

Unemployment Persistence: Theoretical and Empirical Developments

by

Stephen Knights

A thesis submitted in satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Economics

in the

SOCIAL SCIENCES DIVISION
of the
UNIVERSITY OF OXFORD

Balliol College, Hilary Term 2011

Approximate Word Count: 70,000

University of Oxford

Balliol College, Hilary Term 2011

Unemployment Persistence: Theoretical and Empirical Developments

Copyright 2011

by

Stephen Knights

Abstract

Unemployment Persistence: Theoretical and Empirical Developments

by

Stephen Knights

Thesis for Doctor of Philosophy in Economics

University of Oxford

This thesis presents three chapters on the subject of unemployment persistence. Two of the chapters are empirically focussed and the other is a purely theoretic work. Unemployment persistence is defined as the existence of serial correlation in individual employment outcomes.

The first chapter finds evidence for unemployment persistence among men and women in the Australian youth labour market. Individual labour market dynamics are analysed using the Australian Longitudinal Survey. The analytic framework used is a Random Effects Probit model, incorporating lagged employment status as an explanatory variable status. Results support a “scarring” effect of unemployment upon individuals’ future employment prospects.

The second chapter provides decision-theoretic foundations for unemployment persistence, based upon heterogeneous intrinsic productivity among workers. A representative firm is assumed to receive an imperfectly precise signal of worker ability every period, and

re-forms its beliefs every period using a Bayesian updating method. A model of the dynamic behaviour of optimal employment decisions by the firm is constructed. It is shown that under certain circumstances workers of all productivities may be “scarred” in the eyes of the firm by past unemployment, due to the firm’s being unwilling to hire from an unemployment pool of dubious quality.

The third chapter presents a detailed investigation into how to measure unemployment persistence within the UK. The chapter presents several modelling strategies capable of being used to analyse panel data of a binary nature, and discusses how to decide which methods are most appropriate in particular environments. Panel data on men from the British Household Panel Survey are used to estimate a structural state dependence equation in employment status, where lagged employment status is used as an explanatory variable. Particular attention is given to controlling for unobserved heterogeneity between individuals. The empirical results indicate strong evidence of unemployment persistence.

Contents

List of Figures	iv
List of Tables	v
1 Introduction	1
2 Unemployment Persistence in Australian Youth Labour Markets	6
2.1 Introduction	6
2.2 Background	9
2.3 Modelling Unemployment	10
2.3.1 Theoretical Model	10
2.3.2 Empirical Model and State Dependence versus Heterogeneity	13
2.4 The Data	17
2.5 Estimation Results	21
2.5.1 High-Education Equations	23
2.5.2 Low-Education Equations	28
2.6 Conclusions	32
3 Towards a Dynamic Information-based Theory of Unemployment Persistence	34
3.1 Introduction	34
3.2 Background	36
3.2.1 Traditional Explanations	36
3.2.2 Information-Based Explanations	37
3.2.3 Evaluation Criteria	40
3.3 Simple Model	41
3.3.1 Exogenous Unemployment Composition	43
3.3.2 Endogenous Unemployment Composition	51
3.3.3 Summary of Results	70
3.4 Extensions to the Simple Model	71
3.4.1 Uncertainty about Productivity of Both Worker Types	72
3.5 Conclusion	79

4	Measuring Unemployment Persistence for the UK	83
4.1	Introduction	83
4.2	Background	84
4.3	Estimation Methodology	87
4.3.1	Empirical Model	87
4.3.2	Estimation of Nonlinear Specifications	92
4.3.3	Estimation of Linear Specifications	106
4.3.4	Summary of Models to be Estimated	113
4.4	Sample Selection Methodology	114
4.4.1	Variable Selection	115
4.4.2	Shape of Data	118
4.4.3	Sample Characteristics	119
4.5	Estimation Results	124
4.5.1	Results from Nonlinear Specifications	125
4.5.2	Results from Linear Models	139
4.6	Robustness Analysis	151
4.6.1	Sample Selection	152
4.6.2	Time Period	154
4.6.3	Comparison of Results with those of other Studies	156
4.7	Conclusion	163
5	Conclusion	165
	Bibliography	169
A	Appendix for Chapter 2	178
A.1	Data Appendix	178
B	Appendix for Chapter 3	183
C	Appendices for Chapter 4	185
C.1	Data Appendix	185
C.2	Further Details of Estimation Procedures used	189
C.2.1	Nonlinear Models	189
C.2.2	Linear Models	190
C.3	Tables of Summary Statistics for Auxiliary Samples E1 and F1	196

List of Figures

2.1	Probability of unemployment: high education females	26
2.2	Probability of unemployment: high education males	27
2.3	Probability of unemployment: low education females	30
2.4	Probability of unemployment: low education males	31
3.1	Composition of firm's labour force using a hire-fire policy	55
4.1	Taxonomy of Random Effects Models	103

List of Tables

2.1	High education Australian born males and females 1985-88 (sample means)	18
2.2	Low education Australian born males and females 1985-88 (sample means)	19
2.3	Within sample marginal effects: high education Australian born individuals, 1985-88	23
2.4	Initial conditions: high education Australian born individuals, 1985-88 . .	25
2.5	Within sample marginal effects: low education Australian born individuals, 1985-88	28
2.6	Initial conditions: low education Australian born individuals, 1985-88 . . .	29
3.1	Persistence thresholds (simple model)	63
3.2	Persistence thresholds (small firm) for Simple Model and Extended Model	76
3.3	Summary of Models' Predictions	79
4.1	Sample Characteristics for BHPS 1991-2000	120
4.2	Probability of Labour Force Transitions	122
4.3	Results from Dynamic Logit Models (Main sample: N=1285, T=1991-2000)	126
4.4	Marginal Effects: Nonlinear Models (Main sample: N=1285, T=1991-2000)	134
4.5	Effect of Controls upon Estimated State Dependence (Main sample: N=1285, T=1991-2000)	136
4.6	Results from Linear Models: Difference GMM (Main sample: N=1285, T=1991-2000)	141
4.7	Results from Linear Models: System GMM (Main sample: N=1285, T=1991-2000)	143
4.8	Diagnostic Analysis on Standard Errors from Linear Models (Main sample: N=1285, T=1991-2000)	147
4.9	Results from Linear Models using Invalid or Implausible Restrictions (Main sample: N=1285, T=1991-2000)	148
4.10	Marginal Effects: Nonlinear and Linear Models (Main sample: N=1285, T=1991-2000)	150

4.11	Marginal Effects over different Sample Selection Rules	
	153	
4.12	Estimated State Dependence over different Time Periods	
	155	
4.13	Estimated Marginal Effects by Studies	157
4.14	Effects of Age and Educational Partitioning upon Estimated Marginal Effects (Sample: T=1991-1995, except for Column VI where T=1991-2000)	159
4.15	Sample Characteristics, BHPS 1991-1995, among men younger than 25 years in 1991	160
4.16	Sample Characteristics, BHPS 1991-1995, among men 25 years or older in 1991	161
C.1	Sample Characteristics for BHPS: Sample E1 1991-1995	196
C.2	Sample Characteristics for BHPS: Sample F1 1991-1995	197

Acknowledgments

Thanks are due to many people, including Professor Neville Norman for encouraging me to embark upon this course of study and Dr. Mark Harris for the initial idea for the topic. I am also grateful to Raunaq Pungaliya for many discussions and much work on this topic and to Maxim Bouev for programming and technical assistance. During this process, my wife Gabrielle has provided incalculable encouragement and support.

