\sigma_{\delta_0}^2), && k=1,\dots,K, \quad q=1,\dots,K,\\
\delta_{1kq} &\sim N(0,\sigma_{\delta_1}^2), && k=1,\dots,K, \quad q=1,\dots,K,
\end{aligned}
$$

$$
\begin{aligned}
\sigma_{\delta_0}^2 &\sim \sigma_{\delta_{00}}^2 \times \text{Inv-}\chi^2_{\nu_{\delta_0}},\\
\sigma_{\delta_1}^2 &\sim \sigma_{\delta_{10}}^2 \times \text{Inv-}\chi^2_{\nu_{\delta_1}}.
\end{aligned}
$$

Transitions from a state to itself are not allowed, so all values of δ_{0kq} and δ_{1kq} for which $k=q$ are fixed at 0. After some minor exploration, we set $\sigma_g^2=25$ to allow the main effects to be effectively unconstrained and set $\sigma_{\alpha_0}^2=0.75$ and $\nu_\alpha=2$ to allow a reasonably wide range of variances for the random effects (80% of the prior falls within the interval (0,3) whereas 95% falls within (0,15); the posterior variances that we have obtained tend to lie closer to 0 than to the prior mean). For parameters such as these, the estimation and inference are reasonably robust to choices of hyperparameter. For the affinity parameters, we set $\sigma_{\delta_{00}}^2=\sigma_{\delta_{10}}^2=0.5$ and $\nu_{\delta_0}=\nu_{\delta_1}=5$ (again, these primarily set a reasonable range).

We fit the model by using an MCMC algorithm, iterating over the parameters above in Metropolis-within-Gibbs fashion. In other words, we sample from the conditional distributions of each set of parameters, above, sequentially, given all others. We use a random-walk proposal and allow the user to specify a tuning period and target acceptance rate. During the initial burn-in phase, our algorithm periodically adjusts the scale in the jumping rule to attempt to match the target rate. Given the high dimensionality of the parameter space, the MCMC algorithm was implemented in C with an R interface. This code and a test data set are available on the author's Web page `http://homepages.nyu.edu/~ms184/Software/affinity`.

For each iteration, we first compute $\lambda|\nu,\eta,\beta,\alpha,d$, by using expressions (3) and (4), for every spell associated with every subject. Subscripts have been dropped, so the symbols represent the full vector of parameters. In the model with drift, time t is the start of the interval for the transition d_{kqt} from k to q. We might have averaged the affinity over spell duration, but this approach simplifies computation and is a reasonable first approximation. In our data, time is the quarter spanning from age 20 to 36 years (numbered $0,1,\dots,63$), so a 28-year-old has $t=32$. Set $\Theta=(\nu,\eta,\beta',\alpha,\sigma_\alpha^2,\delta_0,\delta_1,\sigma_{\delta_0}^2,\sigma_{\delta_1}^2)$.

The conditional posteriors for the most part have one of two forms:

$$\nu|\text{others} \propto \varphi_K(\nu;0,\sigma_g^2 I_K)\mathcal{L}(\Theta|Y), \tag{5}$$

$$\eta|\text{others} \propto \varphi_K(\eta;0,\sigma_g^2 I_K)\mathcal{L}(\Theta|Y), \tag{6}$$

$$\beta|\text{others} \propto \varphi_P(\beta;0,\sigma_g^2 I_P)\mathcal{L}(\Theta|Y), \tag{7}$$

$$\alpha|\text{others} \propto \varphi_R(\alpha;0,\sigma_\alpha^2 I_R)\mathcal{L}(\Theta|Y), \tag{8}$$

$$\sigma_\alpha^2|\text{others} \sim (\nu_\alpha\sigma_{\alpha_0}^2 + Rs_\alpha^2) \times \text{Inv-}\chi^2_{\nu_\alpha+R}, \tag{9}$$

$$\delta_0|\text{others} \propto \varphi_{K(K-1)}(\delta_0; 0, \sigma^2_{\delta_0} I_{K(K-1)}) \mathcal{L}(\Theta|Y), \tag{10}$$

$$\delta_1|\text{others} \propto \varphi_{K(K-1)}(\delta_1; 0, \sigma^2_{\delta_1} I_{K(K-1)}) \mathcal{L}(\Theta|Y), \tag{11}$$

$$\sigma^2_{\delta_0}|\text{others} \sim \{\nu_{\delta_0}\sigma^2_{\delta_{00}} + K(K-1)s^2_{\delta_0}\} \times \text{Inv-}\chi^2_{\nu_{\delta_0}+K(K-1)}, \tag{12}$$

$$\sigma^2_{\delta_1}|\text{others} \sim \{\nu_{\delta_1}\sigma^2_{\delta_{10}} + K(K-1)s^2_{\delta_1}\} \times \text{Inv-}\chi^2_{\nu_{\delta_1}+K(K-1)}, \tag{13}$$

where φ_N is the N-dimensional normal density and

$$s^2_\alpha = \frac{1}{R}\sum_r \alpha^2_r,$$
$$s^2_{\delta_0} = \frac{1}{K(K-1)}\sum_{k,q} \delta^2_{0kq},$$
$$s^2_{\delta_1} = \frac{1}{K(K-1)}\sum_{k,q} \delta^2_{1kq},$$

and $\mathcal{L}(\Theta|Y)$ is the unlogged likelihood, based on equations (2)–(4). Note throughout that we list the dimension of affinity parameters as $K(K-1)$, because in practice K terms are identically set to 0.

In our implementation, we use a symmetric proposal, and the algorithm proceeds as follows (note that i is used here to reflect iteration step).

Step 1: sample ν_{i+1}.

(a) Propose $\nu_{i+1} = \nu_i + \tau_\nu z_\nu$, where $z_\nu \sim N(0, I_K)$ and τ_ν is a column vector of K scaling factors controlling the size of proposed moves.

(b) Accept the proposal with probability

$$\frac{\mathcal{L}(\nu_{i+1}, \Theta^*_i|Y)\, \varphi_K(\nu_{i+1}; 0, \sigma^2_g I_K)}{\mathcal{L}(\Theta_i|Y)\, \varphi_K(\nu_i; 0, \sigma^2_g I_K)}, \tag{14}$$

where it is understood that $\Theta^*_i = \Theta_i \backslash \nu_i$, in this instance.

