

# State dependence in unemployment incidence:

## Evidence for British men revisited

Wiji Arulampalam<sup>+</sup>

November 2004

### Abstract

The issues of persistence in the observed labour market status of men are investigated using the British Household Panel Survey for the period 1991-97. The paper extends previous work in many directions. In particular, problems of endogenous initial conditions, and unobserved heterogeneity, are addressed within the context of different definitions of unemployment. In addition, allowance is also made to accommodate the 'stayer' phenomenon in the state of employment. All these were found to be very important in the estimation of the effect of scarring.

*JEL Classification:* C15, C23, C25, J64.

*Keywords:* dynamic binary panel models, unemployment, state dependence, unobserved heterogeneity, initial conditions.

+ Also IZA, Bonn, Germany.

*Address for Correspondence:* Professor Wiji Arulampalam, Department of Economics, University of Warwick, Coventry, CV4 7AL. Tel: +44 (0)24 76 523471; Fax: +44 (0)24 76 523032  
Email: [wiji.arulampalam@warwick.ac.uk](mailto:wiji.arulampalam@warwick.ac.uk)

The BHPS data used in this paper were collected by the ESRC Research Centre on Micro-Social Change at the University of Essex and made available through the ESRC Data Archive. I should like to thank Mark Taylor for providing help with the creation of some of the variables used in the study and to the Institute for Social and Economic Research Centre for providing the data on Travel-To-Work-Area unemployment and vacancy rates. I should also like to thank Alison Booth, Norman Ireland, Robin Naylor, Mark Stewart, Mark Taylor, and participants at the 8th International Conference on Panel Data held in Gotenborg in June 1998, Royal Economic Society Conference held in St. Andrews, Scotland in July 2000, and two anonymous referees for useful comments. This paper was produced as part of the project on *Unemployment and Technical and Structural Change*, which was funded by the Leverhulme Trust. The views in the paper are those of the author, and do not necessarily reflect those of the Leverhulme Trust. This paper is a revised version of 'State dependence in unemployment incidence: evidence for British men 1991-95' (1998). Any errors remain my responsibility.

## Abstract

The issues of persistence in the observed labour market status of men are investigated using the British Household Panel Survey for the period 1991-97. The paper extends previous work in many directions. In particular, problems of endogenous initial conditions, and unobserved heterogeneity, are addressed within the context of different definitions of unemployment. In addition, allowance is also made to accommodate the 'stayer' phenomenon in the state of employment. All these were found to be very important in the estimation of the effect of scarring.

## I. INTRODUCTION

The extent to which a previous unemployment experience increases the probability of an individual being unemployed in the future has important implications for labour market policies. Repeat incidences of unemployment can also lead to considerable poverty, social exclusion and distress.<sup>1</sup> If there is considerable persistence, short run policies such as job creation schemes and wage subsidies to employers, may be used to alter the equilibrium unemployment rate. Hence, not only an early successful intervention is required but the identification of the correct target groups also becomes necessary.

Correct identification of any ‘persistence’ in the experiences of unemployment requires one to first address the issue of unobservables. This is crucial since unobserved individual characteristics that are stable over time, when not correctly accounted for in the model specification and estimation, can result in the investigator wrongly concluding that there is a causal effect even when there is none. The basic model that has been used in order to disentangle the effects of unobservables and the causal effect coming via a previous spell of unemployment has been the dynamic random effects discrete choice model. An important issue that requires explicit consideration is the problem of how best to address the issue of endogeneity of ‘initial conditions’.

The initial conditions problem arises when the start of the observation period does not coincide with the start of the stochastic process generating individuals’ unemployment experiences. An individual who is observed in the state of unemployment at the start of the observation period may be there because of an earlier history of unemployment (state

---

<sup>1</sup> Arulampalam (2001) finds that there is a wage loss associated with post-unemployment job. Stewart (2002) finds that individuals who are unemployed go back to low-pay jobs, thus leading to additional income and welfare losses. Clark and Oswald (1994) find that unemployment causes a lot of distress and being unemployed is one of the most important causes of unhappiness.

dependence) or because of some unobserved characteristic affecting the job-offer or job retention rates facing that individual. In order to unravel these two effects, the initial condition needs to be explicitly modelled rather than assumed exogenously given. This paper looks at the sensitivity of the estimated results to different methods of addressing the initial conditions problem. ABT used a two-step technique proposed by Orme (1997) to address this problem. This technique is only valid if the problem caused by ‘endogeneity’ of the initial conditions is not severe. Following Heckman (1981), the model for the ‘initial’ observation is explicitly specified and estimated in this paper.

A recent response of the Government has been the introduction of the ‘New Deal’ program in 1998. One of the main aims of this program is to provide incentives to unemployed individuals on benefits to move to employment. Impositions of tighter conditions on benefit eligibility and job search are important parts of this program. But targeting such policies using very strict definition of ‘unemployment’ ignores a large group of unemployed individuals who may not satisfy these strict criteria but are also affected by repeat incidences of unemployment.

Although the basic model that has been used in order to estimate the persistence effect has been the dynamic random effects discrete choice model, the researchers have used various modifications to the specifications in order to address the common econometric problems of unobserved heterogeneity and initial conditions inherent in these models. This paper revisits the issue of persistence and explicitly looks at the dependence of estimated results to various model specification and estimation. In addition the paper also checks for sensitivity of the estimated persistence effect to the actual definition of ‘unemployment’ that is used.

The data used for this purpose is the longitudinal data on a group of men over the period 1991 to 1997 from the British Household Panel Survey (BHPS).<sup>2</sup> Although this is the same dataset that was used in Arulampalam, Booth, and Taylor (2000) (henceforth called ABT) this paper extends the work of ABT and others in several important ways.

Second specification issue concerns the distribution of the unobservables. The usual assumption of normally distributed unobservables is extended to accommodate the fact that there are many individuals who never experience a spell of unemployment. As shown below, the estimated magnitude of the persistence effect is indeed sensitive to this. The model estimated in this paper allows for this by allowing for empirically determined masses at the two extremes, i.e. plus and minus infinity of the Normal mixing distribution.

Next issue addressed concerns the definition of ‘unemployment’ used in the analysis. One of the most commonly used definitions of unemployment is unemployment with search. For example, one popular definition is that of the International Labour Organisation (ILO). Under this definition an individual is classified as being unemployed if (i) the individual does not have a job; (ii) s/he has searched for a job in the past four weeks; and (iii) s/he is available to start work immediately. This is the definition used by ABT. But sometimes researchers restrict the period of search to one week. But not all individuals who claim to be unemployed in the sample satisfy the search or the availability to start work criteria. Some of the individuals who do not satisfy these criteria are found to be in employment later on in the sample.<sup>3</sup> Looking at persistence in terms of a stricter definition of search unemployment may

---

<sup>2</sup> The cut-off of 1997 was used in order to avoid complications with regard to ‘persistence’ measure, arising from the introduction of the ‘New Deal’ program in 1998, which was aimed at individuals who had been unemployed for six months or more.

<sup>3</sup> Gregg and Wadsworth (1998) find that “those excluded from unemployment because they are not available to start in two weeks but are searching for work are more likely to enter employment than the long term unemployed. The data set used was the British Labour Force Survey for the period 1981 to 1997.

therefore be of little value for some policy purposes. For example, if one needs to look at the effect of wage pressures coming via possible increases to labour supply from these individuals who do not satisfy the standard search criterion, it is important to work with this group of individuals included in the analysis.

