APPLYING HETEROGENEOUS TRANSITION MODELS IN LABOUR ECONOMICS: THE ROLE OF YOUTH TRAINING IN LABOUR MARKET TRANSITIONS[1]

Fabrizia Mealli
Department of Statistics, University of Florence

Stephen Pudney
Public Sector Economics Research Centre
Department of Economics, University of Leicester

revised August 1999

[1] We are grateful to the Economic and Social Research Council for financial support under grants R000234072 and H519255003. We would also like to thank Martyn Andrews, Steve Bradley and Richard Upward for their advice and efforts in constructing the database used in this study. Pinuccia Calia provided valuable research assistance. We received helpful comments from seminar participants at the University of Florence.

ABSTRACT: We illustrate the difficulties raised by four features of realistic transition models in labour economics: dimensionality, institutional constraints, persistence and sample attrition. We estimate a multi-spell multi-state transition model using longitudinal data on the 1988 cohort of male school-leavers in North-West England. The model predicts transitions between college, the government Youth Training Scheme (YTS), employment and unemployment, allowing for endogenous sample attrition and persistent cross-correlated heterogeneity. We simulate the impact of YTS, allowing for endogenous YTS selection induced by heterogeneity. The main findings are a strong positive effect of YTS participation on employment prospects and a large negative impact of early drop-out from YTS.

KEYWORDS: Transition models, youth labour market, training, heterogeneity

JEL CLASSIFICATION: C41, C51, C52, J24

ADDRESS FOR CORRESPONDENCE: Stephen Pudney, Department of Economics, University of Leicester, University Road, Leicester LE1 7RH, UK.

1 Introduction

Measuring the impact of youth training programmes on the labour market continues to be a major focus of microeconometric research and debate. In countries such as the UK, where experimental evaluation of training programmes is infeasible, research is more reliant on tools developed in the literature on multi-state transitions, using models which predict simultaneously the timing and destination state of a transition. Applications include Ridder (1986), Gritz (1993), Dolton Makepeace and Treble (1994) and Mealli and Pudney (1999).

There are several major specification difficulties facing the applied researcher in this area. One is the problem of scale and complexity that besets any realistic model. Active labour market programmes like the British Youth Training Scheme (YTS) and its successors are embedded in the youth labour market, which involves individual transitions between several different states: employment, unemployment and various forms of education or training. In principle, every possible type of transition introduces an additional set of parameters to be estimated, so the dimension of the parameter space rises with the square of the number of separate states, generating both computational and identification difficulties. This is the curse of dimensionality that afflicts many different areas of economics, including demand analysis (Pudney, 1981) and discrete response modelling (Weeks, 1995). A second problem is generated by the institutional features of the training and education system. YTS places are normally limited in duration and college courses are also normally of standard lengths. Conventional duration modelling is not appropriate for such episodes, but flexible semi-parametric approaches (such as that of Meyer, 1990) may introduce far too many additional parameters in a multi-state context. A third important issue is the persistence that is generally found in observed sequences of individual transitions. There are clearly very strong forces tending to hold many individuals in a particular state, once that state has been entered. This is a consideration that motivates the widely-used Goodman (1961) mover-stayer model, which captures an extreme form of persistence. A fourth problem is sample attrition. This is important in any longitudinal study, but particularly so for the youth labour market, where many of the individuals involved may have weak attachment to training and employment, and some resistance to monitoring by survey agencies.

Given the scale of the modelling problem, there is no approach which offers an ideal solution to all these problems simultaneously. In practice we are seeking a model specification and an estimation approach which gives a reasonable compromise between the competing demands of generality and flexibility on the one hand and tractability on the other.

An important economic focus of the analysis is the selection problem. It is well known that selection mechanisms may play an important role in this context: that people who are (self) assigned to training may differ in terms of their unobservable characteristics, and these unobservables may affect also their subsequent labour market experience. If the fundamental role of training per se is to be isolated from the effects of the current pattern of (self-) selection, then it is important to account for the presence of persistent unobserved heterogeneity in the process of model estimation. Once suitable estimates are available, it then becomes possible to assess the impact of YTS using simulations which hold constant the unobservables that generate individual heterogeneity.

2 YTS and the LCS dataset

With the rise of unemployment, and especially youth unemployment, in Britain during the 1980s, government provision of training took on an important and evolving role in the labour market. The one-year YTS programme, introduced in 1983, was extended in 1986 to two years; this was modified and renamed Youth Training in 1989. The system was subsequently decentralised with the introduction of local Training and Enterprise Councils. Later versions of the scheme were intended to be more flexible, but the system remains essentially one of two-year support for young trainees. Our data relate to the 1988 cohort of school-leavers and thus to the two-year version of YTS. In exceptional circumstances, YTS could last longer than 2 years, but for a large majority of participants in the programme, the limit was potentially binding. YTS participants may have had special status as trainees (receiving only a standard YTS allowance), or be regarded as employees in the normal sense, being paid the normal rate for the job. Thus YTS had aspects of both training and employment subsidy schemes. Additional funds were available under YTS for payment to the training providers (usually firms, local authorities and organisations in the voluntary sector) to modify the training programme for people with special training needs who needed additional support.

The data used for the analysis are drawn from a database held on the computer system of the Lancashire Careers Service (LCS), whose duties are those of delivering vocational guidance and a placement service of young people into jobs, training schemes or further education. It generates a wide range of information on all young people who leave school in Lancashire and on the jobs and training programmes undertaken. Andrews and Bradley (1997) give a more detailed description of the dataset.

The sample used in this study comprises 3791 males who entered the labour market with the 1988 cohort of school-leavers, and for whom all necessary information (including postcode, school identifier, etc.) was available. This group was observed continuously from the time they left school, aged 16, in the spring or summer of their fifth year of secondary school, up to the final threshold of 30 June 1992, and every change of participation state was recorded. We identify 4 principal states: continuation in formal education, which we refer to as college (C); employment (E); unemployment (U); and participation in one of the variants of the government youth training scheme (all referred to here as YTS). Note that 16- and 17-year-olds are not eligible for unemployment-related benefits, so unemployment in this context does not refer to registered unemployment, but is the result of a classification decision of the careers advisor. We have also classified a very few short unspecified non-employment episodes as state U. There is a fifth state which we refer to as 'out of sample' (O). This is a catch-all classification referring to any situation in which either the youth concerned is out of the labour force for some reason, or the LCS has lost touch with him or her. Once state O is encountered in the record of any individual, the record is truncated at that point, so that it is an absorbing state in the sense that there can be no subsequent recorded transition out of state O. In the great majority of cases, a transition to O signifies the permanent loss of contact between the LCS and the individual, so that it is, in effect the end of the observation period and represents the usual phenomenon of sample attrition. However, it is important that we deal with the potential endogeneity of attrition, so transitions into state O are modelled together with other transition types.

Many of the individual histories begin with a first spell corresponding to a summer 'waiting' period before starting a job, training or other education. We have excluded all such initial spells recorded by the LCS as waiting periods, and started the work history instead from the succeeding spell. After this adjustment is made, we observe for each individual a sequence of episodes, the final one uncompleted, and for each episode we have an observation on two endogenous variables: its duration and also (for all but the last) the destination state of the transition that terminates it. The data also include observations on explanatory variables such as age, educational attainment and a degree of detail on occupational category of the YTS place and its status (trainee, employee, special funding). Summary statistics for the sample are given in appendix table A1.