Step 2: sample η_{i+1}—the proposals parallel those of ν, with a similar updating rule. The range of the jumping rule is again tuned with parameter-specific scaling factors.

Step 3: sample $\delta_{0,i+1}$, $\delta_{1,i+1}$, $\sigma^2_{\delta_0,i+1}$ and $\sigma^2_{\delta_1,i+1}$.

(a) Propose $\delta_{0,i+1} = \delta_{0,i} + \tau_{\delta_0} z_{\delta_0}$, and $\delta_{1,i+1} = \delta_{1,i} + \tau_{\delta_1} z_{\delta_1}$, where $z_{\delta_0} \sim N(0, I_{K(K-1)})$ and $z_{\delta_1} \sim N(0, I_{K(K-1)})$ and τ_{δ_0} and τ_{δ_1} are each scalars controlling the size of proposed moves. Thus, a single tuning parameter is used for each type of affinity (this decision was based on the sheer number of these effects).

(b) Accept proposals and update Θ_i one component at a time. The $2K(K-1)$ components can be updated in a different order, each iteration, by preshuffling the index. The (k, q)th component δ_{0kq} is updated with probability

$$\frac{\mathcal{L}(\delta_{0kq,i+1}, \Theta^*_i|Y)\, \varphi(\delta_{0kq,i+1}; 0, \sigma^2_{\delta_0})}{\mathcal{L}(\Theta_i|Y)\, \varphi(\delta_{0kq,i}; 0, \sigma^2_{\delta_0})}, \tag{15}$$

where it is understood that $\Theta^*_i = \Theta_i \backslash \delta_{0kq,i}$. The updating rule for δ_{1kq} is analogous.

(c) Sample $\sigma^2_{\delta_0,t+1}$ and $\sigma^2_{\delta_1,t+1}$: the posterior is given by distributions (12) and (13) in closed form and may be sampled exactly.

Step 4: sample β_{i+1}—propose $\beta_{i+1}=\beta_i+\tau_\beta z_\beta$, where $z_\beta \sim N(0, I_P)$ and τ_β is a column vector of scaling factors controlling the size of P proposed moves. The remaining implementation mirrors that of η or ν.
Step 5: sample α_{t+1}—the proposals mirror those of the components of β, with R tuning parameters, one for each α_r.
Step 6: sample $\sigma^2_{\alpha;t+1}$—the posterior is given by distribution (9) in closed form and may be sampled exactly.

We implemented a tuning phase that adjusts the scale of the proposals $(\tau_\nu, \tau_\eta, \ldots, \tau_{\delta_0}, \tau_{\delta_1})$ to aim for a 10–20% acceptance rate, in accordance with the dimensionality of the problem (Gelman *et al.*, 2004). We run a minimum of 20000 post-tuning iterations, checking for convergence of all parameters by using Gelman and Rubin's between–within R criterion BWR (Gelman and Rubin, 1992) and adding additional blocks of 10000 iterations if warranted. All models satisfied the convergence criterion BWR <1.2, suggesting adequate mixing of the chain. We thin the posterior samples somewhat to compensate for any less-well-tuned parameter.

4.4. Model selection

Two aspects of model selection have a bearing on the substance of our problem: the 'need' for affinities or for affinity drift parameters and, given their potential difference, whether or not male and female labour markets may be pooled. The latter involves testing a large interaction at the level of random effects. With the former, we included these terms because we thought that pairwise relationships would reveal important information, but we would like to evaluate this formally. For model comparison and selection, we shall use the deviance information criterion DIC (Spiegelhalter *et al.*, 2002), which is useful for evaluating Bayesian models with random effects. In particular, it penalizes the posterior deviance by using an effective number of parameters that may be substantially smaller than the number of random effects in the model. The effective number of parameters is calculated as

$$p_D=\overline{D(\Theta)}-D(\bar{\Theta}) \tag{16}$$

where $D(\Theta)=-2l(\Theta|Y)$. $\overline{D(\Theta)}$ is the posterior expectation of the deviance and $D(\bar{\Theta})$ is the deviance evaluated at the expected value of the parameter vector. These quantities are all derived from our MCMC samples. Then DIC is given by

$$\text{DIC}=\overline{D(\Theta)}+p_D=D(\bar{\Theta})+2p_D. \tag{17}$$

Comparison of DIC and p_D for several models will guide our understanding of the process under investigation and the parsimony of our models. Given that these values are estimated from MCMC samples, we require some sense of their variability before making comparisons across models. By running multiple chains, we can compute the magnitude of the difference in DICs across different MCMC *samples*. Approximate confidence bounds for a specific DIC are based on twice this magnitude; non-overlapping DIC-intervals will be taken as evidence of one model over another. Using this criterion, we can be fairly sure that differences are not due to sampling variation in DIC.

4.5. Model summary

In terms of our application, the two aspects of the model that are most important are covariate effects and the affinities. The former can be summarized by posterior distributions or functions thereof, whereas the latter is best visualized in some manner. Matrices can be thought of as a spatial map, where row and column numbers specify co-ordinates. As such, each small region (reflecting the relationship between two types of job) that is defined by this map can be shaded on the basis of its magnitude. The ordering of the rows and columns is essentially arbitrary in our labour market data, which is likely to obscure relationships in the data. Since one of the interesting questions that we pose about these data is whether affinities (and thus mobility) are organized more by industry or by occupation, we order the rows and columns first by industry code and then by occupation code. This is a specific *seriation*, or ordering, of the $I \times O$s that will reveal the presence or absence of structure through the extent to which the map looks roughly block diagonal. Seriation is often approached as an optimization problem (see Gale *et al.* (1984), Sarkar (2008) and Hahsler *et al.* (2008) for a discussion).

5. Findings

Our source of data is the National Longitudinal Survey of Youth. This survey is a large representative national sample of more than 12 000 young men and women in the USA, with oversampling among populations of colour and the economically disadvantaged. They were aged 14–21 years when first surveyed in 1979 and continue to be surveyed into their 40s, with the most recently available year being 2008. Sampling weights are used to adjust for the oversampling. With the majority of lifetime wage growth and job changes occurring in the first decade of work experience (Topel and Ward, 1992), they provide key information on career formation at a crucial point in the life course. The oldest workers in our data are aged 42 years in 2000—we do not include future years in this study, as we limit the sequence analysis to workers aged 20–36 years.