Another reason for looking at the impact of using different definitions for ‘unemployment’ is that most of the studies that look at persistence commonly use a two-state model of the labour market because of lack of observations to enable one to specify and estimate a general multi-state model.<sup>4</sup> Since a dynamic model of this type requires individuals to have continuous observations, an inevitable outcome of using a stricter definition of ‘unemployment’ is to discard individuals who do not satisfy the criteria. This may potentially bias the estimated effects of persistence. This study therefore looks at the issue of persistence using different classifications of the variable of interest. These are (i) self reported unemployment where no search criterion is used, (ii) search unemployment where the search period is four weeks (ILO definition), and (iii) search unemployment where the search period is one week.<sup>5</sup> These are important distinctions that result in various degrees of attrition in the sample used for the analysis with the last classification being the one with the least attrition.

The remainder of this paper is set out as follows. Section II presents the dynamic panel data model and discusses the various estimation issues. Section III discusses the data and the sample used. The estimates of ‘scarring’ are presented and discussed in Section IV. Final section summarises and concludes.

---

<sup>4</sup> Few examples of studies that have looked at the issue of persistence in unemployment are, Narendranathan, et al (1985), Flaig et al (1993), Muhleisen and Zimmermann (1994), Narendranathan and Elias (1995), McCulloch and Dex (1996) and, Arulampalam et al (2000).

<sup>5</sup> Because of small cell sizes, the analyses presented in this paper are conducted in a two-state model instead of a multi-state model of the labour market.

## II MODEL AND ESTIMATION

### II.1 Background

The extent to which a previous unemployment experience increases the probability of an individual being unemployed in the future has important implications for labour market policies. An individual may be stigmatised by long or repeated spells of unemployment that are beyond his control. Prospective employers may use the experience of unemployment as a signal of ‘unemployability’. It may also result in depreciation of acquired human capital. Hence he may receive fewer or no job offers the longer is his experience of unemployment or the greater the number of spells.

Of course, an unemployed individual may also become less choosy as unemployment spell lengthens. Consequently, he may revise downwards his reservation wage, and accept a poor quality job that is not expected to last long. For this reason, men who have experienced unemployment in one period may be more likely to be unemployed subsequently. While this study is unable to distinguish between these various competing hypotheses as to the causes of state dependence because of the reduced form nature of the model estimated, it is able to establish empirically whether genuine state dependence exists in this sample.

Consider the state of unemployment. The proportion of time spent by an individual in unemployment will depend on the probabilities of entry into and then exit from the state of unemployment. This requires some theory of labour turnover as a framework for the analysis. Since the study is concerned with reduced form models of unemployment incidence, only a few salient remarks are presented instead of a detailed discussion of various theories of labour turnover. First consider the determinants of the probability of entry into unemployment.

In a world where there are only two states, employment and unemployment, an individual can enter unemployment either voluntarily or involuntarily. An individual will

quit a job and enter unemployment if the value of outside prospects including the expected spell of unemployment exceeds the expected earnings in the current job. On the other hand, the probability of involuntary separation will depend on the individual's productivity relative to the wage.

As far as the determinants of the probability of exit out of unemployment is concerned, this will depend on the arrival rate of acceptable job offers which in turn will depend on how the individual searches for a suitable job and how the prospective employer views the individual as a suitable candidate.

Hence, the usual set of control variables such as demographic and family variables, level of education, and variables to proxy labour market tightness are included in the reduced form model of unemployment incidence. These variables reflect individual search intensity, and job-offer arrival or job-retention rates.

Not only observable but unobservable individual characteristics may also affect the propensity of certain individuals to be unemployed. Individuals may have undesirable attributes such as lack of motivation that, while unobservable to the statistician, may be used by employers to affect the rates of job-offers and job- retentions. A proper control of this is necessary in order to avoid biasing the estimated state dependence effect in the model of unemployment incidence.

## II.2 The Econometric Model

The observed dependent variable is binary, taking the value of one if the individual is unemployed at the time of the interview, and zero otherwise. This variable is observed, in the sample, at most seven consecutive separate interview dates. The model for individual  $i$  at the interview date at time  $t$  is specified as

$$y_{it}^* = \mathbf{x}_{it}'\boldsymbol{\beta} + \gamma y_{it-1} + \alpha_i + u_{it}, \quad i=1,2,\dots,n \text{ and } t=2,\dots,T_i \quad (1)$$



where  $y^*$  denotes the unobservable individual propensity to be unemployed,  $\mathbf{x}$  is a vector of exogenous observable characteristics that influence  $y^*$ ,  $\boldsymbol{\beta}$  is the vector of coefficients associated with  $\mathbf{x}$ ,  $\alpha_i$  denotes the individual-specific time invariant unobservable effect and  $u_{it}$  is assumed to be an innovation distributed as  $iidN(0, \sigma_u^2)$ . It is also assumed that  $\alpha_i \sim iidN(0, \sigma_\alpha^2)$  and is distributed independently of the  $u_{it}$  for all  $i$  and  $t$ .

An individual is observed to be unemployed when his propensity to be unemployed crosses a threshold (zero in this case), that is, if  $y_{it}^* > 0$  and  $= 0$  else. It is also assumed that the unobservable individual's propensity to be unemployed at time  $t$  is a function of the *observed* unemployment status of the individual in the previous period; that is, it is the actual *experience* of an unemployment spell, rather than the *propensity* to be unemployed, that affects the current incidence of unemployment. The inclusion of the lagged dependent variable on the right hand side of (1) allows one to test for the presence of state dependence.<sup>6</sup>

Before proceeding with the estimation of the model, some assumptions regarding the initial observation  $y_{i1}$  are required. As discussed earlier, the start of the observation period need not necessarily coincide with the start of the stochastic process generating the individuals' unemployment experiences. This is the case in the BHPS data set that is used. In general, the initial observation and the unobservable heterogeneity term will be correlated, leading to the so-called 'initial conditions' problem. To account for this problem, following Heckman (1981c), a reduced form equation for the initial observation is first specified as follows:

$$y_{i1}^* = \boldsymbol{\lambda}'\mathbf{z}_i + \eta_i \quad i=1, \dots, n \quad (2)$$

---

<sup>6</sup> For a survey of some of these models, see Hsiao (1986) and Maddala (1987).

where  $z_i$  is a vector of strictly exogenous instruments,  $var(\eta_i) = \sigma_\eta^2$  and  $corr(\alpha_i, \eta_i) = \rho$ . Here, the vector of variables in  $z$  will include variables that were relevant in period  $I$  along with some pre-sample information, which might affect the probability of unemployment in period  $I$ . Note, for identification purposes, one needs at least one variable in  $z$ , which is not in  $x$ .

The next step involves a linear specification, in terms of orthogonal error components, to account for the possibility of non-zero  $\rho$ :

$$\eta_i = \theta \alpha_i + u_{i1} \quad (3)$$

By construction,  $\alpha_i$  and  $u_{i1}$  in (3) will be orthogonal to one another,  $\theta = \rho\sigma_\eta/\sigma_\alpha$  and  $var(u_{i1}) = \sigma_\eta^2(1 - \rho^2)$ . It is also assumed that  $u_{it}$  satisfies the same distributional assumptions as  $u_{i1}$  for all  $i$  and  $t=2, \dots, T$ , and  $E(u_{it} y_{i1}) = 0$  for all  $i$  and  $t$ .