I wish to thank my examiners, Professor Wiji Arulampalam and Professor Margaret Stevens, for their detailed and helpful suggestions to improve the content and exposition of the thesis.

Most of all, I am deeply grateful to both my supervisors, Steve Bond and Mary Gregory, and have learned immensely from both.

Chapter 1

Introduction

The issue of unemployment presents many challenges for economists and policy makers alike. This is not merely due to the existence of high overall levels of unemployment in many industrialised countries. There is also significant concern over the composition of the unemployed. As the authors of a leading book on unemployment assert:

“Unemployment rates differ widely between occupations and between regions, as well as across age, race, and (sometimes) sex groups. The striking thing is how *stable* these differences are. In all countries, unskilled people have much higher unemployment rates than those with skills. Similarly, youths have much higher rates than adults. In addition, in most countries (though not in the USA) regional differences are highly persistent - with unemployment always above average for example in the North of England and the South of Italy.”
(Layard, Nickell and Jackman 2005, italics added.)

This thesis investigates unemployment persistence at the individual level, where unemployment persistence is defined as the existence of serial correlation in employment outcomes. It is structured around three distinct chapters, each of which makes a contribution to better understanding of this phenomenon.

The thesis begins with an empirical investigation (Chapter 2), which looks into employment outcomes in the Australian youth labour market over 1985-1988. In this study, individual labour market dynamics are analysed using the Australian Longitudinal Survey. A random utility framework for analysing discrete choices is adopted. In this context, a model incorporating a state dependent relationship between employment outcomes is estimated. The influence on individual employment outcomes of additional variables including education, gender and unemployment benefits is also investigated. It is found that, even after controlling for observable and unobservable differences between individuals, there is strong evidence of unemployment persistence. In certain key respects, mainly related to the use of panel data, the findings of this study differ markedly from those of other Australian labour market studies.

Historically, Chapter 2 was written before the other two substantive chapters and is a relatively short empirical study. The chapter was jointly authored and later published as Knights *et al* (2002): Stephen Knights was the lead author, and was responsible for developing the empirical methodology, interpreting results and drafting the actual text of the paper. Mark N. Harris was responsible for developing numerical routines, especially those relating to Gaussian quadrature. Joanne Loundes was responsible for handling the data and coordinating discussion.

While the findings of Chapter 2 provide further insight into the causes of contemporary unemployment, the question of explaining the structural relationships this study identifies still remains to be dealt with. Put another way, Chapter 2 presents evidence for the existence of unemployment persistence, but does not explain why it should have arisen

in the first instance. This theoretical issue is addressed in Chapter 3. It is argued that heterogeneous intrinsic productivity among workers can explain unemployment persistence, by examining the dynamic behaviour of optimal employment decisions in an environment where uncertainty over worker types exists. A representative firm is assumed to receive a 'noisy' signal of worker ability every period, and re-forms its beliefs anew every period using Bayesian updating. Under certain circumstances, it can be shown that persistent unemployment results, even for high-productivity workers, due to the firm's being unwilling to hire from an unemployment pool of dubious quality.

This is an important result. Despite the existence of a large economic literature on the effects of incomplete and/or imperfect information upon market outcomes, little formal work exists using information-theoretic explanations to model unemployment persistence. This is perhaps surprising, as anecdotal evidence exists that employers often view recent unemployment spells poorly when evaluating potential employees, and may be reluctant to hire those recently unemployed. Chapter 2 provides a rationale for unemployment persistence consistent with this, and points to further theoretical research in this direction.

Having established that unemployment persistence possesses both an empirical basis and a theoretical rationale, the task remains to document the extent to which it is observed. Chapter 2 partly accomplishes this; however, since that study uses relatively old data which are restricted to the youth labour market, a natural extension is to pursue an empirical investigation using a more up to date and representative data set. Chapter 4 accomplishes this: it estimates unemployment persistence within the UK by considering BHPS data during 1991-2000.

Both in scope of data used, and analytical methods embraced, Chapter 4 is a more detailed work than Chapter 2. First, in terms of time period used, Chapter 2 only uses a 4 year panel, while Chapter 4 uses a 10 year panel. Second, Chapter 4 estimates results using several linear and nonlinear models to compare and contrast estimation results, compared with Chapter 2 which restricts analysis to a random effects probit model.

Chapter 4 has other contributions to make, beyond providing evidence of contemporary unemployment persistence. In particular, it sets out a methodological survey of panel data methods capable of analysing binary data more generally. Although books such as Hsiao (2003) have sections dealing with such issues, Chapter 4 goes into considerable detail to explain the suitability of several estimation methods in specific situations. There is also discussion of model selection within the context of measuring unemployment persistence. Therefore, Chapter 4 can be seen as an empirical and methodological work, as well as an up to date work of reference on panel data econometrics.

Several nonlinear models are examined in Chapter 4, where disturbances are posited to follow a logistic distribution. Based upon this principle, standard random effects logit models can identify parameters of interest through maximum likelihood estimation (MLE), where distributional restrictions upon individual effects are imposed. To further relax these distributional assumptions, it is desirable to estimate a conditional fixed effects model. However, if individual effects are fixed then MLE is inconsistent. To address this issue, a conditional fixed effects model is estimated using a method proposed by Honore and Kyriazidou (2000): this method identifies the underlying parameters through examining those consecutive observations where an individual has gone from employment to

unemployment, or vice-versa. Linear models are also used, both as a means to performing diagnostic checks on the results from the nonlinear models and as potentially alternative specifications.

The results from all models estimated show that, after controlling for a rich set of observed characteristics, and allowing for unobserved differences between individuals, there is strong evidence of unemployment persistence. These findings are consistent with results from the earlier analysis in Chapter 2 and with the theory put forward in Chapter 3.

Chapter 2

Unemployment Persistence in Australian Youth Labour Markets

2.1 Introduction

It is important to know at an individual level whether “state dependence” drives unemployment persistence, or conversely, whether it is “heterogeneity”. Namely, is the cause of the relative disadvantage of an unemployed worker the mere *experience* of being unemployed (i.e. state dependence), or alternatively is this due to intrinsic differences between individuals (i.e. heterogeneity)?