Recall that we consider two different factors in our analysis: gender and education. Through this lens, we shall evaluate whether the careers of men and women follow statistically similar pathways, i.e. whether the affinities are about the same, regardless of gender, controlling for education. If they are, a pooled affinity model can be used to evaluate remaining covariate effects. If the affinities are different, we shall determine the particular $I \times O$s that characterize gender differences and what these imply about the labour market.

To evaluate potential gender differences, we fit a more complex model than described above. Namely, we fit a fully interacted model, in which parameters have gender-specific components. For example, we include the vectors ν and ν_s and use coding -1 for males and 1 for females in the model, so that $\nu - \nu_s$, $\nu + \nu_s$ and ν are the vectors of source effects for males, females and their average respectively. All parameters are represented in terms of this coding. Broadly speaking, then, we compare two competing models with parameterizations

$$\Theta_{\text{pool}} = (\nu, \eta, \beta', \alpha, \sigma_\alpha, \delta_0, \delta_1, \sigma_{\delta_0}, \sigma_{\delta_1}),$$
$$\Theta_{\text{inter}} = (\nu, \nu_s, \eta, \eta_s, \beta', \beta'_s, \alpha, \sigma_\alpha, \sigma_{\alpha_s}, \delta_0, \delta_{0s}, \delta_1, \delta_{1s}, \sigma_{\delta_0}, \sigma_{\delta_{0s}}, \sigma_{\delta_1}, \sigma_{\delta_{1s}}),$$

where the absence of a subscript denotes the average that is associated with the $(-1, 1)$ coded sex effect for a model parameter, and subscript 's' denotes the differential. Variance components are expressed in terms of standard deviations to maintain additivity in the differential coding. This coding scheme allows us to compare the parameters in pooled models with average effects in the interaction models. We use DIC to assess the fit of pooled and interaction models.

Rather than introduce a new set of priors for the differentials, we note that these differential codings correspond to sex-specific levels for all the parameters. Priors for these can be based on priors for the pooled model, and they can be adjusted for either sex. Practically speaking, this is only important for the variance components, as the other priors are effectively flat. If we believe, for example, that females have greater variance in individual effects α, we could inflate the level of σ_{α_0} for females. In the software, we implement these as proportionate differences; in our National Longitudinal Survey of Youth example, we use equivalent priors for males and females, reflecting our expectation of homogeneity.

We also contrast workers who have completed more and less education by age 24 years. Those who have at least a 2-year degree form one group and those with a high school degree or less form the other. We do not pool across education groups for substantive reasons. Age 24 years was chosen because most traditionally schooled workers complete their schooling by that age, and it is sufficiently close to when we start the career clock to provide a meaningful partition. We track whether additional schooling is obtained through a time-dependent covariate. Specifically, we include indicators for when high school, HS, or 2-year, AA, or 4-year, BA, college is completed for the less educated workers and 4-year college, BA, or a Master's degree, MA, is completed for more educated workers.

Procedurally, we first stratify by education group; within each group, we compare two sets of models, pooled and interaction with respect to gender; within each of these specifications, we examine three sets of models, which capture increasing levels of complexity with respect to affinities. Thus, our baseline model includes source, destination and covariate effects ν, η and β', as well as individual-specific effects α. The next two affinity models include δ_0- and δ_1-terms and their corresponding variance components, sequentially. All models contain a fixed effect for gender, education and time, the last capturing whether on average job change rates are increasing or decreasing as workers age, net of all else. Theory would suggest that rates decrease as workers complete an early career job shopping process. The stylized fact that job changing will diminish over time may be driven by composition, in that workers may simply move to $I \times O$s with typically longer tenures as they age. We can net out this aspect of labour market sorting by using affinity models. Fully interacted models mirror these sequential model specifications. Note, however, given our model formulation and assumptions (source, destination, affinity and affinity drift effects are mean 0), the gender effect β_{sex} for interaction models represents the differential at time 0. Contrast this with the gender effect for pooled models, which represents the average differential across time. Thus, the interaction of gender with time and individual differences in transition times suggests that the β_{sex}-effects for interaction and pooled models are not directly comparable. Note as well that the constant is suppressed in our table.

The posterior means for fixed effects and variance components along with DIC for the gender-specific models are given in Table 1. Time has been rescaled to reflect annual rather than quarterly changes. A 95% quantile-based confidence bound for the posterior of each parameter is used to assess significance in Table 1; intervals that do not cross zero are indicated.

Examining the pooled, high school or lower education models first, we have the interesting finding that the only significant covariate effect is time. The effects are in terms of log-rates, so we can exponentiate them to obtain an approximate percentage change in rate. In the baseline and affinity models the time effect is positive, suggesting about a 3% increase per year; workers change $I \times O$s more often as they age, net of all else. This is unexpected, but it can be understood as time's effect after netting out certain compositional changes in the type of job that is held at different points in the career. In contrast, in the affinity drift model, the effect changes sign and increases in magnitude, to about -10%. This model allows $I \times O$ relationships to 'reorganize' over time, so that the effects of compositional differences (types of job held) change over time. By modelling the effect of compositional differences in this dynamic fashion, we return to the stylized fact that job change rates are decreasing with age, but now we have the stronger finding that this holds under more nuanced controls for types of job. We note that DIC would

Table 1. Fixed effects and variance components (posterior means) and DIC for pooled and gender interaction models, baseline and affinity models, by education†