Hence, it follows from above discussions that the following two equations specify a complete model for the unemployment process,

$$y_{it}^* = \lambda' z_i + \theta \alpha_i + u_{it} \quad i=1, \dots, n \text{ and } t=1 \quad (4)$$

$$y_{it}^* = x_{it}' \beta + \gamma y_{it-1} + \alpha_i + u_{it} \quad i=1, \dots, n \text{ and } t=2, \dots, T_i \quad (1)$$

Note, in the above model,  $variance(u_{it}) = \sigma_u^2$  for  $t=2, \dots, T_i$  and  $variance(u_{i1}) = \sigma_\eta^2(1 - \rho^2)$ . As shown in Heckman (1981a, 1981b), under the additional assumption that  $\alpha_i$  and  $u_{it}$  are jointly multi-variate normal, this model can be easily estimated by noting that the distribution of  $y_{it}^*$  conditional on  $\alpha_i$ ,  $x_{it}$  and  $y_{it-1}$ , is independent normal. Marginalising the likelihood with respect to the  $\alpha$  gives the likelihood function for individual  $i$ ,

$$L_i = \int_{-\infty}^{\infty} \left\{ \prod_{t=2}^T \Phi \left[ \frac{x_{it}' \beta}{\sigma_u} + \frac{\gamma}{\sigma_u} y_{it-1} + \frac{\sigma_\alpha}{\sigma_u} \tilde{\alpha} \right] (2y_{it} - 1) \right\}$$

$$\Phi\left[\left(\mathbf{z}_i' \frac{\lambda}{\sigma_\eta \sqrt{1-\rho^2}} + \frac{\rho}{\sqrt{1-\rho^2}} \tilde{\alpha}\right) (2y_{it} - 1)\right] \phi(\tilde{\alpha}) d\tilde{\alpha} \quad (5)$$

where  $\tilde{\alpha}_i = \alpha_i / \sigma_\alpha$  and,  $\phi$  and  $\Phi$  are the density and the distribution function of the standard normal variate.

A normalisation is required next because of the binary nature of the variable under study. A convenient normalisation is that the variance of the error term  $u_{it} = 1$  for all  $t$ . Equation (5), thus becomes,

$$L_i = \int_{-\infty}^{\infty} \left( \prod_{t=2}^T \Phi[(\mathbf{x}_{it}' \boldsymbol{\beta} + \gamma y_{it-1} + \sigma_\alpha \tilde{\alpha}) (2y_{it} - 1)] \right. \\ \left. \Phi[(\mathbf{z}_i' \boldsymbol{\lambda} + \theta \sigma_\alpha \tilde{\alpha}) (2y_{it} - 1)] \right) \phi(\tilde{\alpha}) d\tilde{\alpha} \quad (6)$$

An obvious weakness in the above specification is the assumption of normally distributed unobserved heterogeneity. This assumption does not allow enough flexibility to model the phenomenon that some individuals are always observed to be in the same state in the sample (Narendranathan and Elias (1982)). A very large positive (negative) value for the unobservable  $\alpha$  will give a very large (small) value for  $y^*$  and hence a very large (small) probability of being in unemployment. This is accommodated by allowing for empirically determined masses at the two extremes, i.e. plus and minus infinity of the Normal mixing distribution. This gives the following likelihood for individual  $i$ ,

$$L_i^* = \frac{\Psi_0}{1 + \Psi_0 + \Psi_1} \left[ \prod_{t=1}^{T_i} (1 - y_{it}) \right] + \frac{\Psi_1}{1 + \Psi_0 + \Psi_1} \left[ \prod_{t=1}^{T_i} y_{it} \right] + \frac{L_i}{1 + \Psi_0 + \Psi_1} \quad (7)$$

where,  $L_i$  is given by equation (6) and  $\Psi_0$  and  $\Psi_1$  are the unknown end point parameters. Hence, the proportion of individuals who are assumed to have a very large or a very small unobserved component are given by  $p_0$  and  $p_1$ , where,

$$p_0 = \frac{\Psi_0}{1 + \Psi_0 + \Psi_1} \quad \text{and} \quad p_1 = \frac{\Psi_1}{1 + \Psi_0 + \Psi_1} \quad \text{respectively.} \quad (8)$$

**Some testable restrictions on the Model**

1. Exogeneity of the Initial Condition:  $\theta=0$  in (4) is a test of this hypothesis that the initial conditions can be treated as exogenous. Also note, under the assumption that the initial condition is exogenous, the above model reduces to a simple random effects probit model over  $t=2, \dots, T_i$ .

2. Unemployment process observed from the beginning: This is equivalent to a test of  $\theta=1$  in (4). This model can be estimated simply by creating a time dummy (*dum*) which equals one in  $t=1$  and zero otherwise. Equations (4) and (1) together now become

$$y_{it}^* = [\mathbf{x}_{it} * (1 - dum)]' \boldsymbol{\beta} + \gamma(1 - dum)y_{it-1} + (\mathbf{z}_i * dum)' \boldsymbol{\lambda} + \alpha_i + u_{it} \quad (9)$$

Equation (11) can then be estimated using all seven years of data with standard software packages, which allow estimation of random effects probit models.

3. No unobservable time invariant individual characteristics:

$$\text{Note } \text{corr}(v_{it}, v_{is}) = \text{corr}(\alpha_i + u_{it}, \alpha_i + u_{is}) = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_u^2} = r \text{ say, for all } t \neq s \neq 1. \quad (10)$$

Hence, a test of  $H_0 : \sigma_\alpha^2 = 0$  (which is a test that there are no unobservable characteristics in the sample and therefore the model collapses to a simple probit<sup>7</sup>) is equivalent to a test of  $H_0 : r = 0$  in equation (10). This can be tested as a likelihood ratio test but the test statistic will not be a standard chi-sq test since the parameter restriction is on the boundary of the parameter space. The standard likelihood ratio test statistic has a probability mass of 0.5 at zero and 0.5  $\chi^2(1)$  for positive values. Thus a one-sided 5% significance level test requires the use of the 10% critical value (Lawless (1987)).

**Interpretation of  $\gamma$**

A convenient way to interpret the estimated persistence effect  $\gamma$  is required. One such method is to look at the change in the predicted probabilities conditional on previous labour market status. This is the one provided in Chamberlain (1984). This is the standard marginal effect calculation in qualitative dependent variable models but accounts for the distribution of unobserved heterogeneity in the population. For each individual, predicted probabilities are calculated conditional, first on unemployment in the previous period and secondly on employment in the previous period. The difference between the first and the second averaged over the sample gives an estimate of state dependence or the persistence effect. This is the

---

<sup>7</sup> Comparisons of ordinary probit results with those from a random effects probit model need to account for different normalisations used by commercially available software (Arulampalam (1999)).

probability of observing a randomly chosen individual in unemployment in the current period conditional on previous labour market status.

In addition to the above measure of ‘scarring’, the ratio of the two conditional probabilities is also reported. This would tell us how large the conditional probability of being in unemployment this period conditional on previous unemployment relative to previous employment. As seen below, both measures are informative.