There are two obvious features revealed by inspection of the data, which give rise to non-standard elements of the model we estimate. The first of these is shown in figure 1, which plots the smoothed empirical cdf of YTS spell durations. The cdf clearly shows the importance of the 2-year limit on the length of YTS spells and the common occurrence of early termination. Nearly 30% of YTS spells finish within a year and nearly 50% before the two-year limit. Conventional transition model specifications cannot capture this feature, and we use below an extension of the limited competing risks (LCR) model introduced by Mealli, Pudney and Thomas (1996). Figure 2 plots the smoothed empirical hazard function of durations of college spells, and reveals another non-standard feature in the form of peaks at durations around 0.9 and 1.9 years (corresponding to educational courses lasting 1 and 2 academic years). Again, we make an appropriate adaptation to our model to cope with this.

**** FIGURES 1 AND 2 HERE ****

3 A correlated random-effects transition model

Longitudinal data of the type described above covers each individual from the beginning of his work history to an exogenously-determined date at which the observation period ends. This generates for each sampled individual a set of k observed episodes (note that k is a random variable). Each episode has two important attributes: its duration; and the type of episode that succeeds it (the destination state). In our case there are 4 possible states that we might observe. We write the observed endogenous variables $r_0; \delta_1; r_1; \ldots; \delta_k$, where δ_s is the duration of the sth episode, and r_s is the destination state for the transition that brings it to an end. Thus the econometric model can be regarded as a specification for the joint distribution of a set of k continuous variables (the δ_s) and k discrete variables (the r_s). For each episode there is a vector of observed explanatory variables, x_s, which may vary across episodes but which is assumed constant over time within episodes.

The model we estimate in this study is a modified form of the conventional heterogeneous multi-spell multi-state transition model (see Pudney (1989) and Lancaster (1990) for surveys). Such models proceed by partitioning the observed work history into a sequence of episodes. For the first spell of the sequence, there is a discrete distribution of the state variable r_0 with conditional probability mass function $P(r_0|x_0, v)$. Conditional on past history, each successive episode for $s = 1 \ldots k-1$ is characterised by a joint density/mass function $f(\delta_s, r_s|x_s, v)$, where x_s may include functions of earlier state and duration variables, to allow for lagged state dependence. The term v is a vector of unobserved random effects, each element normalised to have unit variance; v is constant over time, and can thus generate strong serial dependence in the sequence of episodes. Under our sampling scheme, the final observed spell is usually an episode of C, E, U or YTS, which is still in progress at the end of the observation period. For this last incomplete episode, the eventual destination state is unobserved, and its distribution is characterised by a survivor function $S(\delta_k|x_k, v)$ which gives the conditional probability of the kth spell lasting at least as long as δ_k. Conditional on the observed covariates $X = \{x_0 \ldots x_k\}$ and the unobserved effects v, the joint distribution of $r_0, \delta_1, r_1, \ldots, \delta_k$ is then:

$$f(r_0, \delta_1, r_1 \ldots \delta_k|X, v) = P(r_0|x_0, v) \times \left[\prod_{s=1}^{k-1} f(\delta_s, r_s|x_s, v)\right] S(\delta_k|x_k, v) \qquad (1)$$

For the smaller number of cases where the sample record ends with a transition to state O (in other words attrition), there is no duration for state O and the last component of (1) is $S(\delta_k|x_k, v) \equiv 1$. There is a further complication for the still fewer cases where the record ends with a YTS $\rightarrow$ O transition, and this is discussed below.

The transition components of the model (the pdf f and the survivor function S) are based on the notion of a set of origin- and destination-specific transition intensity functions for each spell. These give the instantaneous probability of exit to a given destination at a particular time, conditional on no previous exit having occurred. Thus, for any given episode, spent in state i, the jth transition intensity function $h_{ij}(t|z;v)$ is given by:

$$\Pr(r = j, \delta \in (t, t+dt) | \delta \geq t; x; v) = h_{ij}(t|x;v)dt \qquad (2)$$

where x and v are respectively vectors of observed and unobserved covariates which are specific to the individual but may vary across episodes for a given individual. Our data are constructed in such a way that an episode can never be followed by another episode of the same type, so the i,ith transition intensity h_{ii} does not exist. The joint probability density/mass function of exit route, r, and realised duration, δ, is then constructed as:

$$f(r, \delta|x; v) = h_{ir}(\delta|x;v) \exp\left[-\sum_{j \neq i} I_{ij}(\delta|x;v) \right] \qquad (3)$$

where $I_{ij}(\delta|x;v)$ is the i,jth integrated hazard:

$$I_{ij}(\delta|x;v) = \int_0^{\delta} h_{ij}(t|x;v)dt \qquad (4)$$

Since the random effects v are unobserved, (1) cannot be used directly as the basis of an estimated model. However, if we assume a specific joint distribution function, $G(v)$, for the random effects, they can be removed by integration and estimation can then proceed by maximising the following log-likelihood based on (1) with respect to the model parameters:

$$\ln L = \sum_{n=1}^{N} \ln \left\{ \int P(r_0|x_0;v) \left[\prod_{s=1}^{k-1} f(\delta_s, r_s|x_s;v) \right] S(\delta_k|x_k;v)dG(v) \right\} \qquad (5)$$

where the suffix $n = 1...N$ indexes the individuals in the sample.

It is important to realise that, for estimation purposes, the definition (2) of the transition intensity function is applicable to any form of continuous-time multi-state transition process. It is also possible to think of such a process in terms of a competing risks structure, involving independently-distributed latent durations for transition to each possible destination, with the observed duration and transition corresponding to the shortest of the latent durations. These two interpretations are observationally equivalent in the sense that it is always possible to construct a set of indpendent latent durations consistent with any given set of transition intensities. This aspect of the interpretation of the model therefore has no impact on estimation. However, when we come to simulating the model under assumptions of changed policy or abstracting from the biasing effects of sample attrition, then interpretation of the structure becomes important. For simulation purposes the competing risks interpretation has considerable analytical power, but at the cost of a very strong assumption about the structural invariance of the transition process. We return to these issues in section 5 below.

The specifications we use for the various components of the model are described in the following sections.

3.1 Heterogeneity

We now turn to the specification of the persistent random-effects. First note that there has been some debate about the practical importance of heterogeneity in applied modelling. Ridder (1987) has shown that neglecting unobserved heterogeneity results in biases that are negligible, provided a sufficiently flexible baseline hazard is specified. However, his results apply only to the simple case of single-spell data with no censoring. In the multi-spell context where random effects capture persistence over time as well as inter-individual variation, and where there is a non-negligible censoring frequency, heterogeneity cannot be assumed to be innocuous. We opt instead for a model in which there is are reasonable degrees of flexibility in both the transition intensity functions and the heterogeneity distribution.