		β_{sex}	β_{HS}	β_{AA}	β_t	σ_α	σ_{δ_0}	σ_{δ_1}	*DIC*
Initial education: high school or lower—pooled models									
Baseline		0.00	−0.01	0.02	0.03†	0.22	—	—	418247
Affinities		0.00	−0.01	0.02	0.03†	0.22	1.96	—	370239
Affinities with drift		0.00	−0.01	0.00	−0.10†	0.22	1.97	0.24	366208
Initial education: high school or lower—interaction models									
Baseline	Average		0.01	0.00	0.03†	0.21	—	—	401611
	Differential	−0.03	−0.03†	0.00	0.01†	0.00	—	—	
Affinities	Average		0.01	0.00	0.03†	0.21	2.09	—	355521
	Differential	−0.18†	−0.03†	0.00	0.01†	0.00	0.03	—	
Affinities with drift	Average		0.01	−0.04	−0.15†	0.21	2.12	0.28	350034
	Differential	−0.23†	−0.03†	−0.01	0.00	0.00	0.07†	0.01	
		β_{sex}	β_{BA}	β_{MA}	β_t	σ_α	σ_{δ_0}	σ_{δ_1}	*DIC*
Initial education: 2-year college or higher—pooled models									
Baseline		−0.01	−0.07†	0.03	0.02†	0.22	—	—	119993
Affinities		−0.01	−0.07†	0.03	0.02†	0.21	2.15	—	101893
Affinities with drift		−0.01	0.04	−0.02	−0.20†	0.23	2.16	0.31	99428
Initial education: 2-year college or higher—interaction models									
Baseline	Average		−0.09†	0.05	0.02†	0.24	—	—	122858
	Differential	−0.55†	0.00	−0.03	0.01†	0.02	—	—	
Affinities	Average		−0.09†	0.05	0.02†	0.23	2.19	—	103910
	Differential	−0.59†	0.00	−0.03	0.01†	0.02†	0.00	—	
Affinities with drift	Average		0.05†	−0.02	−0.26†	0.25	2.20	0.36	100573
	Differential	−0.62†	−0.02	−0.01	0.03	0.02	0.00	−0.01	

†Interval does not cross zero.

recommend the affinity drift model over its competitors. Gender and education do not seem to matter for these models, but this may be an artefact of the pooling.

Moving down Table 1, the next set of models is based on a full interaction with gender. The average effects are given in the first row for each model, whereas the second row lists differential effects (i.e. those with '*s*'-indices). For the baseline model, the differential β_{sex} is not significant, but a gender-specific education effect emerges in these models, with high school completion associated with fewer transitions for females (3% lower than average) and thus more frequent transitions for men. The average effect of time is similar to that of the pooled models, but now a small but significant (1%) differential effect emerges, with transition rates for females increasing over time, effectively catching up to their high-school-educated, male counterparts. As we move to the affinity model, differential β_{sex} is large and significantly negative, indicating that females are expected to make fewer transitions at the start of their careers. The other differentials are identical to those in the baseline model. In the affinity drift interaction model, as in the corresponding pooled model, the fixed effect for time is large and significantly negative (here −15%), but the differential has become non-significant. The large, negative sex differential (which is associated with career start) still holds at −23%. The fact that pooled models did not obtain a significant sex differential under any model specification appears to be an artefact of the pooling, but we remind the reader that this effect is not directly comparable with that from the interaction model. DIC again recommends the affinity drift model, which is perhaps more striking in these interaction models, with their larger parameter space.

Variance components for these models describe the amount of variation that is between subjects, which is captured in the σ_α-term, as well as that between $I \times O$s, which is captured in the σ_{δ_0}-term. The affinity drift variance component σ_{δ_1} captures the between-$I \times O$ differences that emerge over time. Average and differential effects are presented for these parameters, and significance of the differentials is indicated. Given a rescaling of time t to an annual metric, we can describe the average variance that is associated with $I \times O$s as a function of years past age 20 years: $\sigma^2_{\delta_0} + t^2\sigma^2_{\delta_1}$. Notably, we find no significant differences between average variance components as we increase model complexity via affinity parameters. However, we do find one significant differential: in the affinity drift model, females apparently have somewhat greater heterogeneity in their δ_0-terms. Differences in the variance components or lack thereof reflect heterogeneity of effects, so similar heterogeneity across sex does not imply that each affinity is the same. Differences in affinities across sex are explored in greater detail in the visualization to follow.

The magnitude of individual differences is relatively small: the posterior mean for σ_α is 0.21 or 0.22 in all the lower education models. Affinities vary greatly across $I \times O$s, with average σ_{δ_0} estimated to be close to 2 for these models. The magnitude of this variance component translates into starkly different transition rates—some $I \times O$s simply do not make a transition from one to another, whereas others are strongly related to one another. The affinity drift variation σ_{δ_1} has apparently low average levels between 0.24 and 0.36 in these models, but the corresponding terms become increasingly meaningful over time. In the affinity drift interaction model, by age 30 years, with elapsed $t = 10$, $t\sigma_{\delta_1}$ approaches 3 for females, which exceeds the variation in initial differences, σ_{δ_0}. Affinities can shrink to zero or double in magnitude, over time.

For more educated workers, the pooled models initially tell a different story with respect to the effect of additional education. Here, workers are receiving Bachelor and Masters degrees, and the former is associated with a decreased rate of change (a reduction of 7% in baseline and affinity models). However, this effect loses significance in the affinity drift model, which is recommended by DIC. This suggests that, for college graduates, decreasing rates of change are driven by movement to types of job with longer tenures. The only other significant effect is for time, which mirrors the findings for lower educated workers, namely small but significant increases in rates of change for baseline models are reversed to substantial negative effects (-20%) in the affinity drift models. The variance components in this model are of the same magnitude and can be interpreted in the same manner as for the lower educated workers.

Turning to the interaction models, we find larger and more consistent differentials for β_{sex}, with values hovering around -0.60. Again, this reflects gender differences at the start of the career. In the simpler models, there are significant sex differentials in the effect of time, but these fade in the affinity drift model. The overarching, common finding of small but significant positive effects for time being reversed to large, significantly negative effects in the affinity drift model holds in these models as well. The only significant variance component differential is for σ_{α_s}, in the affinity model, but it is not particularly large.

We now evaluate the need for pooled or interaction models for more educated workers using DIC. Given that DICs are based on MCMC samples, we check the potential variation in DICs across samples; in all the models evaluated, twice the absolute difference of a DIC *across different simulations* (for the same model) never exceeded 125. The DIC-differences *across models* (in the model selection process) were always in the thousands, so our DIC-differences are unlikely to be due to sampling variation. We have that, within any education group and type of model, DIC always selects models with affinities and drift over time. We also find that pooled models are recommended by DIC for higher educated workers, whereas interaction models are recommended for lower educated workers. This is consistent regardless of model complexity.