### **III THE DATA AND THE SAMPLE**

The data are from the first seven waves of the British Household Panel Survey (BHPS), a nationally representative survey of households randomly selected south of the Caledonian Canal.<sup>8,9</sup> The first wave of the BHPS was conducted from September to December 1991, and annually thereafter (see Taylor (1996) for details). The sample used includes (i) all men who were interviewed in 1991 (the first wave)<sup>10</sup>, (ii) aged 16 or over and under 55 at the initial wave, and (iii) had a non-missing labour market status and not in full-time education. Since the nature of the problem being analysed (lagged dependent variable is a regressor) requires one to have consecutive information on labour market, the sample was further restricted to only include men until they missed a direct interview. Then men were also required to be active in the labour market continuously to be included in the sample. The number of eligible sample members in each wave for this sample of 3024 men is provided in Table 1 Column [1].

---

<sup>8</sup> Thus the north of Scotland is excluded.

<sup>9</sup> Also see footnote 2.

<sup>10</sup> About 4% of the sample was lost because of either giving a proxy-interview or a telephone interview at each wave. Only direct interview sample were selected because of missing values for many variables of interest for individuals who did not give a direct interview.

Of the original 3024 men, 15.3% of men did not appear in 1992 wave, and a further 10% of the original sample was not in by the 1993 wave. By the end of the sample period, only 61% of the sample members had been present in all seven waves after applying the selection criteria as outlined above. Column [2] to [4] lists the number of individuals who contributed in each wave to the various analyses conducted. The two labour market statuses considered are, (i) employment including self-employment, and (ii) unemployment. The least restrictive assumption in terms of the definition used for the dependent variable is 'unemployment' and information for this is provided in Column [2]. Column [3] uses the standard ILO definition and thus excludes all those in Column [2] who said that they have not searched in the last four weeks. This is the classification of 'unemployment' that was used by ABT. Column [4] restricts the sample further to only consider individuals who had been searching in the past one week.

Looking across the columns of Table 1, it is interesting to note that the rate of attrition over waves is very similar across various definitions used in the analyses. From the start to the end year, approximately 40 to 45% of the original sample members are lost. The largest loss was from wave 1 to wave 2. About 18% of wave 1 men did not appear in wave 2 sample. Approximately another 10% did not appear in wave 3 sample. But rate of attrition has been less than 5% per wave since then. This rate of attrition is also remarkably close to the original 'interview' sample attrition that required the individuals to be present at consecutive interviews (plus other selections as discussed above).

Table 2 gives some information about the distribution of unemployment spells across waves for all men with valid data for variables used in the analysis. The average rate of unemployment over this seven-year period was 6.8%. Restricting the sample to those who satisfy the ILO definition of unemployment, the average rate of unemployment goes down to 4.8%. Applying a very strict definition of search over the last week reduces this even further

to 4%. Thus, the average unemployment rate amongst our sample members over this period is about 40% more than the average rate of 'search unemployment' in the sample. Unemployment rates are found to decline during the sample period with the first decline taking place in 1992 and another one in 1994.

Table 2 also gives the raw data conditional probabilities. There is a lot of persistence over this period as measured by the probability of being in 'unemployment' at the time of interview in a year given the same status in the previous interview period. The state dependence for the case of 'unemployment' is about 56 percentage points over the sample period with the probability of being in unemployment for an individual with previous unemployment (at last year's interview period) is 22 times higher compared to someone without this previous unemployment. Restricting the sample by imposing search criterion reduces the scarring effect to about 51 to 47 percentage points depending on how one defines search, and also increases the differential effect. The reason for this is that as one imposes the search criterion, the probability of unemployment for a previously unemployed individual decreases, but at the same time, this probability for a previously employed individual decreases by a larger amount.

The most interesting observation that emerges from this table is the fact that although the aggregate unemployment rate had been coming down over this period, the scarring effect as measured as the differences in the two conditional probabilities has not changed much. But what has happened is that the unemployment incidence among the previously employed people has been going down at a faster rate compared to previously unemployed people.

The interesting question is how much of the observed persistence in the raw data is due to observed characteristics, unobserved characteristics, and genuine state dependence. This is addressed through the multivariate analysis presented next.



#### IV. EMPIRICAL RESULTS

The models estimated are reduced form models of ‘unemployment’ incidences where three different definitions of the state of ‘unemployment’ are considered. These are, (i) unemployment with and without search, (ii) unemployment with search over the last four weeks (ILO definition), and (iii) unemployment with search over the past one-week. The last of these is the most restrictive definition with the first being the least restrictive.

The models include, as discussed previously, the usual set of control variables such as demographic and family variables, level of education, and variables to proxy labour market tightness. These variables reflect individual search intensity, job-offer arrival and job retention rates. The persistence effect is accounted for by the inclusion of the previous labour market status variable along with observable characteristics and, with the allowance for unobservable individual characteristics in the model.

The analysis presented does not enable one to distinguish between these various competing hypotheses as to the causes of state dependence. But it does allow one to establish empirically, whether state dependence exists in the sample used and to investigate the sensitivity of the results to different definitions of ‘unemployment’ used.

Definitions of the explanatory variables and the respective sample means are shown in Appendix Table A1. All the variables that come under the heading of ‘*Pre-sample Information*’ are the variables that were used to identify the initial condition model. Not surprisingly, as one starts to impose various search criterion restrictions on the selection, there are: (i) fewer local authority renters; (ii) more owner-occupiers with mortgage; (iii) fewer men with health problems and disability; (iv) more men with a degree or higher, and (v) fewer men with no qualifications.

Given the growing concern among the policy makers over the deteriorating labour market positions of lower educated young people, in all the models estimated, the persistence

effect was allowed to vary with age as well as qualifications in a very broad manner. One distinction is those men who are under 25 years of age (the age for eligibility for the ‘New Deal’ program) from those who are older than 25 and the other is those who have at least an ‘O’ Level qualification level (obtained at or around the minimum school-leaving age) from those who do not have.<sup>11</sup>

Before the discussion of the results, an important issue needs to be addressed. In the data used, successive interviews were carried out approximately one year apart. One possible concern that may arise with respect to the interpretation of the estimated persistence effects is to do with the possibility that the persistence effects that are estimated are picking up the effects coming via one continuous spell of unemployment instead of being two separate unemployment spells. It is therefore important to bear in mind the following. First, since unemployment is treated as a random variable, however far apart the consecutive observations are, one would always find some individuals to be in a continuous spell of unemployment. As long as the average unemployment duration is less than the time interval between two successive interviews, on average, one would expect most of the individuals who are observed to be in unemployment at two consecutive interviews, to be in two different spells and therefore not drive the results one obtains.<sup>12</sup> Second, although for some policy purposes it is important to find out whether an individual is experiencing a very long spell of unemployment or just very short duration repeat spells, both cases should be of concern to policy makers because of the associated welfare costs associated with unemployment. A

---

<sup>11</sup> The education variable was originally measured by four dummy variables, indicating the highest qualification attained by the individual by 1991. These variables were Degree (a university degree), Other-higher (other higher qualification equivalent to a degree), A-level (one or more advanced-level qualifications representing university entrance-level qualifications usually taken at or around the age of 18) and, O-level (one or more ordinary-level qualifications obtained at or around the minimum school-leaving age). Based on the estimated results, some of these were combined in the analyses presented in this paper.

<sup>12</sup> For an extensive treatment of this issue, see Arulampalam et al (2000).

different type of analyses will be required in order to disentangle the effects of duration dependence coming via long spells from those coming via repeat spells. Whatever the cause of scarring, scarring itself should be of concern.