The same problem of dimensionality is found here as in the observable part of the model. Rather than the general case of 16 unobservables, each specific to a distinct origin-destination combination, we simplify the structure by using persistent heterogeneity to represent those unobservable factors which predispose individuals towards long or short stays in particular states. Our view is that this sort of state-dependent 'stickiness' is likely to be the main dimension of unobservable persistence - an assumption similar to, but less extreme than, the assumption underlying the familiar mover-stayer model (Goodman, 1961). Thus we use a four-factor specification, where each of the four random effects is constant over time and linked to a particular state of origin rather than destination. We assume that the observed covariates and these random effects enter the transition intensities in an exponential form, so that in general $h_{ij}(t|x;v)$ can be expressed as $h^*_{ij}(t;x;\omega_i v_i)$ where

ω_i is a scale parameter. There is again a conflict between flexibility and tractability, in terms of the functional form of the distribution of the unobservables v_i. One might follow Heckman and Singer (1984) and Gritz (1993) by using a semi-parametric mass-point distribution, where the location of the mass-points and the associated probabilities are treated as parameters to be estimated. Van den Berg (1997) has shown in the context of a 2-state competing risks model that this specification has an advantage over other distributional forms (including the normal) in that it permits a wider range of possible correlations between the two underlying latent durations. However, in our 4-dimensional setting, this would entail another great expansion of the parameter space. Since there is, in any case, a fair amount of informal empirical experience suggesting that distributional form is relatively unimportant provided the transition intensities are specified sufficiently flexibly, we are content to assume a normal distribution for the v_i.

We introduce correlation across states in the persistent heterogeneity terms in a simple way which nevertheless encompasses the two most common forms used in practice. This is done by constructing the v_i from an underlying vector ξ as follows:

$$v_i = (1 - \lambda)\xi_i + \lambda \sum_{p=1}^{4} \xi_p \tag{6}$$

where λ is a single parameter controlling the general degree of cross-correlation. Under this specification, the correlation between any pair of heterogeneity terms, $\omega_i v_i$ and $\omega_j v_j$, is $2\,\mathrm{sgn}(\omega_i \omega_j)\lambda(1+\lambda)/(1+3\lambda^2)$. Note that one of the scale parameters should be normalised with respect to its sign, since (with the ξ_i symmetrically distributed) the sample distribution induced by the model is invariant to multiplication of all the ω_i by -1. There are two important special cases of (6); $\lambda = 0$ corresponds to the assumption of independence across (origin) states; $\lambda = 1$ yields the one-factor specification discussed by Lindeboom and van den Berg (1994).

3.2 The initial state

Our model for the initial state indicator r_0 is a 4-outcome multinomial logit (MNL) structure of the following form:

$$\Pr(r_0 = j | x_0; v) = \frac{\exp(x_0 \gamma_j + \psi_j v_j)}{\sum_{i=1}^{4} \exp(x_0 \gamma_i + \psi_i v_i)} \tag{7}$$

where γ_1 is normalised at 0. The parameters ψ_j are scale parameters that also control the correlation between the initial state and successive episodes.

3.3 The transition model

For a completely general model, 16 transition intensities should be specified – a practical impossibility, since this would lead to an enormously large parameter set. This dimensionality problem is very serious in applied work, and there are two obvious solutions, neither of which is ideal. The most common approach is to reduce the number of states, either by combining states (for example college and training) into a single category, or by deleting individuals who make certain transitions (such as those who return to college or who leave the sample by attrition). The consequences of this type of simplification are potentially serious and obscure, since it is impossible to test the implicit underlying assumptions without maintaining the original degree of detail. In these two examples, the implicit assumptions are respectively: that transition rates to and from college and training are identical; and that the processes of transition to college or attrition are independent of all other transitions. Neither assumption is very appealing, so we prefer to retain the original level of detail in the data, and to simplify the model structure in (arguably) less restrictive and (definitely) more transparent ways.

We adopt as our basic model a simplified specification with separate intensity functions only for each exit route, letting the effect of the state of origin be captured by dummy variables included with the other regressors, together with a few variables which are specific to the state of origin (specifically SPECIAL and YTCHOICE describing the nature of the training placement in YTS spells, and YTMATCH, CLERK and TECH describing occupation and training match for $E \to U$ transitions).

The best choice for the functional form $h_{ij}(t|x; v)$ is not obvious. Meyer (1990) has proposed a flexible semi-parametric approach which is now widely used in simpler contexts. It entails estimating the dependence of h_{ij} on t as a flexible step function, which introduces a separate parameter for each step. In our case, with 16 different transition types, this would entail a huge expansion in the dimension of the parameter space. Instead, we adopt a different approach. We specify a generic parametric functional form for $h_{ij}(t|x;v)$, which is chosen to be reasonably flexible (in particular, not necessarily monotonic). However, we also exploit our a priori knowledge of the institutional features of the education and training systems to modify these functional forms to allow for the occurrence of 'standard' spell durations for many episodes of YTS and college. These modifications to the basic model are described below. The basic specification we use is the Burr form:

$$h_{ij}(t|x;v) = \frac{\exp(z_i\beta_j + \omega_i v_i)\alpha_{ij}t^{\alpha_{ij}-1}}{1+\sigma_j^2\exp(z_i\beta_j + \omega_i v_j)\,t^{\alpha_{ij}}} \qquad (8)$$

where z_i is a row vector of explanatory variables constructed from x in some way that may be specific to the state of origin, i. Note that the absence of a subscript i on β_j is not restrictive: any form $z_i^*\beta_{ij}^*$ can be rewritten $z_i\beta_j$ by defining z_i appropriately, using origin-specific dummy variables in additive and multiplicative form. The form (8) is non-proportional and not necessarily monotonic, but it has the Weibull form as the special case $\sigma_{ij} = 0$. The parameters α_{ij} and σ_{ij} are specific to the origin-destination combination i,j, and this gives the specification considerable flexibility. The Burr form has the following survivor function:

$$S_{ij}(\delta|x;v) = \left[1+\sigma_{ij}^2\exp(z_i\beta_j + \omega_i v_i)\,t^{\alpha_{ij}}\right]^{-1/\sigma_{ij}^2} \qquad (9)$$

Note that the Burr model can be derived as a Weibull-gamma mixture, with the gamma heterogeneity spell-specific and independent across spells, but such an interpretation is not necessary and is not in any case testable without further a priori restrictions, such as proportionality.

3.4 YTS spells

There are two special features of YTS episodes that call for some modification of the standard transition model outlined above. One relates to attrition (YTS$\to$O transitions). Given the monitoring function of the LCS for YTS trainees, it is essentially impossible for the LCS to lose contact with an individual while he remains in a YTS place. Thus a YTS $\rightarrow$ O transition must coincide with a transition from YTS to either C, E or U, where the destination state is unobserved by the LCS. Thus a transition of this kind is a case where the observed duration in the $k-1$th spell (YTS) is indeed the true completed YTS duration, δ_{k-1}, but the destination state r_{k-1} is unobserved. For the small number of episodes of this kind, the distribution (3) is:

$$f(\delta_{k-1}|x;v) = \sum_{j \neq i} h_{ij}(\delta_{k-1}|x_{k-1};v) \exp\left[- \sum_{j \neq i} I_{ij}(\delta_{k-1}|x_{k-1};v) \right] \qquad (10)$$

A second special feature of YTS episodes is the exogenous 2-year limit imposed on them by the rules of the system. Mealli, Pudney and Thomas (1996) proposed a simple model for handling this complication which figure 1 shows to be so important in our data. The method involves making allowance for a discontinuity in the destination state probabilities conditional on YTS duration at the 2-year limit. The transition structure operates normally until the limit is reached, at which point a separate MNL structure comes into play. Thus, for a YTS episode:

$$\Pr(r = j|\delta = 2;x;v) = \frac{\exp(w\pi_j + \theta_j v_{YTS})}{\sum_p \exp(w\pi_p + \theta_p v_{YTS})} \qquad (11)$$

where w is a vector of relevant covariates. In the sample there are no cases at all of a college spell following a full term YTS episode; consequently the summation in the denominator of (11) runs over only two alternatives, E and U. The π and θ parameters are normalised to zero for the latter.