On this issue, disagreement exists in the literature. For example, Phelps (1972) argues that true state dependence exists. By contrast, Cripps and Tarling (1974) argue that what appears to be state dependence is, in fact, “spurious” state dependence due to unobserved heterogeneity. Although the issue of state dependence versus heterogeneity has been examined in several international studies, the only Australian studies that touch on this issue are those by Le and Miller (2001) and Buddelmeyer *et al* (2010). The contribution of this Chapter is to investigate whether such dynamic relationships exist in the Australian youth labour market.

To assess the validity of these competing explanations, it is necessary to define a

dynamic model of unemployment. Econometrically, this necessitates inclusion of a lagged dependent variable as an explanatory variable. Moreover, theoretical arguments have been put forward to justify the existence of dynamic relationships within labour markets. If the event of being unemployed leads to a slowdown, or even a reversal, in the growth of human capital, an individual's chances of again finding employment would progressively worsen (Heckman, 1981a). This would be theoretically consistent with state dependence. Alternatively, as we argue later in Chapter 3, if employers have incomplete information about individuals' respective marginal revenue products, they may use past unemployment records as a screening device. Again, this could account for such observed behaviour. Finally, if individuals' time horizons are shortened by past spells of unemployment, they may systematically substitute away from present consumption towards leisure, implying that their opportunity cost of being unemployed would fall (Hotz *et al*, 1988). This, too, would be consistent with state dependence.

These theoretical arguments all share the feature that they support an explicit dynamic process; that is, where individuals' employment prospects change over time. However, the theoretical arguments are also consistent with purely static explanations for unemployment. Most notable among these are the insider-outsider hypotheses advanced by Layard and Nickell (1985) and Lindbeck and Snower (1986), where the labour force is modelled as being partitioned according to whether one is in employment or not.

The potential policy implications of the existence of dynamic relationships are important. A typical goal of policy making is to increase employment incidence within disadvantaged groups in the labour force. However, effectively attaining such a goal depends upon recognising distinctions such as the above, which relate to the causes - and thus composition - of unemployment. If employment outcomes are strongly state dependent, future policies which increase aggregate employment may have different impacts across cohorts of workers, depending on the existing employment incidence within each cohort. Under such circumstances, it is likely that the benefits of greater employment will be concentrated disproportionately among those already employed, to the exclusion (and further disadvantage) of those currently unemployed. Therefore, if the goal of policy-makers is to increase employment among disadvantaged groups, merely increasing the amount of employment opportunities may not be sufficient.

The structure of the study is as follows. Section 2.2 surveys the characteristics of dynamic models of unemployment, paying particular attention to differences between, and similarities within, the contemporary literature. Section 2.3 develops a simple model

of state-dependence within a random utility equilibrium framework. Section 2.4 describes the data used for the empirical work, and Section 2.5 discusses the results. Section 2.6 concludes.¹

¹The Chapter is joint work with Mark N. Harris and Joanne Loundes. A later version of this Chapter has been published as Knights *et al* (2002).

2.2 Background

The techniques used in the literature to model state dependence are highly influenced by whether the dynamics are modelled in discrete or continuous time. To model transitions in and out of the state (as opposed to the duration) of unemployment, then discrete time modelling is accomplished by using binary variables to denote employment outcomes. This is the approach followed in this Chapter, and is typified in Flaig *et al* (1993) and Arulampalam (2004). For an overview of these and other modelling strategies, see Lancaster (1990, Ch.3).

Many of the empirical models rely exclusively on cross-sectional data, where individuals are not identified over time. This was particularly prevalent in earlier papers, including Heckman and Borjas (1980), Narendranathan *et al* (1985) and Miller (1989); however, it usually owed less to theoretical or empirical advantages than to the greater ease with which such data could be obtained. A very small proportion of models, such as Jackman and Layard (1990), use pure time series data. However, a consensus has emerged in favour of - wherever possible - utilizing the greater flexibility and informational content of longitudinal (i.e. panel) data. A key advantage of panel data is that it allows the researcher to adequately control for unobservable individual heterogeneity, and hence to test the hypothesis of state dependence versus heterogeneity.² The only Australian studies that adopt a similar approach to the current paper are those by Le and Miller (2001), who use the *Survey of Employment and Unemployment Patterns* to calculate an index of the risk of becoming unemployed, and Buddelmeyer *et al* (2010). The Le and Miller (2001) analysis involves a 3-year panel, where a conventional model of unemployment (i.e. one that cannot account for unobserved characteristics) is supplemented by incorporating measures of the individual's labour market history and family characteristics to proxy for unobserved influences. The Buddelmeyer *et al* (2010) analysis uses longitudinal data from the *Household, Income and Labour Dynamics Survey* to examine the extent to which relatively high rates of transition from low-paid employment into unemployment are the result of disadvantageous personal characteristics.

²Heckman (1981c, p.92) uses the term spurious state dependence to denote the erroneous labelling as state dependence of what is, in fact, unobservable heterogeneity. The issue is further discussed in Section 3.1 below.

2.3 Modelling Unemployment

2.3.1 Theoretical Model

On the supply side, individuals make optimal decisions to allocate available time between work and leisure.³ Individuals may also have access to a financial endowment, which may either be unrelated to the work-leisure decision (such as existing wealth), or alternatively may be contingent upon it (such as unemployment benefits). In the presence of wage rates which are low relative to prices of goods and services, it may be optimal for individuals to allocate no time for work. Therefore, each individual may be said to have a certain reservation wage. If wages fall below reservation wages, employment will not be chosen.

The determinants and functional form of utility potentially vary - perhaps widely - across individuals; therefore, a need for caution exists when compiling a list of variables which may determine utility, and hence also affect reservation wages. However, past studies have indicated that characteristics such as education, marital status, level of past indebtedness acquired and experience broadly influence work-leisure choices across society.

In the subsequent estimations for men, experience is proxied by current year minus the year the individual left school. Due to this parameterisation, a negative relationship between age and employment is expected. A slightly different method is used to calculate female potential experience, as this is more likely to suffer from interrupted spells in the labour force due to the presence of children. This is done by weighting experience (i.e. age minus year left school) by the number of children present.⁴

Labour market theory also suggests that, at least in the medium run, any regional imbalances in unemployment rates - up to the cost of inter-regional migration differentials - will be erased. However, it is quite possible that, in the short run, one's place of abode affects employment prospects. That is, place of residence may also affect the intensity of job search.

On the demand side, firms make hiring decisions based on maximising profits. This problem's optimal solution involves hiring only workers whose estimated marginal

³It should be noted that 'work' in this context is a simplification, which includes both time spent in actual employment, and time spent searching for employment when a job cannot instantly be found. A substantial literature on the latter exists: for a useful economic and econometric overview of search theory see Zaretsky & Coughlin (1995).