Given the focus of the paper, only the main estimates of interest are provided in Table 3, and the detailed estimates in the Appendix Table A2. Turning now to Table 3, as discussed earlier, a more flexible distribution for unobserved heterogeneity was introduced to account for the fact that some individuals will never be observed to change status. In all of the models estimated, the predicted proportion of individuals who will always be found in unemployment (with or without search) was zero. As a result, the models were estimated with this restriction imposed. The raw data state dependence effects (that is the scarring effects) are provided in the first panel, followed by predicted state dependence effects for the models without end points and with end points respectively. The state dependence effects were calculated as discussed in the previous section, and have been averaged over the sample period and provided for the four sub-categories of men. These calculations give the probability of observing someone in ‘unemployment’ this period given ‘unemployment’ in the previous period, relative to someone who was employed in the previous period. The ratios of these two conditional probabilities are also reported.

First consider the state dependence or the persistence observed in the raw data. The estimated scarring effect of previous period’s unemployment is broadly similar across the three samples in the raw data. As far as the raw data is concerned, it does not seem to matter how unemployment is defined except in the case of men aged less than 25 without any educational qualifications. Generally, young and the old men with some qualification exhibit similar persistence probabilities in the raw data. The same applies to those without qualifications too. For example, taking the full unemployment sample over this period, it is estimated that a man (regardless of his age) with some educational qualification would

experience a scarring effect of around 40 percentage points. This figure increases to about 60 percentage points if he does not possess any qualifications. But an interesting picture emerges from the ratio calculations. The ratio of the two conditional probabilities are much lower for men less than 25 compared to men older than 25. In addition the relative probability changes are higher for men with educational qualifications than for those without any qualifications. The reason for this is as follows. Unemployment among the young is generally higher compared to older men. In addition, 'job shopping behaviour' when one is young is generally expected to be acceptable. But at the same time employers generally do not expect educated men to be unemployed and when they actually do then it sends out the wrong signal. Hence the scarring effect in terms of the ratio will be higher for educated men relative to non-educated, and also higher for older men compared to younger men.

It is important to control for possible observed and unobserved characteristics that may affect the probability of someone becoming unemployed. Attention is drawn next to the bottom two panels where estimated state dependence effects are presented for models with and without end-points. Inclusion of end points does have an effect on the estimated scarring effect. Marginal effects are higher but the ratio measures are slightly smaller in models with end points. The reason for this is that when end points are included the predicted conditional probabilities are higher compared to the models without the end points. But also at the same time, the models with end points seem to predict a relatively higher probability of unemployment for previously unemployed people compared to previously employed people. In addition, the proportion of variance attributed to unobserved heterogeneity in the total error variance is also reduced when end points are included. In spite of this, the predicted mass at the extreme was not significantly different from zero implying that the normally distributed unobserved heterogeneity was sufficient to capture the variations. Likelihood ratio statistic for testing the null of zero heterogeneity variance was also rejected in all the models.

As expected, controlling for observable and unobservable characteristics, reduce the scarring effect. Whether one uses a ‘one-week’ or ‘four-week’ search criterion for the definition of unemployment seems to produce broadly similar results. Young people are still found to be less scarred, relative to older men by their experience of unemployment in terms of ratio measure. Interestingly, significant differences in the ratio of estimated probabilities between those who have some educational qualification and those who do not in the models with end points are only found when a very broad definition of unemployment is used. In all of the other models, the ratio of the estimated conditional probabilities is not dependent on educational qualification.

Table 4 provides some information about what percentage of the raw persistence effects are being attributed to genuine state dependence when allowance is made for observable and unobservable characteristics of the individuals. Allowing for end points is found to explain more of the observed persistence in the raw data. The model with end-points attributes a massive 80% of the observed persistence to genuine scarring for older educated men in the unemployment sample. That is, once they experience an unemployment spell, they are really scarred by their experience. In contrast, in the search unemployment sample, a larger proportion of observed persistence is coming via genuine state dependence for the young men compared to the older men.

As discussed earlier, various hypotheses may be tested regarding the initial observation using the estimated coefficient of  $\theta$ . The model with  $\theta=0$  implies that the initial condition is exogenous and  $\theta=1$  implies that the model specification allows us to treat the process as if it had been observed from the beginning.<sup>13</sup> The estimated  $\theta$  and the t-ratio for

---

<sup>13</sup> The same unobservable component enters both the first observation equation and the other equations.

the testing of  $\theta=0$  and  $\theta=1$  null are also given in Table 3. In the absence of end points, the null of  $\theta=0$  is rejected in all the models implying the importance of allowing for the endogeneity of initial conditions in these models. But interestingly, when end points are included in the model, the null of exogenous initial conditions is rejected in the models which use search criterion in the definition of unemployment.

The null of  $\theta=1$  is rejected in all the models estimated regardless of whether end points are included or not except in the model which uses end points and the ILO definition of unemployment.

## V. CONCLUSIONS

This paper extended the results of some of the previously published work on unemployment persistence for British men, using the British Household Panel Survey for the period 1991-97, in several important ways. First, the issue of how the estimated persistence effects varied according to the definitions of unemployment used was considered. The definitions considered were, (i) unemployment with and without search, (ii) unemployment with search over the last four weeks (ILO definition), and (iii) unemployment with search over the past one-week. The last of these is the most restrictive definition with the first being the least restrictive. Next, endogeneity of the initial observation, and unobserved heterogeneity were explicitly modelled. More importantly, the common models that are routinely estimated in the literature on persistence was extended to accommodate the phenomenon that some individuals never become unemployed, by including two empirically determined mass points at the two extremes of a continuous distribution for the unobserved heterogeneity.

Given the growing concern over the deteriorating labour market positions of less educated people, four broad groups were identified in order to allow for different persistence effects. These were, men aged less than 25 with no education, men aged less than 25 with

some qualification, men aged 25 or more with no education, and men aged 25 with some qualification.

The results confirm the earlier finding that strong state dependence effects do exist with respect to previous unemployment. This finding is consistent with the “scarring” theory of unemployment - an individual’s previous unemployment experience has implications for his future labour market behaviour, perhaps because of depreciation of human capital, or because employers use an individual’s previous labour market history as a screening device about his productivity. In addition, this study found the scarring effect to be sensitive to the definition of unemployment used. The main findings are as follows:

- Although the aggregate unemployment rate had been generally coming down over the period considered, the scarring effect of a spell of unemployment in the raw data had not changed much. But, unemployment incidence among the previously employed people had been going down at a faster rate compared to previously unemployed people.
- In general the estimated scarring effects over the years in the raw data were broadly similar across the different definitions of unemployment used. But interestingly, the ratio of the two conditional probabilities was much higher in the case of men older than 25 compared to men aged less than 25. This is consistent with the view that although the incidence of unemployment is generally higher among the younger men relative to older men, the younger men are less scarred by their experience in terms of relative probabilities since ‘job shopping behaviour’ is expected to be acceptable when one is young.
- Inclusion of end points in the distribution of unobserved heterogeneity was found to be very important. The estimated scarring effects were not only higher compared to models without end points but the proportion in the raw data attributed to genuine state

dependence was also higher. Including end points also reduced the variance of the unobserved heterogeneity component relative to the total error variance.

- The exogeneity of ‘initial conditions’ was easily rejected in most of the models estimated.
- The proportion of raw data persistence attributed to genuine scarring was found to be sensitive to the definition of the sample used in the case of men aged less than 25 without any educational qualifications, and men aged more than 25 with some educational qualifications. A massive 81% of the observed persistence was estimated to come from genuine scarring for older men with qualifications when the full unemployment sample was used. The figure for younger men without qualification was the lowest in this sample (29%). For the other two groups of men, the proportion attributed to scarring was broadly similar across the different samples.