3.5 Bunching of college durations

To capture the two peaks in the empirical hazard function for college spells, we superimpose two spikes uniformly across all transition intensity functions for college spells. Thus, for origin C and destinations j = E, U, O, YTS the modified transition intensities are:

$$h^{*}_{Cj}(t|x;v) = h_{Cj}(t|x;v) \exp(\rho_1 A_1(t) + \rho_2 A_2(t))$$

where $A_1(t)$ and $A_2(t)$ are indicator functions of $(0.8 \leq t \leq 1)$ and $(1.8 \leq t \leq 2)$ respectively.

3.6 Simulated maximum likelihood

The major computational problem involved in maximising the log-likelihood function (5) is the computation of the 4-dimensional integral which defines each of the N likelihood components. The approach we take is to approximate the integral by an average over a set of Q pseudo-random deviates generated from the assumed joint standard lognormal distribution for ξ. We also make use of the antithetic variates technique to improve the efficiency of simulation, with all the underlying pseudo-random normal deviates re-used with reversed sign to reduce simulation variance. Thus, in practice we maximise numerically the following approximate log-likelihood.

$$L(\theta) = \sum_{n=1}^{N} \ln\left(\frac{1}{Q} \sum_{q=1}^{Q} \frac{l_n(v^q) + l_n(-v^q)}{2} \right)$$

where $l_n(v^q)$ is the likelihood component for individual n, evaluated at a simulated value v^q for the unobservables.

This simulated ML estimator is consistent and asymptotically efficient as Q and $N \to \infty$ with $\sqrt{N}/Q \to 0$ (see Gourieroux and Monfort 1996). Practical experience suggests that it generally works well even with small values of Q (see Mealli and Pudney 1996) for evidence on this). In this study, we use $Q = 40$. The asymptotic approximation to the covariance matrix of the estimated parameter vector $\hat{\theta}$ is computed via the conventional OPG formula, which gives a consistent estimate in the usual sense, under the same conditions on Q and N.

4 Estimation results

4.1 Estimation strategy

Our preferred set of estimates is given in appendix tables A4-A7. These estimates are the outcome of a process of exploration which of necessity could not follow the 'general-to-specific' strategy that is usually favoured, since the most general specification within our framework would have approximately 450 parameters, with 16 separate random effects, and would certainly not be possible to estimate with available computing and data resources. Even after the considerable simplifications we have made, there remain 117 parameters in the model. Apart from the constraints imposed on us by this dimensionality problem, we have adopted throughout a conservative criterion, and retained in the model all variables with coefficient t-ratios in excess of 1.0. Some further explanatory variables were tried in earlier specifications, but found to be insignificant everywhere. These were all dummy variables, distinguishing those who: took a non-academic subject mix at school; were in a technical/craft occupation when in work; and those who had trainee rather than employee status when in YTS. Thus the sparse degree of occupational and training detail in the final model is consistent with the available sample information.

The relatively low frequencies of certain transition types have made it necessary to impose further restrictions to achieve adequate estimation precision. In particular, the α_{ij} and σ_{ij} parameters for destination $j = 0$ (sample attrition) could not be separately estimated, so we have imposed the restrictions $\alpha_{i0} = \alpha_{k0}$ and $\sigma_{i0} = \sigma_{k0}$ for all i, k.

4.2 The heterogeneity distribution

Table A7 gives details of the parameters underlying the joint distribution of the persistent heterogeneity terms appearing in the initial state and transition structures. Heterogeneity appears strongly significant in the initial state logit only for the exponents associated with employment and YTS. The transition intensities have significant origin-specific persistent heterogeneity linked to C, U and YTS. The evidence for heterogeneity associated with employment is weak, although a very conservative significance level would imply a significant role for it, and also in the logit that comes into force at the 2-year YTS limit (θ_E in Table A6).

The estimated value of λ implies a correlation of $\pm$ 0.30 between any pair of scaled heterogeneity terms, $\omega_i v_i$, in the transition part of the model. There is a positive correlation between the heterogeneity terms associated with the E, U and YTS states, but these are all negatively correlated with the heterogeneity term associated with college. This implies a distinction between those who are predisposed towards long college spells and those with a tendency towards long U and YTS spells. Note that the estimate of λ is significantly different from both 0 and 1 at any reasonable significance level, so both the one-factor and independent stochastic structures are rejected by these results, although independence is clearly preferable to the single-factor assumption.

Wherever significant, there is a negative correlation between the random effect appearing in a branch of the initial state logit, $\psi_j v_j$, and the corresponding random effect in the transition structure, $\omega_i v_i$. This implies, as one might expect, that a high probability of starting in a particular state tends to be associated with long durations in that state.

The logit structure which determines exit route probabilities once the 2-year YTS limit is reached involves a random effect which is correlated with the random effect in the transition intensities for YTS spells. However, this is of doubtful significance.

4.3 Duration dependence

The functional forms of the destination-specific transition intensities are plotted in figures 3-6 conditional on different states of origin. In constructing these plots, we have fixed the elements of the vector of observed covariates x at the representative values listed in Table 1 below. The persistent origin-specific random effects, v, are fixed at their median values, 0. The relative diversity of the functional forms across states of origin gives an indication of the degree of flexibility inherent in the structure we have estimated. There are several points to note here.

Firstly, the estimated values of the thirteen σ_{ij} parameters are significant in all but four cases, implying that the restricted Weibull form would be rejected against the Burr model that we have estimated; thus, if we were prepared to assume proportional Weibull competing risks, this would imply a significant role for destination-specific Gamma heterogeneity uncorrelated across spells. Secondly, the transition intensities are not generally monotonic; an increasing then falling pattern is found for the transitions $C \to YTS$, $E \to YTS$ and $YTS \to U$. The aggregate hazard rate is non-monotonic for exits from college, employment and unemployment. Thirdly, transition intensities for exit to college are very small for all states of origin except unemployment, where there is a sizeable intensity of transition into education at short unemployment durations. The generally low degree of transition into state C reflects the fact that, for most people, formal post-16 education is a state entered as first destination after leaving school, or not at all. However, the fact that there are unobservables common to both the initial state and transition parts of the model implies that the decision to enter college after school is endogenous and cannot be modelled separately from the transitions among the other three states.

Figure 8 shows the aggregated hazard rates, $h_{i:}(t|x;v) = \sum_j h_{ij}$, governing exits from each state of origin, i. The typical short unemployment durations imply a high hazard rate for exits from unemployment, but declining strongly with duration, implying a heavy right-hand tail for the distribution of unemployment durations. For the other three states of origin, the hazard rates are rather flatter, except for the 1- and 2-year peaks for college spells. Note that we cannot distinguish unambiguously between true duration dependence and the effects of non-persistent heterogeneity here, at least not without imposing restrictions such as proportionality of hazards.