This suggests that the labour market dynamics—how workers move through the market over time—are quite similar, given comparable initial conditions, for graduates of 2-year colleges and higher, regardless of gender. Although we did find some interesting differentials in the interaction model, the additional parameters that were required to identify them were not justified under the DIC-criterion.

We next visualize the 103×103 affinity matrices $\{\delta_0\}_{kq}$ and $\{\delta_1\}_{kq}$ for lower educated men and women, based on the affinity drift interaction model. Parameter δ_0 represents the affinities in the market at age 20 years, whereas δ_1 captures their change over time. Our gender encoding establishes an average and differential set of effects, and we again report posterior means. The extent to which these affinities are positive or negative and their patterning relative to industries and occupations will be the focus of this analysis. Since confidence bounds for the majority of affinities cross zero, we use blank space for all of these; since some affinities are positive and some are negative, we shade with black any negative affinity and use two shades of grey for positive effects, darker for higher levels. The cut points for the graduations were data driven, using the median of the positive significant effects to split them. We examine $I \times O$s sorted by industry code, and we briefly comment on those sorted by occupation code. The advantage to using a common seriation is that cross-comparisons can be made; block diagonal structure reflects the organizing principle that is used in the seriation.

In Fig. 1(a), we display the δ_0-affinities for lower educated workers and contrast these with sex differentials in Fig. 1(b). The abbreviations that are used for industries are given in alphabetical order in Table 2, which also lists their frequency (proportions of person–periods) across the

Table 2. Industry abbreviations and their relative frequency (across all workers and jobs) in the National Longitudinal Survey of Youth

Abbreviation	*Description*	%
agr&	Agriculture, forestry	3.5
autS	Automative and repair services	2.1
bldS	Building services	2.7
bzSv	High end business services	1.7
cnst	Construction	7.2
comm	Communications	1.3
eatD	Eating and drinking	5.7
educ	Education	6.1
FIRE	Finance, insurance, real estate	6.6
hlth	Healthcare	8.7
hotl	Hotels and laundry service	2.0
mnfD	Manufacturing, durable	11.1
mnfN	Manufacturing, non-durable	8.5
nonp	Non-profit	1.6
PAdm	Public administration	4.9
prfS	Professional services	2.7
prsl	Personal services	1.9
recr	Entertainment, recreation services	1.7
retA	Retail automotive	1.4
retF	Retail food	2.5
retH	Retail hard goods (except automotive)	7.9
tmps	Temporary work agencies	0.7
tran	Transportation	3.1
util	Utilities, sanitary services	1.4
whol	Wholesale trade	3.0

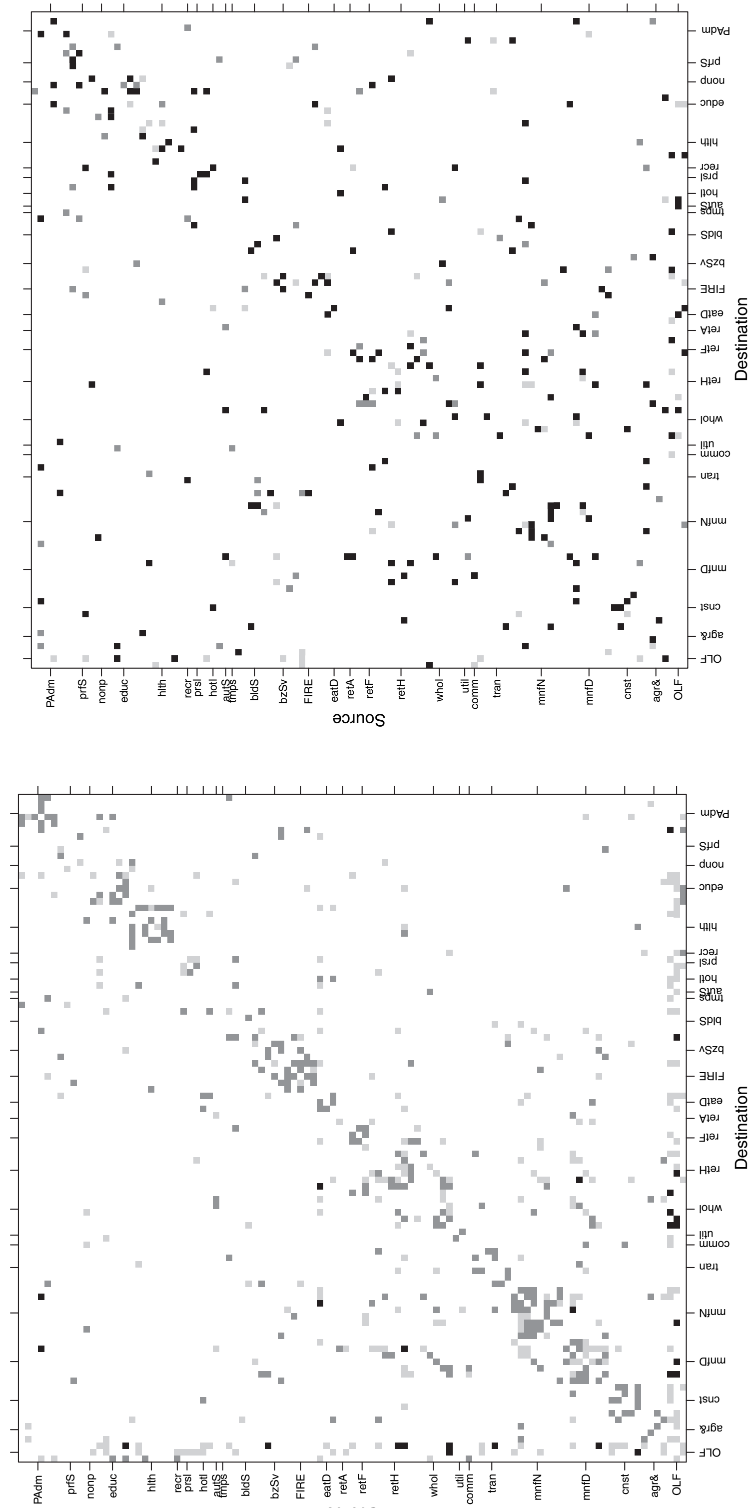

Fig. 1. Workers with no post-secondary degree, aged 24 years: (a) full model, average affinities; (b) full model, differential affinities