## REFERENCES

- Arulampalam, W. (1999) “A Note on estimated effects in random effect probit models”, Oxford Bulletin of Economics and Statistics, 61(4), 597-602.
- Arulampalam, W. (2001) “Is unemployment really scarring? Effects of unemployment experience on wages”, Economic Journal, 111(475), 585-606.
- Arulampalam, W., Booth, A.L and Taylor, M. P. (2000) “Unemployment persistence”, Oxford Economic Papers, 52, 24-50.
- Chamberlain, G. (1984) “Panel data”, in S. Griliches and M. Intriligator, eds., Handbook of Econometrics, North-Holland, Amsterdam, 1247-1318.
- Clark, A.E. and Oswald, A. J. (1994) “Unhappiness and unemployment”, Economic Journal, 104, 648-59.
- Flaig, G., Licht, G. and Steiner, V. (1993) “Testing for state dependence effects in a dynamic model of male unemployment behaviour”, in H. Bunzel, P. Jensen and N. Westergaard-Nielsen (eds), Panel Data and Labour Market Dynamics, North Holland, Amsterdam.
- Gregg, P. and Wadsworth, J. (1998) “Unemployment and non-employment: unpacking economic inactivity”, Employment Policy Institute Economic Report, 12 (6), June.
- Heckman, J. J. (1981a) - “Statistical models for discrete panel data”, in Structural Analysis of Discrete Data with Econometric Applications, ed. C. F. Manski and D. McFadden, 114-178, Cambridge: MIT Press.
- Heckman, J. J. (1981b) “Heterogeneity and state dependence”, in Studies in Labor Markets, ed. S. Rosen, Chicago, Chicago Press.

- Heckman, J. J. (1981c) "The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process", in Structural Analysis of Discrete Data with Econometric Applications, eds. C. F. Manski and D. McFadden, 114-178, Cambridge: MIT Press.
- Hsiao, C. (1986) Analysis of Panel Data, Econometric Society Monograph, Cambridge University Press.
- Lawless, J. F. (1987) "Negative binomial and mixed Poisson regression", Canadian Journal of Statistics, 15, 209-25.
- Maddala, G. S. (1987) "Limited dependent variable models using panel data", Journal of Human Resources, XX11,3, 307-337.
- McCulloch and S Dex (1996) "Modelling male unemployment persistence in Britain", mimeo, University of Cambridge.
- Mühleisen, M. and Zimmermann, K. F. (1994) "New patterns of labour mobility: A panel analysis of job changes and unemployment", European Economic Review, 38, 793-801.
- Narendranathan, W. and Elias, P. (1993) "Influences of past history on the incidence of youth unemployment: Empirical findings for the UK", Oxford Bulletin of Economics and Statistics, 55 (2), 161-185.
- Narendranathan W, Nickell, S. J. and Metcalf, D. (1985) "An investigation into the incidence and dynamic structure of sickness and unemployment in Britain, 1965-75" Journal of the Royal Statistical Society, Series A, 148(3), 254-267.
- Narendranathan, W. and Stewart, M. B. (1993a), "Modelling the probability of leaving unemployment: competing risks models with flexible base-line hazards" Journal of the Royal Statistical Society, Series C, Applied Statistics, 42(1), 63-83.
- Narendranathan, W. and Stewart, M. B. (1993b) "How does the benefit effect vary as unemployment spells lengthen?", Journal of Applied Econometrics, 8(4), 361-382.

- Orme, C. D. (1997) “The initial conditions problem and two-step estimation in discrete panel data models”, mimeo, University of Manchester.
- Stewart, M. B. (2002) “The inter-related dynamics of unemployment and low pay”, mimeo, University of Warwick.
- Taylor, M. F. (1996) (ed. with J Brice et al) British Household Panel Survey User Manual Volumes A and B, Wivenhoe Park: University of Essex.

Table 1: Number of individuals in the Interview and Estimation Samples  
[Cumulative percentage lost from wave to wave]  
Waves 1 – 7 (1991 – 1997)

Wave	Interview Sample [1]	Employment/ Unemployment [2]	Employment/ Search (4weeks) Unemployment [3]	Employment/ Search (1 week) Unemployment [4]
1991	3024	2772	2675	2619
1992	2560 [15.3]	2286 [17.5]	2192 [18.1]	2139 [18.3]
1993	2268 [25.0]	2003 [27.7]	1919 [28.3]	1882 [28.1]
1994	2114 [30.1]	1839 [33.7]	1761 [34.2]	1727 [34.1]
1995	1967 [35.0]	1700 [38.7]	1626 [39.2]	1599 [39.1]
1996	1905 [37.0]	1613 [41.8]	1542 [42.4]	1518 [42.0]
1997	1840 [39.2]	1549 [44.1]	1472 [45.0]	1449 [44.7]

**Table 2: Incidence and Persistence in the Raw Data (1991 – 1997)**

	Wave 1 (1991)	Wave 2 (1992)	Wave 3 (1993)	Wave 4 (1994)	Wave 5 (1995)	Wave 6 (1996)	Wave 7 (1997)	Total
<b>Employment vs Unemployment</b>								
Prob(unemp <sub>t</sub> )	0.093	0.083	0.074	0.069	0.047	0.048	0.035	0.068
Prob(unemp <sub>t</sub>  unemp <sub>t-1</sub> )		0.590	0.602	0.584	0.520	0.644	0.563	0.584
Prob(unemp <sub>t</sub>  emp <sub>t-1</sub> )		0.039	0.036	0.031	0.017	0.020	0.012	0.027
State dependence - Difference		0.551	0.566	0.553	0.503	0.624	0.541	0.557
Ratio		15	17	19	31	32	47	22
<b>Employment vs Search (past 4 weeks) Unemployment</b>								
Prob(unemp <sub>t</sub> )	0.076	0.059	0.049	0.050	0.033	0.028	0.018	0.048
Prob(unemp <sub>t</sub>  unemp <sub>t-1</sub> )		0.507	0.580	0.610	0.471	0.537	0.400	0.529
Prob(unemp <sub>t</sub>  emp <sub>t-1</sub> )		0.029	0.024	0.023	0.013	0.015	0.008	0.019
State dependence Difference		0.478	0.456	0.587	0.458	0.522	0.392	0.510
Ratio		18	24	27	36	36	50	28
<b>Employment vs Search (past 1 week) Unemployment</b>								
Prob(unemp <sub>t</sub> )	0.062	0.047	0.038	0.041	0.026	0.022	0.016	0.039
Prob(unemp <sub>t</sub>  unemp <sub>t-1</sub> )		0.456	0.521	0.597	0.411	0.516	0.385	0.487
Prob(unemp <sub>t</sub>  emp <sub>t-1</sub> )		0.026	0.019	0.020	0.012	0.011	0.008	0.017
State dependence Difference		0.430	0.502	0.577	0.399	0.505	0.377	0.470
Ratio		18	28	30	34	47	48	29

Notes: (i) State dependence: difference = Prob(unemp<sub>t</sub>|unemp<sub>t-1</sub>) - Prob(unemp<sub>t</sub>|emp<sub>t-1</sub>).  
(ii) State dependence: Ratio = Prob(unemp<sub>t</sub>|unemp<sub>t-1</sub>) / Prob(unemp<sub>t</sub>|emp<sub>t-1</sub>).