**** FIGURES 3 - 8 HERE ****

4.4 Simulation strategy

The model structure is sufficiently complex that it is difficult to interpret the parameter estimates directly. Instead we use simple illustrative simulations to bring out the economic implications of the estimated parameter values. The 'base case' simulations are performed for a hypothetical individual who is average with respect to quantitative attributes and modal with respect to most qualitative ones. An exception to this is educational attainment, which we fix at the next-to-lowest category (GCSE2), to represent the group for whom YTS is potentially most important. Thus our representative individual has the characteristics listed in Table 1.

TABLE 1 Attributes of illustrative individual

Attribute	Assumption used for simulations
Date of birth	28 February 1972
Ethnic origin	white
Educational attainment	one or more GCSE passes, none above grade D
Subject mix	academic mix of school subjects
Health	no major health problem
School quality	attended a school where 38.4% of pupils achieved 5 or more GCSE passes
Area quality	lives in a ward where 77.9% of homes are owner-occupied
Local unemployment	unemployment rate in ward of residence is 10.3%
Date of episode	current episode began on 10th March 1989
Previous YTS	no previous experience of YTS
Occupation	when employed, is neither clerical nor craft/technical
Special needs	has no special training needs when in YTS

Simulations are conducted in the following way. For the representative individual defined in Table 1, 500 5-year work histories are generated via stochastic simulation of the estimated model.[1] These are summarised by calculating the average proportion of time spent in each of the four states and the average frequency of each spell-type. To control for endogenous selection and attrition, we keep all the random effects fixed at their median values of zero, and reset all transition intensities into state 0 to zero. We then explore the effects of the covariates by considering a set of hypothetical individuals with slightly different characteristics from the representative individual. These explore the effects of ethnicity, educational attainment and the nature of the locality. For the last of these, we change the SCHOOL, AREA and URATE variables to values of 10%, 25% and 20% respectively.

[1] The simulation process involves sampling from the type I extreme value distribution for the logit parts of the model, and from the distribution of each latent duration for the transition part. In both cases, the inverse of the relevant cdf was evaluated using uniform pseudo-random numbers.

TABLE 2 Simulated effects of the covariates for a hypothetical individual

Simulated individual	Spell type	Proportion of time (%)	Proportion of non-college time	Frequency of spells (%)	Mean no. of spells
	C	19.7	–	18.2	
Base case	E	56.9	70.9	39.0	2.59
(see table 1)	U	7.2	9.0	24.1	
	YTS	16.2	20.2	18.7	
	C	58.7	–	55.3	
Non-white	E	24.2	58.6	17.8	2.01
	U	9.3	22.5	18.8	
	YTS	7.8	18.9	8.0	
	C	30.6	–	29.9	
1-3 GCSEs at	E	47.7	68.7	33.1	2.16
grade C or	U	9.3	13.4	23.2	
better	YTS	12.4	17.9	13.8	
	C	65.6	–	64.8	
More than 3	E	24.3	70.6	17.2	1.55
GCSEs at grade	U	4.5	13.1	12.0	
C or better	YTS	5.6	16.3	6.0	
	C	22.2	–	21.3	
Major health	E	52.4	67.4	33.6	2.57
problem	U	4.8	6.2	22.3	
	YTS	20.6	26.5	22.8	
	C	18.4	–	17.0	
Poor school	E	52.8	64.7	35.9	2.75
& area quality	U	9.1	11.2	24.4	
	YTS	19.6	24.0	22.7	

Note: 500 replications over a 5-year period; random effects fixed at 0

Table 2 reveals a large impact for the variables representing ethnicity and educational attainment, in comparison with the variables used to capture the influence of social background. An individual identical to the base case, but from a non-white ethnic group (typically south Asian in practice) is predicted to have a much higher probability of remaining in full-time education (59% of the 5-year period on average, compared to 20% for the reference white individual). However, for ethnic minority individuals who are not in education, the picture is gloomy. Non-whites have a much higher proportion of their non-college time (22% compared to 9%) spent unemployed, with a roughly comparable proportion spent in YTS.

The effect of increasing educational attainment at GCSE is to increase the proportion of time spent in post-16 education from 20% to 31% and 66% for the three GCSE performance classes used in the analysis. Improving GCSE performance has relatively little impact on the amount of time predicted to be spent in unemployment and its main effect is to generate a substitution of formal education for employment and YTS training.

There is a moderate estimated effect of physical and social disadvantage. Individuals identified as having some sort of (subjectively defined) major health problem are predicted to spend a greater proportion of their first five post-school years in college or YTS (43% rather than 36%) compared with the otherwise similar base case. This displaces employment (52% rather than 57%), but also reduces the time spent unemployed by about two and a half percentage points. In this sense, there is evidence that the youth employment system was managing to provide effective support for the physically disadvantaged, if only temporarily. After controlling for other personal characteristics, there is a significant role for local social influences as captured by the occupational, educational and housing characteristics of the local area, and the quality of the individual's school. Poor school and neighbourhood characteristics are associated with a slightly reduced prediction of time spent in college and employment, with a corresponding increase in unemployment and YTS tenure. Nevertheless, compared to race and education effects, these are minor influences.

4.5 The effects of unobserved heterogeneity

To analyse the effects of persistent heterogeneity specific to each state of origin, we conduct simulations similar to those presented in the previous paragraph. The results are shown in figures 9-12. We consider the representative individual and then conduct the following sequence of stochastic simulations. For each state $i = C, E, U, YTS$ set all the heterogeneity terms to zero except for one, v_i, whose value is varied over a grid of values in the range $[-2; 2]$ (covering approximately 4 standard deviations). At each point in the grid, 500 5-year work histories are simulated stochastically and the average proportion of time spent in each state is recorded. This is done for each of the four v_i, and the results plotted. The plots in figures 9-12 show the effect of varying each of the heterogeneity terms on the proportion of time spent respectively in college, employment, unemployment and unemployment.

The striking feature of these plots is the large impact of these persistent unobservable factors on the average proportions of the 5-year simulation period spent in each of the four states. This is particularly true for college, where the proportion of time spent in education falls from over 20% at $v_C = 0$ to almost zero at $v_C = 2$, with a corresponding rise in the time spent in employment and unemployment. The proportion of time spent unemployed (essentially the unemployment rate among individuals of the representative type) is strongly influenced by all four state-specific random effects, with a 6 percentage point variation in the unemployment rate.

**** FIGURES 9 - 12 HERE ****

5 Simulations of the effects of YTS

We now bring out the policy implications of the model by estimating the average impact of YTS for different types of individual, again using stochastic simulation as a basis. A formal policy simulation can be conducted by comparing the model's predictions in two hypothetical worlds in which the YTS system does and does not exist. The latter (the 'counter-factual') requires the estimated model to be modified in such a way that YTS spells can no longer occur. The results and the interpretational problems associated with this exercise are presented in section 5.2 below. However, first we consider the effects of YTS participation and and of early dropout from YTS, by comparing the simulated labour market experience of YTS participants and non-participants within a YTS world. For this we use the model as estimated, except that the 'risk' of attrition (transition to state O) is deleted.

5.1 The effects of YTS participation

We work with the same set of reference individuals as in sections 4.4-4.5 above. Again, the state-specific random effects are fixed at their median values of 0, so that the simulations avoid the problems of endogenous selection arising from persistent unobservable characteristics. This time the

500 replications are divided into two groups: the first one contains histories with no YTS spell and the second one histories with at least one YTS spell. Thus we have two groups of fictional individuals, identical except that the first happen by chance to have avoided entry into YTS, while the second have been through YTS. To make the comparison as equal as possible, we take the last 3 years of the simulated 5-year history for the non-YTS group and the post-YTS period (which is of random length) for the YTS group. We exclude from each group those individuals for whom there is a college spell in the reference period, thus focusing attention solely on labour market participants.