⁴Potential rather than actual labour market experience is used because the actual labour market experience variable is of poor quality.

revenue products are at least as high as their respective wage rates. Assuming that each worker's individual marginal product does not vary across hours of employment, the firm decides whether to hire using this criterion. The Australian wage determination system is a particularly pertinent issue in this instance, as the survey to be used for this analysis falls in the first half of the Prices and Incomes accord. Between August 1985 and August 1988, real wages fell at an average annual rate of 1.7%, accompanied by strong growth in employment (3.1% per annum). Youth employment growth (individuals aged between 15 and 24) was somewhat lower, but on average it was higher than that experienced by youth over the past decade. In broad terms, youths involved in the survey during the period 1984 to 1988 were faced with falling real wages, but increased employment opportunities.

Simultaneously solving demand and supply will theoretically lead to the establishment of an equilibrium wage for each worker. However, such an approach is likely to prove empirically intractable, due to identification issues.⁵ For example, education may increase an individual's expectations of high remuneration and thus their reservation wage (supply-side) while simultaneously increasing their marginal revenue product (demand-side). In such circumstances, where it is impossible to observationally distinguish supply factors from demand factors, the structural system is said to be under-identified, and individual demand and supply equations cannot be estimated separately.⁶ Nevertheless, to avoid the unrealistic assumption of ignoring demand side issues, there are several variables that can be incorporated which may potentially affect an individual's attractiveness to an employer. These include education, race, age, physical ability and gender.⁷ Labour market participants may be disadvantaged according to race, physical ability and gender. Human capital, as proxied by age and educational attainment, is likely to be viewed as a favourable trait by potential employers.

To avoid the likely under-identification problems inherent in structurally modelling labour demand and supply separately, employment outcomes are assumed to be equilibrium values, and a set of variables which may explain these outcomes is constructed. No attempt

⁵Heuristically, the problem is that changes in equilibrium wage rates could be generated either by shifts in demand or supply. If explanatory variables are correlated with both demand and supply, then such shifts cannot be, empirically, distinguished from each other. For a general discussion of identification issues, see Johnston (1984, pp.452-460).

⁶One way to separately estimate supply and demand equations is to obtain explicit data on individual's reservation wages. Work along these lines has been completed in an Australian context by Heath & Swann (1999).

⁷The sample only includes Australian born individuals, but respondents can be identified as being of Aboriginal or Torres Strait Islander, Western, Asian or 'other' racial background. However, a lack of observations resulted in the race variable being excluded.

is made to further classify any of these variables as belonging, exclusively, to either the supply or demand side of the labour market.

Therefore, *in lieu* of individually estimating demand and supply equations, a probabilistic approach is used, following Flaig *et al* (1993), to model employment outcomes. An individual's employment status (employed or unemployed) is simply modelled as a binary variable, the result of an underlying latent index. This latent index can be viewed as the excess of wage offers for individual i in time period t (w_{it}), over their contemporaneous reservation wages, w_{it}^* thus

$$y_{it}^* = w_{it} - w_{it}^*, \quad i = 1, \dots, N, \quad t = 1, \dots, T \quad (2.1)$$

This index is assumed to be a function of observed characteristics (as well as macroeconomic variables) x_{it} , with unknown weights β , such that

$$y_{it}^* = x_{it}'\beta + v_{it} \quad (2.2)$$

where v_{it} is a stochastic disturbance term, assumed to be normally distributed with zero mean and scalar variance σ_v^2 . However y_{it}^* is not observed for each individual. Instead the realisation of y_{it} is observed, where the individual is employed if y_{it}^* is positive and unemployed otherwise

$$y_{it} = \begin{cases} 1 & \text{if } y_{it}^* \geq 0 \\ 0 & \text{if } y_{it}^* < 0 \end{cases}$$

where unity denotes employed and zero unemployed.⁸

⁸There is no focus in this paper on individuals who are not in the workforce.

2.3.2 Empirical Model and State Dependence versus Heterogeneity

To allow for unobserved heterogeneity, the usual procedure is followed [see, for example, Butler and Moffitt (1982) and Hsiao (2003)] in decomposing v_{it} into a time-invariant individual effect η_i , and a component which varies across individuals and across time periods ε_{it}

$$v_{it} = \eta_i + \varepsilon_{it} \quad (2.3)$$

The ε_{it} are stochastic error terms, assumed to follow an independent normal distribution with zero mean and variance σ_ε^2 . However, an issue arises as to the specification of the unobserved effects η_i . If they are assumed to be random but are in fact correlated with the observed heterogeneity, then this leads to inconsistent parameter estimation if this correlation is not taken into account. Therefore, in the simple case, the literature suggests a “fixed effects” approach if such correlation is thought to be present, and a “random effects” approach if not.

In many cases, however, this is not the case for non-linear models (such as probits). This is because a fixed effects approach yields inconsistent parameter estimation (for all parameters, and not just the η_i) in typical panel datasets characterised by small T [see Greene (2000), Chapter 19]. Hence in such non-linear models most of the literature follows Chamberlain (1984), who uses a random effects approach and explicitly models any potential correlation. This is the approach followed in this Chapter: the idiosyncratic unobserved effects η_i are treated as random draws (from a normal distribution with zero mean and variance σ_η^2 , independent of ε_{it}) and any potential correlation between these and the observed heterogeneity is accounted for, as described below.

Following Butler and Moffitt’s (1982) specification of the unobserved effects, the correlation between composite error terms v_{it} , over time (the *intra-class correlation coefficient*) is

$$\rho \equiv \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}, \quad t \neq s \quad (2.4)$$

This particular specification for the unobserved heterogeneity is somewhat restrictive in that it implies equal correlation over individuals across time periods. It does, however, yield a tractable likelihood function [see Greene (2000, Chapter 19) for discussion].

To test the hypothesis of true state dependence, it is necessary to augment the vector of explanatory variables to include the individual’s previous employment status,

$y_{i,t-1}$. Thus the full model becomes

$$y_{it}^* = x'_{it}\beta + \delta y_{i,t-1} + \eta_i + \varepsilon_{it} \quad (2.5)$$

On the assumption of independent normality and, for identification purposes, normalising σ_η^2 to unity, this yields the generic probability of employment for any individual i and time period t :

$$P(y_{it} = 1 | x_{it}, \beta, \delta, \eta_i) = \Phi \left[\frac{x'_{it}\beta + \delta y_{i,t-1} + \eta_i}{\sigma_\varepsilon} \right]$$

where $\Phi[c] \equiv \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^c \exp(-\frac{1}{2}s^2) ds$ is the cumulative distribution function of the standard normal probability distribution. This probability is conditioned upon unobserved heterogeneity η_i so we can generalise to allow for the assumption that η_i also follows the normal distribution:

$$P(y_{it} = 1 | x_{it}, \beta, \delta) = \int_{-\infty}^{\infty} \Phi \left[\frac{x'_{it}\beta + \delta y_{i,t-1} + \eta_i}{\sigma_\varepsilon} \right] \phi(\eta_i) d\eta_i$$

where $\phi(c) \equiv \frac{1}{(2\pi)^{1/2}} \exp(-\frac{1}{2}s^2)$ is the probability density function of the standard normal probability distribution.