period examined. In Fig. 1(a), black cells are in the range $(-2.2, 0)$, light grey $(0, 2.6)$ and dark grey $(2.6, 7.1)$ (the median is slightly smaller, at 2.2, for gender-specific affinities). The $I \times O$s are seriated on the basis of the industry code number and occupation code within that; this ordering roughly corresponds to the first digit of a census-based classification, for which some allied industries (such as construction and manufacturing) tend to be given nearby codes. The first three rows and columns (counting from the bottom left-hand corner outwards) represent time out of the labour force: in school, not in school or unemployed. A strong block diagonal structure is apparent, suggesting that there are many industry-based affinities for both men and women. Note as well that the darker shading tends to reside on these block diagonal elements. This shading indicates a positive coefficient, so movement within an industry or related industries is more frequent, especially early in the career. The many significant effects in the first three rows and columns suggest that movement in and out of the labour force varies greatly by source or destination $I \times O$. Most negative effects emerge here as well. The differences between men and women are seen in Fig. 1(b) to consist of many negative effects near the diagonal. This suggests that some of the pathways within industry-based blocks are less common for women, or that they 'stay put' more often, at least early in their career.

In Fig. 2, we examine the δ_1-terms. Black shading still reflects negative (significant) values, but the range for both panels is now about $(-0.4, 0)$, whereas light grey represents about $(0, 0.2)$ and dark grey represents values that are a little larger than 0.2. In Fig. 2(a), the majority of significant effects are concentrated in the first three rows and columns of the affinity drift matrix, and all are positive, indicating that, for specific $I \times O$s, the rate of transition in and out of the labour force is increasing over time. Contrast this with the main, fixed effect for time in this model, which was significantly negative. We conclude that certain $I \times O$s are much more likely than others to be associated with such movement, and that this difference grows with time. In addition, most remaining significant effects are concentrated near the diagonal, indicating changes of occupation within allied industries, such as manufacturing (codes mnfD and mnfN located at the lower left-hand side). These transition rates are increasing with time, further delineating which types of transitions form the normative pathways through the labour market. Differentials that are depicted in Fig. 2(b) are even more concentrated in the first three rows and columns. The majority are negative, with the bulk on the rows corresponding to time spent out of the labour force, representing a decrease in the rate of transition to the column destinations over time. Females are staying out of the labour force more than men as they age. Conversely, the negative effects in the first three columns suggest that in specific industries, e.g. finance, insurance and real estate (which are denoted as 'FIRE' in Figs 1 and 2 and are traditionally a higher paying sector), females are less likely to leave the market as they age.

We examined equivalent representations of these affinities organized by occupation and did not find the same sort of patterning. This was true for both average and differential effects for δ_0 and δ_1.

As we have seen in Fig. 2, representing ties between $I \times O$-pairs through affinities not only controls for this aspect of labour market dynamics but also allows us to identify natural groupings, revealing certain features of how the labour market is organized. We can also isolate which career types reveal gender differences. In Table 3, we list the 10 largest affinities, first for the population as a whole, and then followed by the gender-specific effects. Affinities, rather than affinity drifts, were examined in this analysis. In aggregate, movement tends to be within industry, at least for eight of the 10 largest affinities, whereas movement is between personal service occupations in the remaining two. Healthcare and food service clusters emerge as important pathways, and the movement is from less to more skilled occupations, such as manager. The largest affinity is for the transition from protective work in public administration (police) to administrative work