Figure 13 shows, for the base case individual, the difference in simulated unemployment incidence for the two groups. At the median value of the random effects, the difference amounts to approximately 5 percentage points, so that YTS experience produces a substantially reduced unemployment risk. We have investigated the impact of unobservable persistent heterogeneity by repeating the simulations for a range of fixed values for each of the v_i. Figure 13 shows the plot for v_U; broadly similar patterns are found for the other v_i, suggesting that the beneficial effect of YTS participation is more or less constant across individuals with differing unobservable characteristics.

**** FIGURE 13 HERE ****

Table 3 shows the influence of observable characteristics, summarising the results of simulations for the base case and peturbations with respect to ethnicity, education and area/school quality. The beneficial effects of YTS participation are evident in all cases, but are particularly strong for members of ethnic minorities and for those with better levels of school examination achievement. Note that these are the groups with the highest probabilities of full-term YTS spells.

TABLE 3 Simulated effects of YTS participation on employment frequency and duration for hypothetical individuals

Simulated individual	Replications with no YTS spell: % period in work	Replications with no YTS spell: % spells in work
Base case	89.9	86.0
Non-white	65.3	60.5
1-3 GCSEs at grade C	85.1	83.0
> 3 GCSEs at grade C	88.3	85.4
Low school & area quality	86.5	84.0

Simulated individual	Replications containing a YTS spell: % post-YTS period in work	% post-YTS spells n work	Mean YTS duration	% YTS spells full-term
Base case	95.1	89.2	1.47	51.0
Non-white	86.6	80.2	1.56	59.5
1-3 GCSEs at grade C	96.5	91.8	1.62	63.6
> 3 GCSEs at grade C	98.6	96.8	1.75	73.1
Low school & area quality	90.1	81.1	1.47	51.8

Note: 500 replications over a 5-year period; random effects fixed at 0

5.2 Simulating a world without YTS

The ultimate aim of this type of modelling exercise is to say something about the economic effects of implementing a training/employment subsidy scheme such as YTS. The obvious way to attempt this is to compare simulations of the model in two alternative settings: one (the 'actual') corresponding to the YTS scheme as it existed during the observation period; and other (the 'counterfactual') corresponding to an otherwise identical hypothetical world in which YTS does not exist. There are well-known and obvious limits on what can be concluded from this type of comparison, since we have no direct way of knowing how the counterfactual should be designed. Note that this is not a problem specific to the simulations presented in this paper; any attempt to give a policy-oriented interpretation of survey-based econometric results is implicitly subject to the same uncertainties.

The design of a counterfactual case requires assumptions about three major sources of interpretative error, usually referred to, rather loosely, as deadweight loss, displacement and scale effects. Deadweight loss refers to the possibility that YTS (whose objective is employment promotion) may direct some resources to those who would have found employment even without YTS. Since YTS has some of the characteristics of an employment subsidy, this is a strong possibility. It seems likely that if YTS had not existed during our observation period, then some of those who were in fact observed to participate in YTS would have been offered conventional employment instead, possibly on old-style private apprenticeships. Displacement refers to a second possibility that a net increase in employment for the YTS target group might be achieved at the expense of a reduction in the employment rate for some other group, presumably older, poorly qualified workers. Note, however, that displacement effects can also work in the other direction. For example, Johnson and Layard (1986) showed, in the context of a segmented labour market with persistent unsatisfied demand for skilled labour and unemployment amongst unskilled workers, that training programmes can simultaneously produce an earnings increase and reduced unemployment probability for the trainee (which might be detected by an evaluation study) and also make available a job for one of the current pool of unemployed. A third interpretative problem is that the aggregate net effect of a training programme may be nonlinear in its scale, so that extrapolation of a micro-level analysis gives a misleading prediction of the effect of a general expansion of the scheme. This mechanism may work, for instance, through the effect of the system on the relative wages of skilled and unskilled labour (see Blau and Robins (1987)).

The evidence on these effects is patchy. Deakin and Pratten (1987) give results from a survey of British employers which suggests that roughly a half of YTS places may have either gone to those who would have been employed by the training provider anyway or substituted for other types of worker (with deadweight loss accounting for the greater part of this inefficiency). However, other authors have found much smaller effects (see Jones (1988)), and the issue remains largely unresolved. Blau and Robins (1987) found some empirical evidence of a nonlinear scale effect, by estimating a significant interaction between programme size and its effects. The need for caution in interpreting the estimated effects of YTS participation is evident, but there exists no clear and simple method for adjusting for deadweight, displacement and scale effects.

The economic assumptions we make about the counterfactual have a direct parallel with the interpretation of the statistical transition model. To say anything about the effects of removing the YTS programme from the youth labour market requires some assumption about how the statistical structure would change if we were to remove one of the possible states. The simulations we present in Table 4 correspond to the very simplest counterfactual case and, equivalently, to the simplest competing risks interpretation.. In the non-YTS world, we simply force the transition intensities for movements from any state into YTS, and the probability of YTS as a first destination, to be zero. The remainder of the estimated model is left unchanged, so that it generates transitions between the remaining three states. In other words, we interpret the model as a competing risks structure, in which the YTS `risk' can be removed without altering the levels of `hazard' associated with the other possible destination states. This is, of course a strong assumption and avoids the issue of the macro-level effects which might occur if there really were an abolition of the whole state training programme.

As before, we work with a set of hypothetical individuals, and remove the effect of inter-individual random variation by fixing the persistent individual-specific random effects at zero. Table 4 then summarises the outcome of 500 replications of a stochastic simulation of the model. The sequence of pseudo-random numbers used for each replication is generated using a randomly-selected seed specific to that replication; within replications, the same pseudo-random sequence is used for the actual and counter-factual cases. Note that the results are not directly comparable to those presented in section 4.2 which compared YTS participants and non-participants, since we are considering here the whole 5-year simulation period rather than the later part of it. We are also not focusing exclusively on the labour market, since we retain in the analysis individuals who are predicted by the simulations to remain in education. A third major difference is that the analysis of section 4.2 did not consider the effects of differences in YTS participation frequency together with the effects of YTS per participant, whereas the simulations reported here will necessarily show bigger impacts of abolition for groups with high YTS participation rates.

On the basis of these results in Table 4, the effect of the YTS programme on employment frequencies is important but moderate: a fall of no more than 5 percentage points in the proportion of time spent in employment. Instead, the major impact of abolition is on time spent in education and in unemployment. With YTS abolished, the proportion of time spent in unemployment rises for most cases by between 6 and 14 percentage points, although the rise is necessarily much smaller for those with low probabilities of YTS participation (notably non-whites and those with good GCSE results). The simulated degree of substitution between continuing education and YTS is substantial, with the duration rising by 4-9 percentage points in every case. The rise is largest for individuals disadvantaged by ethnicity, health or social/educational background; but also for those with a modestly increased level of school examination achievement relative to the base case.