Therefore, it follows that the joint probability $P(y_i)$ of observing a given sequence of employment outcomes $y_{i1}, y_{i2}, \dots, y_{iT}$ for individual i is:

$$P(y_i) = \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{1/2}} \exp\left(-\frac{1}{2}\eta_i^2\right) \prod_{t=2}^T \Phi \left[\left(\frac{x'_{it}\beta + \delta y_{i,t-1}}{\sigma_\varepsilon} + \eta_i \left(\frac{\rho}{1-\rho} \right)^{1/2} \right) (2y_{it} - 1) \right] d\eta_i$$

which can be rewritten more compactly as

$$P(y_i) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \exp(-\tilde{\eta}_i^2) \prod_{t=2}^T \Phi \left[\left(\frac{x'_{it}\beta + \delta y_{i,t-1}}{\sigma_\varepsilon} + \tilde{\eta}_i \theta \right) (2y_{it} - 1) \right] d\tilde{\eta}_i \quad (2.6)$$

where $\tilde{\eta}_i \equiv \frac{\eta_i}{\sqrt{2}}$ and $\theta \equiv \left(\frac{2\rho}{1-\rho} \right)^{1/2}$

The log likelihood function across all individuals is then defined as:

$$L = \sum_i \log P(y_i) \quad (2.7)$$

The integral in Equation (2.7) has no closed form representation, and has to be evaluated using Gaussian quadrature, using the Hermite integration formula (see Butler and Moffitt, 1982).

As noted above, consistent estimation obtained by maximisation of Equation (2.7) relies on the assumption that all of the explanatory variables are independent of the composite error term. Following Chamberlain (1984), it is possible to allow for correlations of the individual effects and x_{it} , by including the time average of time-varying variables as additional explanatory variables. This is the approach adopted in this Chapter.⁹

However, this procedure is not suitable for the lagged endogenous variable, in which instance the so-called initial conditions problem arises. This problem has been documented in the relevant literature [see Heckman (1981a, b and c), Flaig *et al* (1993), Orme (1997), Arulampalam (2004), Arulampalam and Stewart (2009)].

The consequence of the lagged dependent variable in Equation (2.5) is that, unless the initial values are truly exogenous, they will be correlated with the individual effect of the same equation. Heckman's (1981b) simplified procedure for dealing with this involves approximating the latent variable y_{i1}^* by a linear function of relevant pre-sample information x_{i1} thus

$$y_{i1}^* = x_{i1}'\beta_1 + \psi_i \quad (2.8)$$

To allow for a correlation between the individual effects of equations (2.5) and (2.8), ψ_i is specified as a linear function of η_i such that

$$\psi_i = \gamma\eta_i + \varepsilon_{i1} \quad (2.9)$$

where ε_{i1} has mean zero and standard deviation σ_{ε_1} .

Therefore the extent of correlation between η_i and ψ_i is, by construction, a function of γ

⁹Note that if this correlation exists but it has been modeled incorrectly, inconsistent estimation may still result.

and the individual standard deviations σ_η and σ_ψ :

$$\rho_{\eta,\psi} = \gamma\sigma_\eta/\sigma_\psi \quad (2.10)$$

The vector x_{i1} typically contains all of the original variables in x_{it} plus any additional pre-sample information.

For computational purposes it is assumed that the unobserved effects and the disturbance terms are independent and, moreover, that the disturbance terms are serially independent. If, on the other hand, the disturbances are serially correlated and this is not controlled for appropriately in estimation, it is possible that this may, in part, be reflected in the apparent (erroneously) strong significance of the lagged dependent variable. In other words, if serial correlation is present in the unobserved terms of the model—and this is ignored—this effect may be absorbed by the lagged dependent variable, potentially exaggerating the presence of any state dependence.

However, even in the static panel case, if this is not the case, that is if the variance-covariance matrix of v_{it} is assumed to be *unrestricted* (allowing for potential heteroskedasticity and/or autocorrelation), estimation of the model now requires much more complicated techniques [see, for example, Inkermann (2000)]. These methods are further complicated by the presence of the lagged dependent variable utilised in this study.

Under this set of assumptions, full information maximum likelihood estimation is carried out by augmenting the product of Equation (2.6) by

$$\Phi \left[\left(\frac{x'_{i1}\beta_1 + \eta_i\gamma}{\sigma_{\varepsilon_1}} \right) (2y_{i1} - 1) \right]$$

yielding

$$\begin{aligned} P(y_i) &= \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \exp(-\tilde{\eta}_i^2) \prod_{t=2}^T \Phi \left[\left(\frac{x'_{it}\beta + \delta y_{i,t-1} + \tilde{\eta}_i\theta}{\sigma_\varepsilon} \right) (2y_{it} - 1) \right] d\tilde{\eta}_i \\ &\times \Phi \left[\left(\frac{x'_{i1}\beta_1 + \eta_i\gamma}{\sigma_{\varepsilon_1}} \right) (2y_{i1} - 1) \right] \end{aligned} \quad (2.11)$$

and constructing the log likelihood function using the complete specification of $P(y_i)$ in equation (2.11).

2.4 The Data

The Australian Longitudinal Survey (ALS) has been used extensively in analyses of the Australian labour market and is an adequate data set for the purposes of this study due to the size of the sample, the range of responses recorded, and - most importantly - the panel nature of the data.¹⁰ It contains survey responses from almost 9000 individuals, covering a wide range of demographic and labour market characteristics. In this Chapter, the data used cover the period 1985 to 1988; initially the data set contained 8998 respondents aged 15 to 29 years, which fell to 6151 by 1988. General descriptions of this data set, which is managed by the Bureau of Labour Market Research (BLMR), may be found in McRae (1984) and in Harris (1993). The most pertinent study for this paper is that of Harris (1996) which considers a static panel probit model of unemployment. However, since this model is static rather than dynamic, it could not address the issue of state dependence versus heterogeneity.

The ALS has been found to suffer from sample attrition. For example, Harris (1996) found that attrition could be attributed to education and gender - low-education males and females were more likely to ‘deselect’ themselves from the ALS than were their high-education counterparts. To avoid selection bias, a sample selection model could have been considered. However, identification issues are likely to be problematic, as the same variables that affect employment status will affect attrition rates. Moreover, this would unduly complicate the econometric specification. A more attractive approach is to follow Harris (1996) and simply remove those observations with incomplete records, but then estimate separate parameters for the separate sub-groups defined by the factors most related to the endogenous attrition. Thus the data was subdivided into four subgroups: high-education males; high-education females; low-education males and; low-education females. If the attrition is based on observables only, then consistent inferences can now be drawn from analysis of each subgroup.

As the ALS is an annual survey, to keep the units of measurement consistent, the most appropriate measure of employment status is the derived variable, EMPSTAT. This measure reflects the individual’s employment status as at interview date.¹¹

Tables 2.1 and 2.2 present the sample means of the variables that are used in the

¹⁰For a comprehensive bibliography, see the Department of Employment, Education and Training (1990).