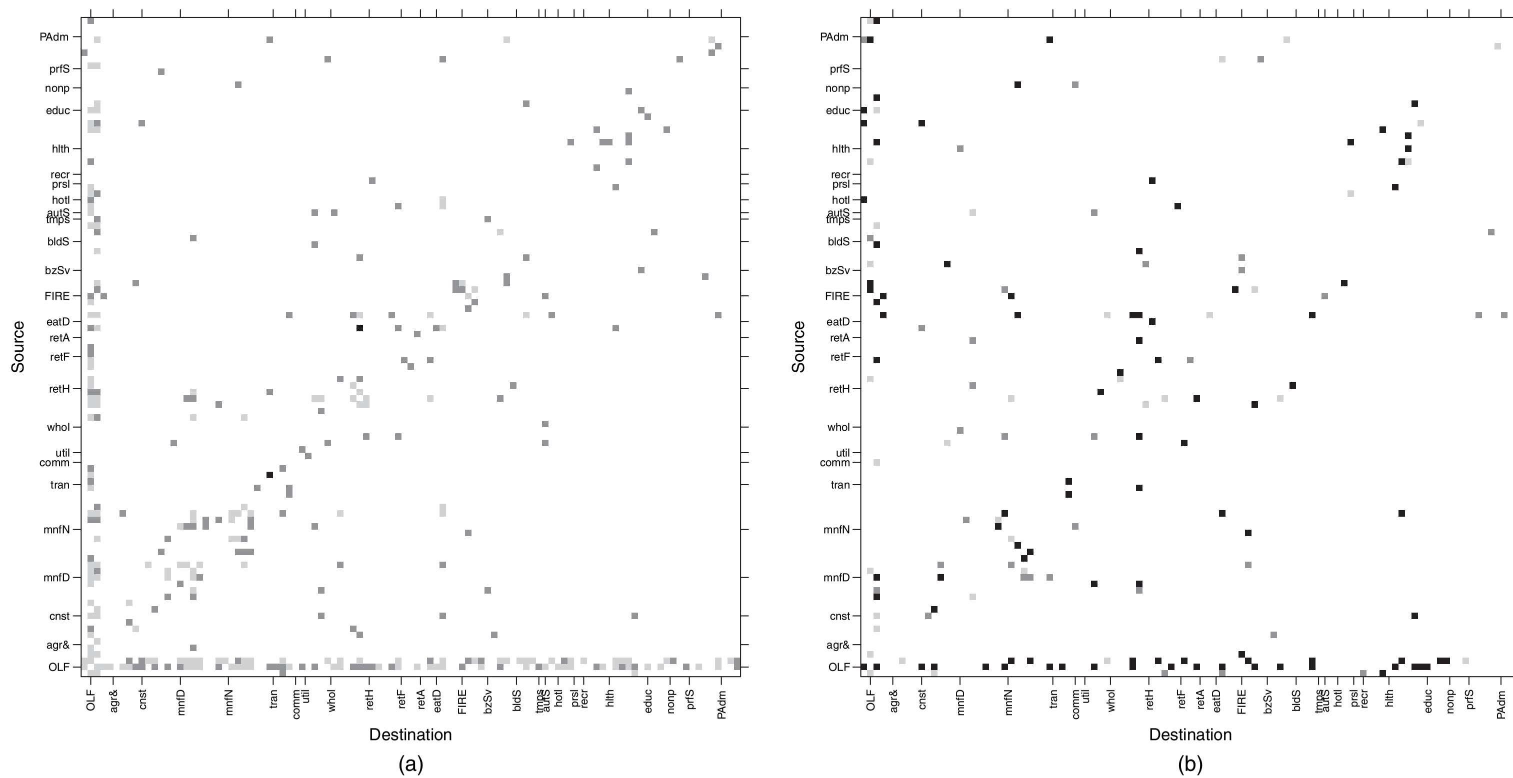

Fig. 2. Workers with no post-secondary degree, aged 24 years: (a) full model, average affinity drifts; (b) full model, differential affinity drifts

Table 3. 10 largest affinities: average, and differential in the interaction model for high school or lower initial education in the National Longitudinal Survey of Youth

Source I × O	*Destination I × O*	δ_0
Average affinities		
Public administration × Protective services	Public administration × Manager or administrator	7.08
Healthcare × Technician	Healthcare × Nurse, Dietician	7.08
Utilities, sanitary services × Clerical or unskilled	Utilities, Sanitation × Craftworker, mechanic	6.27
Hotels and laundry service × Personal service	Personal services × Personal service	6.24
Healthcare × Nurse, Dietician	Healthcare × Technician	6.19
Healthcare × Nurse, Dietician	Healthcare × Manager or administrator	6.08
Eating and drinking × Sales worker	Eating and drinking × Manager or administrator	6.05
Personal services × Personal service	Hotels and laundry service × Personal service	5.99
Retail food × Clerical or unskilled	Retail food × Sales worker	5.89
Retail food × Clerical or unskilled	Retail food × Labourer	5.82
Differential affinities		
Retail food × Labourer	Retail food × Clerical or unskilled	3.93
Retail food × Sales worker	Wholesale trade × Labourer	3.89
Wholesale trade × Labourer	Retail food × Sales worker	3.86
Manufacturing durable × Sales worker	Eating and drinking × Sales worker	3.50
Manufacturing durable × Craftworker, mechanic	Retail automotive × Sales worker	−3.35
Education × Personal service	Education × Manager or administrator	−3.41
Agriculture, forestry × Craftworker, mechanic	Construction × Manager or administrator	−3.51
Building services × Cleaning service	Hotels and laundry service × Cleaning service	−3.58
Manufacturing, durable × Labourer	Education × Clerical or unskilled	−3.68
Retail hard goods (non-automotive) × Labourer	Transportation × Labourer	−4.01

in the same industry. Gender differences play out through increased and decreased affinities. Movement to and from retail food and wholesale trade industries is more common for females. Movement between industries that are loosely aligned, such as manufacturing and automotive, is more common for males, as is movement between agricultural craftwork and the construction industry.

6. Discussion and future directions

These models are highly informative in terms of the substantive questions that we had hoped to answer. Do men and women move through the labour market differently? We have evidence that those without a college degree (early career) do, and this difference emerges over time in the form of both reduced frequency and type of transitions for women. Does the opportunity structure (which $I \times O$-transitions are common and when) change over time? We have strong evidence that it does, at least in some subsectors, regardless of education and gender. Surprisingly, gains in education do not lead to an overall reduction or gain in the number of $I \times O$-transitions, with two exceptions, on the basis of our affinity drift models. Completion of a high school degree is associated with more frequent $I \times O$-transitions for men, whereas completion of a Bachelor degree is associated with more frequent transitions for all more educated workers. For less educated workers, the differential in transition rates is significant, once labour market transition affinities have been controlled, with women making transitions at an overall lower rate. This differential is even larger for college-educated workers, but these are based on an interaction model that is not recommended by DIC. In our DIC-selected, affinity drift models, we find that the overall rate of $I \times O$-transitions declines over time, and this is net of sorting into jobs with longer tenures, so it is robust in a way that has not been thoroughly explored, to date, in the labour economic literature.

Our method provides an important framework for defining what it means to net out the sorting that takes place in the labour market. In addition to the usual regression-like fixed effects, these models include random effects to capture the dynamic nature of job transitions. This structure may be of interest in and of itself, and we present a visualization technique that benefits from the Bayesian methodology, in that only the affinities with the most compelling evidence of strength or difference need to be examined. Through model selection using DIC, different layers of homogeneity and heterogeneity may be explored. The patterning of transitions themselves can be compared across groups, as we did with gender. We note that a hierarchical approach allows modelling extensions at whichever level the researcher deems substantively relevant. So, for example, affinities may be 'explained' through $I \times O$ source- or destination-specific covariates; variation in individual effects may of course be reduced by including important subject level covariates (such as the presence or absence of children). The variance components that we have introduced with this modelling class are fairly generic, in that they are meaningful in any marked point process.

As a final note, we mention latent spatial models (Handcock *et al.*, 2007; Hoff *et al.*, 2002; Gormley and Murphy, 2007), which use a spatial representation of tokens to build what we call affinities in this paper. We have explored this approach with our sequence data with varying degrees of success. Our findings suggest that the symmetry that is implicit in many of the distance-based approaches and the limited explanatory power of lower dimensional embeddings are a little problematic for the more general problem. However, alternative, asymmetric distance metrics that were proposed by Hoff (2005) are quite promising. To the extent that our affinities may be embedded in a lower dimensional space, our work can be characterized as a 'poor man's latent spatial model', in that spatial models can be generated from our affinities, rather than vice versa. In related work (Perrault-Joncas *et al.*, 2010), we are exploring a different class of token embeddings that could encapsulate a portion of the information that is contained in the affinities. The use of Bayesian techniques for our class of models allows a host of modifications to explore ideas such as these.