TABLE 4 Simulated work histories for hypothetical individuals with and without the YTS scheme in existence

Simulated individual	Spell type	Proportion of time for non-YTS world (%)	Increase compared to YTS world (% points)
	C	24.0	+4.3
Base case	E	58.1	+1.2
	U	17.9	+10.7
	C	65.3	+6.6
Non-white	E	21.7	-2.5
	U	13.0	+3.7
	C	38.8	+8.2
1-3 GCSEs at	E	43.1	-4.6
grade C	U	18.1	+8.8
	C	70.2	+4.6
> 3 GCSEs at	E	22.3	-2.0
grade C	U	7.4	+2.9
	C	31.5	+9.3
Major health	E	49.5	-2.9
problem	U	19.0	+14.2
	C	25.7	+7.3
Poor school	E	52.1	-0.7
& area quality	U	22.3	+13.2

Note: 500 replications over a 5-year period; random effects fixed at 0

6 Concluding remarks

We have estimated a large and highly complex transition model designed to address the formidable problems of understanding the role played by government training schemes in the labour market experience of school-leavers. The question "what is the effect of YTS?" is a remarkably complex one, and we have looked at its various dimensions using stochastic simulation of the estimated model. Abstracting from endogenous (self-) selection into YTS, we have found evidence suggesting a significant improvement in subsequent employment prospects for those who do go through YTS, particularly in the case of YTS 'stayers'. This is a rather more encouraging conclusion than that of Dolton, Makepeace and Treble (1994), and is roughly in line with the earlier applied literature, based on less sophisticated statistical models. Our results suggest that, for the first five years after reaching school-leaving age, YTS appears mainly to have absorbed individuals who would otherwise have gone into unemployment or stayed on in the educational system. The employment promotion effect of YTS among 16-21 year olds might in contrast be judged worthwhile but modest. Our estimated model is not intended to have any direct application to a period longer than the 5-year simulation period we have used. However, arguably, these results do give us some grounds for claiming the existence of a positive longer-term effect for YTS. The increased employment probabilities induced by YTS naturally occur in the late post-YTS part of the 5-year history we have simulated. As a result, we can conclude that, conditional on observables and persistent unobservable characteristics, a greater proportion of individuals can be expected to reach age 21 in employment, if YTS has been available during the previous 5 years than would otherwise be the case. On the reasonable assumption of a relatively high degree of employment stability after age 21, this suggests a strong positive long-term effect of YTS on employment probabilities.

References

[1] Ackum, S. (1991). Youth Unemployment, Labour Market Programs and Subsequent Earnings, Scandinavian Journal of Economics, 93, 4, 531-543.

[2] Andrews, M. and S. Bradley (1997). Modelling the transition from school and the demand for training in the United Kingdom, Economica, 64, 387-413.

[3] Blau, D.M. and P.K. Robins (1987). Training programs and wages - a general equilibrium analysis of the effects of program size, Journal of Human Resources 22, 113-125.

[4] Deakin, B.M. and C.F. Pratten (1987). The economic effects of YTS, Employment Gazette, 95, 491-497.

[5] Burdett K., N.M. Kiefer and S. Sharma (1985). Layoffs and duration dependence in model of turnover, Journal of Econometrics, 28, 51-69.

[6] Dolton, P.J. (1993). The economics of youth training in Britain, Economic Journal 103, 1261-1278.

[7] Dolton, P.J., G.H. Makepeace and J.G. Treble (1994). The Youth Training Scheme and the school-to-work transition, Oxford Economic Papers, 46, 629-657.

[8] Goodman, L. (1961). Statistical methods for the mover-stayer model, Journal of the American Statistical Association, 56, 841-868.

[9] Gourieroux, C. and Monfort, A. (1996). Simulation-Based Econometric Methods, Oxford: Oxford University Press.

[10] Gritz, M. (1993). The impact of training on the frequency and duration of employment, Journal of Econometrics, 57, 21-51.

[11] Heckman, J. J. and Singer, B. (1984). A method for minimising the impact of distributional assumptions in econometric models for duration data, Econometrica, 52, 271-320.

[12] Johnson, G.E. and R. Layard (1986). The natural rate of unemployment: explanation and policy. In O. Ashenfelter and R. Layard (eds.) Handbook of Labour Economics, Amsterdam: North Holland.

[13] Jones, I. (1988). An evaluation of YTS, Oxford Review of Economic Policy, 4, 54-71.

[14] Katz L.F. and B.D. Meyer (1990). Unemployment insurance, recall expectations and unemployment outcomes, Quarterly Journal of Economics, 105, 973-1002.

[15] Lancaster, T. (1985). Generalised residuals and heterogeneous duration models: with applications to the Weibull model, Journal of Econometrics, 28, 113-126.

[16] Lancaster, T. (1990). The Econometric Analysis of Transition Data. Cambridge: Cambridge University Press.

[17] Lindeboom, M. and G. van den Berg (1994). Heterogeneity in models for bivariate survival: the importance of the mixing distribution, Journal of the Royal Statistical Society, series B (Methodological), 56, 49-60.

[18] Main, B.G.M. and M.A. Shelly (1988). The effectiveness of YTS as a manpower policy, Economica, 57, 495-514.

[19] Mealli, F. and S.E. Pudney (1996). Occupational pensions and job mobility in Britain: estimation of a random-effects competing risks model, Journal of Applied Econometrics, 11, 293-320.

[20] Mealli, F. and S.E. Pudney (1999). Specification tests for random-effects transition models: an application to a model of the British Youth Training Scheme, Lifetime Data Analysis, 5, forthcoming.

[21] Mealli, F., S.E. Pudney and J.M. Thomas (1996). Training duration and post-training outcomes: a duration limited competing risks model, Economic Journal, 106, 422-433.

[22] Meyer, B. (1990). Unemployment insurance and unemployment spells, Econometrica, 58, 757-782.

[23] Narendranathan W. and M.B. Stewart (1993). Modelling the probability of leaving unemployment: competing risks models with flexible baseline hazards, Journal of the Royal Statistical Society, series C (Applied Statistics), 42, 63-83.

[24] O'Higgins, N. (1994). YTS, employment and sample selection bias, Oxford Economic Papers, 46, 605-628.

[25] Organisation for Economic Co-operation and Development (1991). Evaluating Labour Market Programmes: the Current State of a Complex Art, Paris: OECD.

[26] Pudney S.E. (1981). An empirical method of approximating the separable structure of consumer preferences, Review of Economic Studies, 48, 561–577.

[27] Pudney S.E. (1989). Modelling Individual Choice: The Econometrics of Corners, Kinks and Holes. Oxford: Basil Blackwell.

[28] Pudney S.E. and J.M. Thomas (1995). Specification tests for the competing risks duration model: an application to unemployment duration and sectoral movement, Oxford Bulletin of Economics and Statistics, 57, 323–347.

[29] Ridder, G. (1987). The sensitivity of duration models to misspecified unobserved heterogeneity and duration dependence, mimeo, University of Amsterdam.

[30] van den Berg, G.J. (1997). Association measures for durations in bivariate hazard models, Journal of Econometrics (Annals of Econometrics), 79, 221–245.

[31] Weeks, M. (1995). Circumventing the curse of dimensionality in applied work using computer intensive methods, Economic Journal, 105, 520-530.