¹¹It would be possible to use the work history diaries to construct an employment status series within each year for each individual, as opposed to a single observation. However, this would lend itself more to a duration analysis (see Section 2.2).

Table 2.1: High education Australian born males and females 1985-88 (sample means)

Variable	Male			Female		
	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.
<i>Within Sample</i>						
Lagged employment	3776	0.86	0.35	2732	0.90	0.30
Experience	3776	6.60	2.39	2732	5.73	2.30
Marital status	3776	0.30	0.46	2732	0.31	0.46
Separated	3776	0.01	0.10	2732	0.01	0.11
City	3776	0.76	0.42	2732	0.79	0.41
Rural	3776	0.07	0.26	2732	0.05	0.22
Own house	3776	0.19	0.39	2732	0.17	0.38
Rent free	3776	0.14	0.35	2732	0.12	0.32
Renting	3776	0.66	0.48	2732	0.69	0.46
Year 12	3776	0.29	0.45	2732	0.33	0.47
Degree	3776	0.09	0.28	2732	0.08	0.27
Diploma	3776	0.14	0.34	2732	0.44	0.50
Trade certificate	3776	0.37	0.48	2732	0.05	0.21
Partner's employment	3776	0.20	0.40	2732	0.29	0.46
Health status	3776	0.07	0.25	2732	0.07	0.26
Children	3776	0.19	0.57	2732	0.08	0.36
Average experience	3776	6.60	2.10	2732	5.74	2.01
<i>Initial Conditions</i>						
State school	944	0.74	0.44	683	0.72	0.45
Manufacturing industry	944	0.13	0.34	683	0.04	0.19
Both parents	944	0.90	0.30	683	0.89	0.31
Less than 6 months	944	0.32	0.47	683	0.35	0.48
Six months	944	0.10	0.30	683	0.15	0.35
Employment history	944	0.74	0.44	683	0.70	0.46
Unemployment history	944	0.03	0.16	683	0.04	0.19

Table 2.2: Low education Australian born males and females 1985-88 (sample means)

Variable	Male			Female		
	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.
<i>Within Sample</i>						
Lagged employment	2108	0.75	0.43	1176	0.84	0.36
Experience	2108	6.11	2.94	1176	5.43	2.56
Marital status	2108	0.25	0.43	1176	0.31	0.46
Separated	2108	0.01	0.12	1176	0.02	0.14
City	2108	0.63	0.48	1176	0.63	0.48
Rural	2108	0.12	0.32	1176	0.10	0.30
Own house	2108	0.10	0.30	1176	0.15	0.36
Rent free	2108	0.16	0.37	1176	0.13	0.33
Renting	2108	0.72	0.45	1176	0.70	0.46
Year 10	2108	0.56	0.50	1176	0.50	0.50
Year 11	2108	0.29	0.45	1176	0.39	0.49
Partner's employment	2108	0.12	0.32	1176	0.30	0.46
Health status	2108	0.10	0.30	1176	0.07	0.25
Children	2108	0.21	0.59	1176	0.13	0.47
Average experience	2108	6.12	2.71	1176	5.44	2.31
<i>Initial Conditions</i>						
State school	527	0.89	0.31	294	0.88	0.32
Manufacturing industry	527	0.17	0.37	294	0.09	0.28
Both parents	527	0.86	0.35	294	0.86	0.35
Less than 6 months	527	0.27	0.45	294	0.30	0.46
Six months	527	0.15	0.36	294	0.20	0.40
Employment history	527	0.57	0.50	294	0.60	0.49
Unemployment history	527	0.09	0.28	294	0.10	0.29

estimation for the high-education and low-education groups, respectively.¹² Individuals who have completed at least secondary school (that is, high education individuals) are more likely to be employed over the sample period compared to their lower education counterparts. Within the high-education group, males are more likely to have a trade qualification (37% compared to 5% for females), whereas females are significantly more likely to have completed a diploma (44% versus 14%). Within the low-education group, males are more likely to be single (75% versus 69%) but less likely to have an employed partner. In comparison, low-education females are more likely to have completed year 11 as compared to low-education males.

In terms of the initial conditions variable means, high-education individuals are less likely to have gone to a state school, or to have taken more than six months to find their first job. They are more likely to have been employed at the time of the survey. Males are more likely to have got their first job in the manufacturing industry, especially among those who have not completed a secondary education.

For more information about the data used, including definitions of all variables, see Appendix A.1.

¹²The age range is 15 to 29 years for the low education respondents, and 16 to 29 years for the high education respondents.

2.5 Estimation Results

A balanced panel is used to obtain the empirical results. In a strict sense, it is not necessarily restricted to individuals who had been in the labour force in all four waves of the data. Individuals are allowed to have periods outside the labour force, but have to be either “predominantly employed” or “predominantly unemployed” during each of the four years to remain in our panel. The variable EMPSTAT was used to determine the number of individuals left out of our survey according to these criteria. Just over 5 per cent of the surveyed respondents were not in the labour force for all four years, 81% of who were women. Of those that would have liked to work either full-time or part-time, several reasons were given for not looking for work. For men the overwhelming factor was due to undertaking further study. For women the results were evenly spread between childcare problems, study, other reasons and those that had no stated reason for not wanting work. Trouble finding a job was a listed option, but very few respondents reported that this was the reason they were not looking, suggesting that the ‘discouraged worker’ effect was not a major issue. Therefore, it is unlikely that parameter estimation is strongly influenced by the exclusion of individuals not in the labour force. Additionally, the concern of this paper is the effect of heterogeneity and/or state dependence on labour market outcomes, and as such, the analysis is primarily concerned with those who are part of the labour force, providing justification for why non-participants are excluded.

As evaluation of Equation (2.11) requires approximation methods using Gaussian quadrature, different numbers of quadrature points were considered. If parameter estimates change significantly as the number of points increases, this suggests that the approximation is a poor one, casting doubts as to the validity of the results. Three sets of estimations were undertaken for each group, each involving 8, 12 and 16 point quadrature. Movement across the number of quadrature points did not result in any significant change in any of the parameter estimates, which demonstrates that estimation is robust to number of quadrature points. The results presented were obtained using 16 point quadrature.

The results for the structural parameters β and δ are presented as marginal effects, whereas those for the initial conditions model β_1 are simply the estimated coefficients. The marginal effects are calculated on the assumption that $E(\eta_i) = E(\varepsilon_{it}) = E(\varepsilon_{i1}) = 0$, such that $P(y_{it} = 1) = \Phi(x'_{it}\beta + \delta y_{i,t-1})$. For continuous variables, marginal effects are calculated using $\partial P / \partial x_k = \phi(x'_{it}\beta + \delta y_{i,t-1}) \beta_k$, where $k = 1, \dots, K$ and K is the number of explanatory variables in the structural model. For binary variables, the marginal effects are

calculated as the difference between the implied probabilities when the variable respectively takes the values of one and zero. Expressions are otherwise evaluated at the sample means of each variable. Significance tests and standard errors are, however, based on the underlying estimated coefficients.