**Table 3: Estimated State Dependence (Persistence): Difference [Ratio]**

	Unemployment Sample [1]	Unemployment with Search (4 weeks) Sample [2]	Unemployment with Search (1 week) Sample [3]
<b>RAW DATA</b> <u>State Dependence Effect</u>			
Age < 25 & has some educational qualifications	0.415 [ 9]	0.409 [13]	0.389 [14]
- has NO educational qualifications	0.583 [ 5]	0.634 [ 7]	0.433 [ 6]
Age >=25 & has some educational qualifications	0.401 [22]	0.366 [27]	0.370 [30]
- has NO educational qualifications	0.696 [20]	0.646 [25]	0.642 [31]
<b>MODELS WITHOUT END POINTS</b> <u>Estimated State Dependence Effect:</u>			
Age < 25 & has some educational qualifications	0.182 [ 3]	0.146 [ 3]	0.132 [ 3]
- has NO educational qualifications	0.130 [ 7]	0.326 [ 4]	0.150 [ 3]
Age >=25 & has some educational qualifications	0.322 [ 3]	0.081 [ 6]	0.080 [ 6]
- has NO educational qualifications	0.355 [ 7]	0.239 [ 7]	0.236 [ 8]
$\theta$ - see text (t-ratio for $\theta=0$ ) [t-ratio for $\theta=1$ ]	1.015 (4.51)[0.07]	0.860 (4.39)[0.71]	0.819 (4.08)[0.90]
Proportion of variance of unobserved heterogeneity in the total unexplained variance in periods $t=2, \dots, T_i$ . [r in eq (10)] (standard error)	0.294 (0.060)	0.382 (0.074)	0.396 (0.082)
Likelihood Ratio Test for $\sigma_\alpha^2=0$ [p-value – see text for details]	33.62 [0.00]	33.42 [0.00]	28.62 [0.00]
<b>MODELS WITH END POINTS</b> <u>Estimated State Dependence Effect</u>			
Age < 25 & has some educational qualification	0.209 [ 3]	0.190 [ 3]	0.183 [ 3]
- has NO educational qualifications	0.167 [ 6]	0.341 [ 3]	0.191 [ 2]
Age >=25 & has some educational qualification	0.326 [ 3]	0.088 [ 5]	0.138 [ 6]
- has NO educational qualifications	0.404 [ 6]	0.265 [ 6]	0.349 [ 6]
$\theta$ - see text (t-ratio for $\theta=0$ ) [t-ratio for $\theta=1$ ]	0.841 (3.03)[0.57]	0.265 (1.36)[3.77]	0.514 (1.56)[1.60]
$\psi_0$ (standard error)	0.366 (0.173)	0.101 (0.178)	0.913 (0.389)
Predicted proportion of always employed – $P_0$ (std. Error) <sup>iii</sup>	0.268 (0.210)	0.092 (0.511)	0.478 (0.213)
Proportion of variance of unobserved heterogeneity in the total unexplained variance in periods $t=2, \dots, T_i$ . [r in eq (10)] (standard error)	0.214 (0.065)	0.435 (0.062)	0.247 (0.091)
Likelihood Ratio Test for $\sigma_\alpha^2=0$ [p-value – see text for details]	37.71 [0.00]	96.51 [0.00]	33.31 [0.00]

Notes: (i) The above effects are averaged over waves 1 to 7. (ii) With educational qualifications mean that the highest qualification the person has at least an ‘O’ Level. (iii) Standard Error for  $P_0$  calculated using the delta method.

**Table 4: Estimated State Dependence as a Percentage of Raw Data Persistence**

	<b>Unemployment Sample</b>	<b>Unemployment with Search (4 weeks) Sample</b>	<b>Unemployment with Search (1 week) Sample</b>
	<b>[1]</b>	<b>[2]</b>	<b>[3]</b>
<b>MODELS WITHOUT END POINTS</b>			
<u>Estimated State Dependence Effect</u>			
Age < 25 & has some educational qualifications	44	36	34
- has NO educational qualifications	22	51	35
Age >=25 & has some educational qualifications	80	22	22
- has NO educational qualifications	47	37	37
<b>MODELS WITH END POINTS</b>			
<u>Estimated State Dependence Effect</u>			
Age < 25 & has some educational qualifications	50	47	47
- has NO educational qualifications	29	54	44
Age >=25 & has some educational qualifications	81	24	37
- has NO educational qualifications	58	41	54

Notes: (i) The above percentages are calculated on the basis of the marginal effect calculations provided in Table 3.

**Appendix Table A1: Definitions of variables, Means (Std. Deviation)**

<b>Variable Name</b>	<b>Definition</b>	<b>Employment/ Unemployment Sample</b>	<b>Employment/ Search (4 weeks) Sample</b>	<b>Employment/ Search (1 week) Sample</b>
<i>Number of Observations</i>	Person-year observations used in the estimation	13785	13201	12945
<i>Demographics</i>				
Aged less than 25 – Base case	Aged less than 25 at date of interview.	0.097	0.090	0.088
Aged 25-34	Aged between 25 and 34 at date of interview.	0.291	0.291	0.290
Aged 35-44	Aged between 35 and 44 at date of interview.	0.303	0.307	0.310
Aged 45+	Aged 45 or over at date of interview.	0.309	0.312	0.312
Owner-occupier – no mortgage	Own property outright at wave t.	0.112	0.111	0.112
Owner-occupier – has a mortgage	Has a mortgage on the property at wave t.	0.694	0.705	0.711
Local authority renter	Rents property from local authority at wave t.	0.108	0.099	0.092
Base case – private renters	Rents property privately at wave t.	0.086	0.085	0.085
Health limits work	Health limits type or amount of work at date of interview.	0.064	0.062	0.060
Registered Disabled	Registered disabled at wave t.	0.010	0.010	0.010
<i>Highest qualification</i>				
Degree or above	Holds a University first or higher degree, a teaching, nursing or other higher qualifications (eg. technical, professional qualifications) at wave 1.	0.399	0.406	0.410
A Levels	Holds one or more Advanced level qualifications (or equivalent) representing university entrance-level qualification typically taken at age 18 at wave 1.	0.132	0.133	0.135
O Levels	Holds one or more Ordinary level qualifications (or equivalent) taken at age 16 at end of compulsory schooling at wave 1. Selection mechanism into A level courses.	0.198	0.200	0.201
Base case – none of the above	Less than the above or no qualifications	0.271	0.261	0.254
Apprenticeship completed	Has completed a trade apprenticeship at wave 1.	0.025	0.024	0.023
<i>Family</i>				
Married	Married or cohabiting at wave t	0.767	0.777	0.778
Number of children aged under 5	Number of children under 5 at wave t	0.214	0.212	0.208