TABLE A1 Variables used in the models

Variable	Definition	Mean
Time-invariant characteristics (mean over all individuals)		
DOB	Date of birth (years after 1.1.60)	12.16
WHITE	Dummy = 1 if white; = 0 if other ethnic origin	0.942
GCSE2	Dummy for at least 1 General Certificate of Secondary Education (GCSE) pass at grade D or E	0.263
GCSE3	Dummy for 1-3 GCSE passes at grade C or better	0.185
GCSE4	Dummy for at least 4 GCSE passes at grade C or better	0.413
ILL	Dummy for the existence of a major health problem	0.012
SCHOOL	Measure of school quality = proportion of pupils with at least 5 GCSE passes in first published school league table	0.384
AREA	Measure of social background = proportion of homes in ward of residence that are owner-occupied	0.779
Spell-specific variables (mean over all episodes)		
DATE	Date of the start of spell (years since 1.1.88)	1.11
YTSYET	Dummy for existence of a spell of YTS prior to the current spell	0.229
YTSDUR	Total length of time spent on YTS prior to the current spell (years)	0.300
YTSLIM	Dummy = 1 if two-year limit on YTS was reached prior to the current spell	0.094
YTSMATCH	Dummy = 1 if current spell is in employment and thre was a previous YTS spell in the same industrial sector	0.121
CLERICAL	Dummy = 1 if current spell is in clerical employment	0.036
TECH	Dummy = 1 if current spell is in craft/technical employment	0.135
STN	Dummy = 1 for YTS spell and trainee with special training needs	0.013
CHOICE	Dummy = 1 if last or current YTS spell in desired industrial sector	0.136
URATE	Local rate of unemployment (at ward level)	0.103

TABLE A2 Sample transition frequencies (percent)

(a) Initial spell

State of origin	Destination state: C	E	U	YTS	Attrition	Incomplete	Marginal
C	–	7.7	12.3	4.8	4.9	70.3	47.4
E	3.0	–	17.8	15.4	0.9	62.9	10.6
U	9.1	27.2	–	57.6	5.5	0.7	28.2
YTS	1.6	73.9	11.9	–	10.1	2.5	13.8
Marginal	3.1	21.5	9.7	20.2	5.1	40.5	100

(b) All spells

State of origin	Destination state: C	E	U	YTS	Attrition	Incomplete	Marginal
C	–	8.1	13.4	4.9	4.8	68.8	25.1
E	0.7	–	13.2	5.8	0.4	79.9	30.5
U	5.3	37.8	–	41.2	13.3	2.4	24.5
YTS	1.3	70.6	16.5	–	8.2	3.4	19.9
Marginal	1.8	25.3	10.7	13.1	6.2	42.9	100

TABLE A3 Mean durations (years)

	C	E	U	YTS
Mean duration for completed spells	1.00	0.57	0.27	1.48
Mean elapsed duration for both complete and incomplete spells	2.95	2.17	0.31	1.52

TABLE A4 Estimates: initial state logit component (standard errors in parentheses)

Covariate	Destination state (relative to YTS)		
	C	E	U
Constant	1.575 (0.36)	0.321 (0.38)	3.835 (0.47)
WHITE	–2.580 (0.34)	–	–0.867 (0.37)
GCSE2	0.530 (0.16)	–	–
GCSE3	1.284 (0.20)	0.212 (0.22)	0.307 (0.16)
GCSE4	2.985 (0.20)	0.473 (0.24)	0.522 (0.17)
ILL	–	–2.311 (1.10)	–
SCHOOL	1.855 (0.31)	–	1.236 (0.35)
AREA	–	–	–1.818 (0.33)
URATE	–	–8.607 (2.57)	–12.071 (1.95)

TABLE A5(a) Estimates: transition component (standard errors in parentheses)

Coefficient (β_j)	Destination-specific transition intensities				
	C	E	U	O	YTS
Constant	−1.377 (1.85)	−5.924 (0.57)	−1.820 (0.71)	−6.926 (0.78)	−4.481 (1.11)
DATE	−8.090 (2.99)	–	–	0.795 (0.11)	−1.884 (0.14)
YTSYET	–	–	1.461 (0.43)	–	–
YTSDUR	–	0.762 (0.18)	−1.328 (0.46)	−0.198 (0.17)	–
YTSLIMIT	–	−2.568 (0.71)	−3.234 (0.75)	–	–
YTSMATCH	–	–	−0.610 (0.50)	–	–
CLERICAL	–	–	−0.865 (0.53)	–	–
STN	–	–	1.158 (0.41)	–	–
CHOICE	–	−0.335 (0.15)	–	–	–
WHITE	−1.919 (0.77)	1.433 (0.28)	−0.751 (0.32)	–	1.007 (0.29)
GCSE2	2.150 (0.61)	–	−0.666 (0.20)	–	0.437 (0.18)
GCSE3	2.369 (0.88)	−0.700 (0.17)	−1.233 (0.24)	−1.115 (0.33)	−1.036 (0.45)
GCSE4	3.406 (0.94)	−0.939 (0.18)	−2.046 (0.26)	−2.221 (0.32)	−1.642 (0.45)
ILL	–	–	−0.642 (0.39)	–	0.964 (0.75)
E	–	–	4.782 (0.59)	−0.962 (0.61)	3.469 (1.05)
U	5.530 (0.73)	6.654 (0.54)	–	5.079 (0.58)	6.066 (1.04)
YTS	–	3.558 (0.49)	2.853 (0.41)	–	–
U*(GCSE3/4)	−0.447 (0.72)	0.927 (0.22)	–	1.197 (0.36)	1.635 (0.45)
SCHOOL	–	0.233 (0.32)	−0.690 (0.47)	−1.389 (0.45)	1.451 (0.50)
AREA	1.512 (1.23)	–	−1.628 (0.51)	–	–
URATE	–	−3.231 (1.99)	−2.630 (2.62)	8.488 (3.32)	5.724 (3.20)
College $^{1}_{1}$			0.817 (0.16)		
College $^{1}_{2}$			1.516 (0.16)		

TABLE A5(b) Estimates: Burr shape parameters (standard errors in parentheses)

Origin	C	E	U	O	YTS
	Destination				
	®$_{ij}$				
C	–	1.341 (0.15)	1.852 (0.16)	1.636 (0.13)	1.167 (0.22)
E	0.356 (0.45)	–	1.528 (0.21)	..	1.190 (0.22)
U	2.667 (0.25)	1.601 (0.10)	–	..	1.722 (0.11)
YTS	0.592 (0.58)	1.427 (0.12)	1.100 (0.13)	..	–
	¾$_{ij}$				
C	–	2.494 (0.62)	1.171 (0.47)	0.555 (0.26)	5.547 (1.48)
E	0.414 (1.70)	–	4.083 (0.43)	..	5.465 (0.75)
U	2.429 (0.41)	1.652 (0.13)	–	..	1.508 (0.12)
YTS	5.569 (4.45)	1.018 (0.36)	1.315 (0.40)	..	–

TABLE A6 Estimates: YTS limit logit (standard errors in parentheses)

Parameter	Coefficients for state E
Constant	5.205 (1.25)
Heterogeneity (μ_E)	−1.493 (0.81)

TABLE A7 Estimates: coefficients of random effects and correlation parameter (standard errors in parentheses)

	State			
	C	E	U	YTS
Initial state logit (ψ_i)	0.075 (0.10)	1.106 (0.39)	0.160 (0.23)	0.521 (0.23)
Transition model (ω_i)	3.832 (0.33)	−0.248 (0.21)	−0.335 (0.13)	−0.586 (0.20)
Correlation parameter (λ)	−0.224 (0.04)			