Note that since the employment status is defined as equal to one when an individual is employed, a positive marginal effect for a given variable x means that the probability of employment increases (i.e. that the probability of unemployment decreases) for an increase in the value of x .

2.5.1 High-Education Equations

Table 2.3 presents the estimation results, with respect to the employment status equation, for males and females with high levels of education.

Table 2.3: Within sample marginal effects: high education Australian born individuals, 1985-88

Variable	Male		Female	
	Effect	t-statistic	Effect	t-statistic
Constant	0.025	1.97	0.015	0.33
Lagged employment	0.023	5.48	0.080	4.06
Experience	0.000	-0.13	0.007	1.09
Marital status	0.003	0.65	0.016	0.61
City	0.000	-0.14	0.035	2.28
Rural	-0.020	-4.09	-0.004	-0.21
Own house	0.013	1.52	0.021	0.61
Rent free	0.017	1.82	-0.046	-1.25
Renting	0.000	-0.10	0.008	0.24
Year 12	0.005	0.87	0.042	2.24
Degree	0.017	2.36	0.041	1.47
Diploma	0.007	1.35	0.024	1.29
Trade certificate	0.003	0.71	-0.024	-0.10
Partner's employment	0.002	0.49	-0.012	-0.41
Health status	-0.006	-1.21	-0.037	-2.10
Children	0.001	0.35	-0.029	-2.02
Average experience	-0.003	-1.63	-0.004	-0.51

The key finding is the significance of the coefficient on the lagged dependent variable, lagged employment. This finding applies irrespective of gender and suggests that, even after controlling for observed and unobserved heterogeneity, past employment status significantly predicts present employment outcomes. This suggests the employment prospects of individuals are affected by something intrinsic to the experience of being employed or unemployed.

Each set of results reveals several characteristics that are important in determining individual employment status. However, no variable other than the lagged effect is significant—at a 5% two-tailed significance level—for both genders at once. In particular, marital status, partner's employment and experience appear to have no impact on employment prospects for either gender.

Educational attainment increases the likelihood of employment, even within the high-education groups. However, the effect for males is observed only for those with a degree; while for females, the difference lies between those with at least Year 12 and those

with only Year 11. This suggests some segmentation by gender, with females more likely to find employment not requiring tertiary education (but requiring Year 12), while males may be less likely to find such employment.

Regional differences also appear significant, but again the effects differ across gender. For males, employment prospects are negatively affected by living outside a city or country town, but no other effects can be detected. Females, by comparison, have significantly lower probabilities of being employed *anywhere* outside of a city. Again, this suggests labour market segmentation, with males more likely to work in country towns relative to females, but both less likely to find rurally-based employment (relative to city-based employment).

The only variables which significantly affect employment prospects, therefore, are those relating to past employment, education and place of residence, with an additional negative effect for females who suffer from a disability or who have children. By contrast, Harris (1996) found that age, housing arrangements, marital status and partner's employment status were also significantly correlated with employment status, across gender divisions. These substantially different results suggest that a great deal of correlation attributed to various explanatory variables in the earlier study may be due directly to past differences in employment outcomes across individuals.

Such an interpretation is consistent with the results from the auxiliary Initial Conditions model, reported in Table 2.4. Very few coefficients are significant. Those that are include, for males, experience, although with an unexpected sign; and, for both genders, length of search time for first job, and employment/unemployment history. However, the results from the Initial Conditions model, which is a reduced-form marginal model, are not directly comparable with the results from the main model which is a structural model conditioned upon previous employment status.

From Table 2.4 we note that the estimated value of ρ is significantly different from zero for both males and females, suggesting that there is evidence of individual level unobserved heterogeneity. The estimated value of θ is also significantly different from zero, which indicates that the endogenous initial conditions approach to estimating the employment equation is justified.

Figures 2.1 and 2.2 present, for females and males respectively, the probability of being unemployed over time (measured in percentage point units so that 1.0 refers to a probability of 1% or 0.01). Each probability depends upon whether an individual is likely to be employed or unemployed in the first period (i.e. the pre-sample period $t = 1$).

Table 2.4: Initial conditions: high education Australian born individuals, 1985-88

Variable	Male		Female	
	Coefficient	t-statistic	Coefficient	t-statistic
Constant	0.779	0.72	0.199	0.13
Experience	-0.156	-2.85	-0.058	-0.85
Marital status	0.042	0.10	-0.713	-1.15
City	0.583	2.91	0.321	1.20
Rural	-0.404	-1.34	-0.288	-0.65
Own house	0.565	1.03	-0.250	-0.24
Rent free	0.078	0.14	0.268	0.27
Renting	0.520	1.01	-0.121	-0.12
Western origin	-0.154	-0.19	0.466	0.48
Year 12	0.021	0.07	0.459	1.26
Degree	0.690	1.48	0.990	1.33
Diploma	0.478	1.19	0.595	1.59
Trade certificate	-0.069	0.23	0.406	0.70
Partner's employment	-0.471	-1.12	1.315	1.44
Health status	-0.189	-0.75	-0.307	-0.921
Children	-0.112	-0.41	-0.901	-2.12
State school	0.067	0.22	-0.179	-0.32
Manufacturing industry	-0.910	-4.19	-0.840	-1.59
Both parents	-0.125	-0.42	-0.310	-0.66
Less than 6 months	0.971	2.67	1.127	3.08
Six months	0.818	1.99	1.065	2.67
Employment history	0.417	1.68	0.038	0.14
Unemployment history	-0.196	-0.41	-1.083	-2.61
θ	1.444	5.55	1.013	2.56
ρ^a	0.485	5.42	0.249	2.05

Note: ^aThe significance of rho at the 5% level is measured using the 10% critical value, since the parameter is on the boundary of the admissible parameter space for rho if the null hypothesis (i.e. there is no unobserved heterogeneity) is true. See Arulampalam (2004).

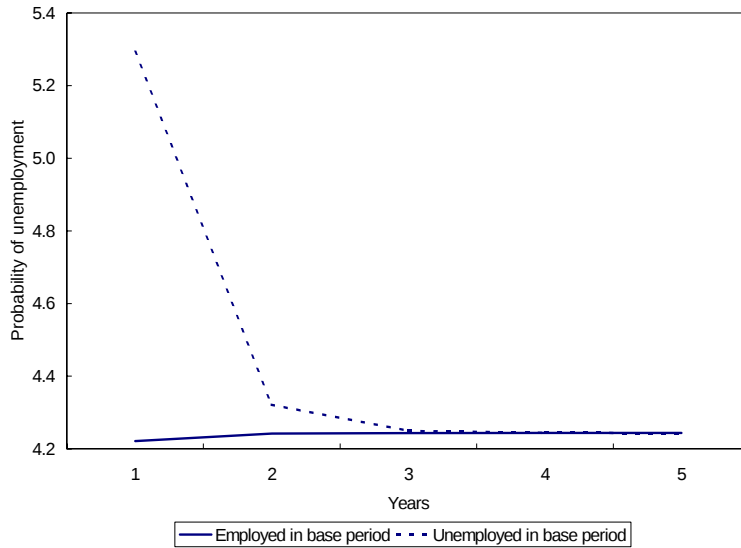


Figure 2.1: Probability of unemployment: high education females

These probabilities effectively illustrate the marginal effect of lagged employment status, with all other variables evaluated at their sample means. The marginal effects assume $E(\eta_i) = E(\varepsilon_{it}) = E(\varepsilon_{i1}) = 0$, and therefore for $t = 1$ the resulting probability is defined as $P(y_{i1} = 1) = \Phi(x'_{i1}\beta_1)$. From the second period onwards, the probability is defined as $P(y_{it} = 1) = \Phi(x'_{it}\beta + \delta y_{i,t-1})$ where the realisation of the lagged dependent variable is taken to be the estimated probability from the previous period. Essentially, this replaces $y_{i,t-1}$ with $y_{i,t-1}^*$.