**Appendix Table A1: Definitions of variables, Means (Std. Deviation) - Continued**

<b>Variable Name</b>	<b>Definition</b>	<b>Employment/ Unemployment Sample</b>	<b>Employment/ Search (4 weeks) Sample</b>	<b>Employment/ Search (1 week) Sample</b>
<i>Demand side factors</i> Local U/V ratio	Unemployment stock divided by the vacancy stock in travel-to-work area of residence at wave t	16.82 (12.3)	16.74 (12.3)	16.70 (12.3)
<i>Presample information</i>				
Father –professional managerial	Father was in a professional-managerial job when the individual was aged 14	0.252	0.255	0.258
Father - skilled manual	Father was in a skilled manual job when the individual was aged 14	0.087	0.088	0.089
Father – unskilled manual	Father was in a semi or unskilled manual job when the individual was aged 14	0.156	0.154	0.153
Father – self employed	Father was self-employed when the individual was aged 14	0.146	0.144	0.145
Base Case -	Father not-employed or Father deceased when individual was 14	0.359	0.359	0.355
Mother –professional managerial	Mother was in a professional-managerial job when the individual was aged 14	0.105	0.105	0.107
Mother - skilled manual	Mother was in a skilled manual job when the individual was aged 14	0.155	0.156	0.157
Mother – unskilled manual	Mother was in a semi or unskilled manual job when the individual was aged 14	0.170	0.170	0.169
Mother – self employed	Mother was self-employed when the individual was aged 14	0.039	0.040	0.040
Base Case -	Mother not-employed or Father deceased when individual was 14	0.531	0.529	0.527

**Appendix Table A2: Random Effects Probit Model Coefficient Estimates (standard errors)  
(Normal Heterogeneity with End Points)**

Variable	Unemployment Sample [1]		Unemployment with Search (4 weeks) Sample [2]		Unemployment with Search (1 week) Sample [3]	
	1992 - 1997	1991	1992 - 1997	1991	1992 - 1997	1991
	<i>Initial condition</i>		<i>Initial condition</i>		<i>Initial condition</i>	
Intercept	-1.094 (0.190)**	-0.529 (0.235)**	-1.485 (0.228)**	-0.787 (0.261)**	-1.083 (0.288)**	-0.401 (0.314)
<i>State dependence</i>						
at t-1*has educ qual * age <25	0.869 (0.215)**		1.085 (0.269)**		0.831 (0.288)**	
at t-1*has NO educ qual * age <25	1.056 (0.229)**		1.416 (0.299)**		0.714 (0.368)**	
at t-1*has educ qual * age >= 25	1.200 (0.126)**		1.056 (0.155)**		1.109 (0.164)**	
at t-1*has NO educ qual *age >= 25	1.656 (0.140)**		1.589 (0.184)**		1.582 (0.200)**	
<i>Demand side factors</i>						
Local U/V ratio /100	0.012 (0.002)**	0.009 (0.004)**	0.011 (0.003)**	0.004 (0.004)	0.013 (0.003)**	0.003 (0.004)
<i>Demographics</i>						
Aged 25-34	-0.537 (0.120)**	-0.385 (0.128)**	-0.595 (0.148)**	-0.550 (0.139)**	-0.645 (0.167)**	-0.524 (0.155)**
Aged 35-44	-0.571 (0.128)**	-0.595 (0.146)**	-0.647 (0.157)**	-0.824 (0.162)**	-0.632 (0.176)**	-0.751 (0.177)**
Aged 45-	-0.430 (0.129)**	-0.500 (0.159)**	-0.555 (0.159)**	-0.692 (0.175)**	-0.528 (0.178)**	-0.593 (0.188)**
Owner-occupier - no mortgage	-0.275 (0.136)**	-0.390 (0.194)**	-0.547 (0.166)**	-0.405 (0.210)**	-0.591 (0.182)**	-0.439 (0.235)**
Owner-occupier - with mortgage	-0.275 (0.106)**	-0.344 (0.142)**	-0.488 (0.120)**	-0.379 (0.155)**	-0.573 (0.142)**	-0.315 (0.169)**
Local authority renter	0.468 (0.123)**	0.659 (0.163)**	0.291 (0.137)**	0.641 (0.175)**	0.187 (0.161)**	0.655 (0.201)**
Health limits work	0.528 (0.100)**	0.417 (0.169)**	0.588 (0.121)**	0.304 (0.188)	0.527 (0.137)**	0.317 (0.214)
Registered disabled	-0.070 (0.274)	0.618 (0.393)	-0.287 (0.353)	0.802 (0.400)**	-0.155 (0.371)	0.680 (0.485)
<i>Highest qualification</i>						
Degree or above	-0.380 (0.093)**	-0.478 (0.126)**	-0.443 (0.117)**	-0.375 (0.142)**	-0.380 (0.127)**	-0.305 (0.150)**
'A' Levels	-0.289 (0.115)**	-0.542 (0.157)**	-0.255 (0.140)**	-0.375 (0.171)**	-0.167 (0.150)	-0.402 (0.185)**
- 'O' Levels	-0.314 (0.104)**	-0.216 (0.123)**	-0.266 (0.124)**	-0.058 (0.134)	-0.222 (0.137)	-0.047 (0.148)
Completed an Apprenticeship	0.001 (0.194)	0.149 (0.259)	-0.021 (0.246)	0.203 (0.297)	-0.051 (0.274)	0.184 (0.318)
<i>Family</i>						
Married or living as married	-0.319 (0.082)**	-0.404 (0.116)**	-0.103 (0.103)	-0.349 (0.129)**	-0.030 (0.114)	-0.325 (0.140)**
Number of children aged under 5	0.133 (0.062)**	0.269 (0.089)**	0.091 (0.078)	0.241 (0.101)**	0.120 (0.084)	0.224 (0.110)**

**Appendix Table A2: Continued**

Variable	Unemployment Sample		Unemployment with Search (4 weeks) Sample		Unemployment with Search (1 week) Sample	
	[1]		[2]		[3]	
	1992 - 1997	1991 <i>Initial condition</i>	1992 - 1997	1991 <i>Initial condition</i>	1992 - 1997	1991 <i>Initial condition</i>
<i>Pre-sample Information</i>						
Father's occupation at age 14						
- Professional & managerial		-0.149 (0.127)		-0.288 (0.144)**		-0.223 (0.152)
- Skilled manual		0.401 (0.161)**		0.305 (0.178)*		0.413 (0.186)**
- Semi or unskilled manual		0.054 (0.125)		0.015 (0.141)		0.025 (0.154)
- Self employed		-0.193 (0.136)		-0.164 (0.151)		-0.063 (0.158)
Mother's occupation at age 14						
- Professional & managerial		-0.343 (0.177)*		-0.295 (0.193)		-0.135 (0.197)
- Skilled manual		-0.136 (0.137)		-0.189 (0.153)		-0.127 (0.164)
- Semi or unskilled manual		-0.206 (0.122)*		-0.239 (0.136)*		-0.223 (0.150)
- Self employed		0.604 (0.222)*		0.780 (0.227)**		0.705 (0.257)**
$\theta$ – see text		0.841 (0.278)**		0.265 (0.195)		0.514 (0.303)*
$r$ = Proportion of variance of unobserved heterogeneity in the total unexplained variance in periods $t=2, \dots, T_i$		0.214 (0.065)**		0.435 (0.062)**		0.247 (0.091)**
$\Psi_0$		0.366 (0.173)**		0.101 (0.178)		0.913 (0.389)**
$\Psi_1$		0.000		0.000		0.000
$P_0$		0.268		0.092		0.478
$P_1$		0.000		0.000		0.000
Model Log-likelihood		-2298.01		-1739.64		-1534.97
Model Log-likelihood with $r = 0$		-2316.87		-1787.89		-1551.63
Number of observations		13785		13201		12945
Number of individuals		2772		2675		2619

Notes: (i) Standard errors in brackets. \*\*, \* denote significance at 5% and 10% respectively.

(ii)  $\psi_1$  was empirically determined to be always zero in all of the above specifications.  $P_0$  and  $P_1$  are the predicted proportion of stayers in state 0 (employment) and state 1 ('unemployment') respectively.