Acknowledgements

This research was supported in part by the Annie E. Casey Foundation. We thank Annette Bernhardt, Claire Gormley, Mark Handcock, Peter Hoff, Brendan Murphy, Adrian Raftery, Russell Steele, Makram Talih, the reviewers and the Joint Editor for their helpful suggestions.

References

Aalen, O. O. (1988) Heterogeneity in survival analysis. *Statist. Med.*, **7**, 1121–1137.

Abbott, A. (1995) Sequence analysis. *A. Rev. Sociol.*, **21**, 93–113.

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993) *Statistical Models based on Counting Processes*. New York: Springer.

Becker, G. S. (1993) *Human Capital: a Theoretical and Empirical Analysis, with Special Reference to Education*. Chicago: University of Chicago Press.

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

Blau, F., Ferber, M. and Winkler, A. (2005) *The Economics of Women, Men, and Work*. New York: Prentice Hall.

Cox, D. R. (1972) Regression models and life-tables (with discussion). *J. R. Statist. Soc.* B, **34**, 187–220.

Crowder, M. J. (2001) *Classical Competing Risks*. Boca Raton: Chapman and Hall–CRC Press.

Dahinden, C., Parmigiani, G., Emerick, M. C. and Bühlmann, P. (2007) Penalized likelihood for sparse contingency tables with an application to full-length cDNA libraries. *BMC Bioinform.*, **8**, 476.

Danziger, S. and Gottschalk, P. (1993) *Uneven Tides: Rising Inequality in America*. New York: Russell Sage Foundation.
Durbin, R., Eddy, S., Krogh, A. and Mitchison, G. (1998) *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. New York: Cambridge University Press.
Gale, N., Halperin, W. C. and Costanzo, C. M. (1984) Unclassed matrix shading and optimal ordering in hierarchical cluster analysis. *J. Class.*, **1**, 75–92.
Gardecki, R. and Neumark, D. (1998) Order from chaos?: the effects of early labor market experiences on adult labor market outcomes. *Industrl Lab. Relns Rev.*, **51**, 299–322.
Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004) *Bayesian Data Analysis*. Boca Raton: Chapman and Hall–CRC Press.
Gelman, A. and Rubin, D. B. (1992) Inference from iterative simulation (with discussion). *Statist. Sci.*, **7**, 457–511.
Gormley, I. C. and Murphy, T. B. (2007) A latent space model for rank data. *Lect. Notes Comput. Sci.*, **4503**, 90–102.
Guha, S., Rastogi, R. and Shim, K. (2000) ROCK: a robust clustering algorithm for categorical attributes. *Inform. Syst.*, **25**, 345–366.
Hahsler, M., Hornik, K. and Buchta, C. (2008) Getting things in order: an introduction to the R package seriation. *J. Statist. Softwr.*, **25**, no. 3, 1–34.
Handcock, M. S., Raftery, A. E. and Tantrum, J. M. (2007) Model-based clustering for social networks (with discussion). *J. R. Statist. Soc.* A, **170**, 301–354.
Hoff, P. D. (2005) Bilinear mixed-effects models for dyadic data. *J. Am. Statist. Ass.*, **100**, 286–295.
Hoff, P., Raftery, A. E. and Handcock, M. (2002) Latent space approaches to social network analysis. *J. Am. Statist. Ass.*, **97**, 1090–1098.
Hu, W., Li, G. and Li, N. (2009) A Bayesian approach to joint analysis of longitudinal measurements and competing risks failure time data. *Statist. Med.*, **28**, 1601–1619.
Huang, Z. (1998) Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Minng Knowl. Discov.*, **2**, 283–304.
Juang, B. H. and Rabiner, L. R. (1991) Hidden Markov models for speech recognition. *Technometrics*, **33**, 251–272.
Le, S. Q. and Ho, T. B. (2005) An association-based dissimilarity measure for categorical data. *Pattn Recogn Lett.*, **26**, 2549–2557.
Mincer, J. (1974) *Schooling, Experience and Earnings*. New York: Columbia University Press.
Mishel, L., Bernstein, J. and Schmitt, J. (2001) *The State of Working America 2000-2001*. Ithaca: Industrial and Labor Relations Press.
Molenberghs, G. and Verbeke, G. (2005) *Discrete Longitudinal Data*. New York: Springer.
Neumark, D. (ed.) (2000) *On the Job: is Long-term Employment a Thing of the Past?* New York: Russell Sage Foundation.
Osterman, P. (ed.) (1984) *Internal Labor Markets*. Cambridge: Massachusetts Institute of Technology Press.
Osterman, P. and Burton, D. (2004) Ports and ladders: the nature and relevance of internal labor markets in a changing world. In *Oxford Handbook on Work and Organization* (eds P. Tolbert and R. Batt), pp. 425–445. Oxford: Oxford University Press.
Perrault-Joncas, D., Meilă, M. and Scott, M. (2010) Building a job landscape from directional transition data. In *Proc. Association for the Advancement of Artificial Intelligence Fall Symp. Manifold Learning and Its Applications*. To be published.
van de Pol, F. and Langeheine, R. (1990) Mixed Markov latent class models. *Sociol. Methodol.*, **20**, 213–247.
Provasnik, S. and Planty, M. (2008) *Community Colleges: Special Supplement to the Condition of Education 2008 (NCES 2008-033)*. Washington DC: National Center for Education Statistics.
Rosenfeld, R. A. and Spenner, K. I. (1992) Occupational sex segregation and women's early career job shifts. *Wrk Occupns*, **19**, 424–449.
Sarkar, D. (2008) *Lattice: Multivariate Data Visualization with R*. New York: Springer.
Scott, M. A. and Kennedy, B. B. (2005) Pitfalls in pathways: some perspectives on competing risks event history analysis in education research. *J. Educ. Behav. Statist.*, **30**, 413–442.
Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion). *J. R. Statist. Soc.* B, **64**, 583–639.
Topel, R. H. and Ward, M. P. (1992) Job mobility and the careers of young men. *Q. J. Econ.*, **107**, 439–479.