Noteworthy is the apparent ‘scarring’ effect that being unemployed has on individuals. High-education females that are unemployed in the base period have a 5.3% chance of being unemployed in the first year (evaluating all other characteristics at their sample means). This declines rapidly in the second year, and the effect eventually wears off by the fourth year. As expected, those who are employed in the base period have a lower probability of being unemployed in the first year (around 4.2%), a result that remains fairly stable over time.

High-education males exhibit a similar scarring effect, although the probability of being unemployed is much lower than that for high-education females. The spread between probabilities depending on whether an individual is employed or unemployed in the base period is only slightly narrower than the female results (1.0 percentage points compared to

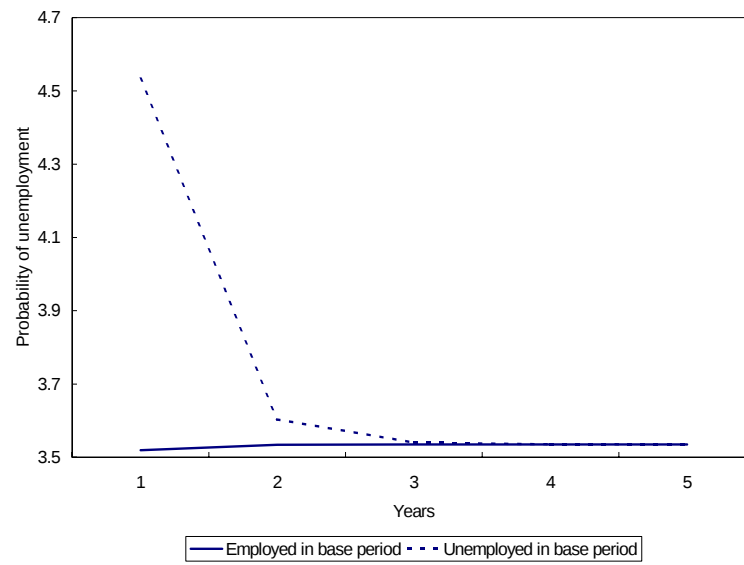


Figure 2.2: Probability of unemployment: high education males

1.1 percentage points) suggesting that high-education males are likely to suffer as much as unemployed females, relative to their employed counterparts.

2.5.2 Low-Education Equations

Table 2.5 presents the estimation results with respect to the employment status equation, for males and females with low-education.

Table 2.5: Within sample marginal effects: low education Australian born individuals, 1985-88

Variable	Male		Female	
	Effect	t-statistic	Effect	t-statistic
Constant	0.108	1.92	-0.073	-0.55
Lagged employment	0.062	3.10	0.108	3.50
Experience	0.005	0.76	0.009	0.87
Marital status	-0.032	1.17	0.032	0.24
Separated	-0.006	-0.20	-0.007	-0.05
City	0.023	1.77	0.007	0.35
Rural	-0.041	-1.89	-0.011	-0.30
Own house	0.006	0.18	0.102	0.82
Rent free	-0.088	-1.81	-0.012	-0.10
Renting	-0.011	-0.19	0.069	0.59
Year 10	0.325	1.49	0.033	1.10
Year 11	0.574	2.31	0.080	2.17
Partner's employment	0.516	1.84	-0.035	-0.28
Health status ^a	-0.335	-1.77	—	—
Children	0.075	0.49	-0.042	-1.65
Average experience	-0.129	-1.59	-0.002	-0.16

Notes: ^aA variable measuring whether or not females had a disability that limited the amount of work they could perform could not be included due to a lack of observations.

For the low-education groups, as is also the case for the high-education groups, the clearest finding is the strong direct correlation between employment outcomes over time, after controlling for observed and unobserved differences between individuals. This effect is significant, at a 5% two-tailed significance level, for both genders.

Educational levels again are significant, even within the low-education groups. The more schooling an individual has attained increases the likelihood that he or she is employed. In particular, completion of Year 11 significantly increases the probability of being employed, regardless of gender. However, there are no significant differences in job prospects among those who have completed Year 10 and below. For males, employment is also positively associated with their partner being employed.

However, males have a significantly lower probability of being employed anywhere outside the city, which is further exacerbated if they live outside a country town. In contrast to their high education counterparts, living in rent free accommodation is negatively

associated with employment. Having a disability that limits the amount of work one can perform also has a negative impact on employment prospects.

Neither individual nor group effects of experience are observed. These findings are difficult to interpret; however, it is reasonable to conclude that experience effects on employment prospects are not strong overall within the low-education groups. This applies to both the structural and initial conditions models (Table 2.6). Similar to the high education case, both ρ and θ are significantly different from zero.

Table 2.6: Initial conditions: low education Australian born individuals, 1985-88

Variable	Male		Female	
	Coefficient	t-statistic	Coefficient	t-statistic
Constant	-1.666	-1.40	-2.894	-1.75
Experience	-0.065	-1.42	-0.015	-0.23
Marital status	-0.412	-1.01	-0.863	-0.70
Separated	0.859	0.08	0.096	0.08
City	0.125	0.57	0.249	1.03
Rural	-0.270	-0.85	0.398	0.92
Own house	2.524	2.72	0.985	0.95
Rent free	0.054	0.07	0.096	0.11
Renting	0.993	1.53	0.179	0.20
Western Origin	-0.177	-0.26	1.260	2.22
Year 10	0.320	1.20	0.671	1.73
Year 11	0.344	1.18	1.030	2.45
Partner's employment	0.599	1.07	0.423	0.34
Health status ^a	0.405	1.45	—	—
Children	-0.021	-0.06	-0.284	-0.89
State school	1.069	2.16	0.970	1.00
Manufacturing industry	-0.486	-1.98	-0.387	-1.02
Both parents	0.329	1.31	0.141	0.50
Less than 6 months	-0.131	-0.42	0.853	2.23
Six months	0.133	0.40	0.694	1.80
Employment history	0.726	3.33	0.228	0.75
Unemployment history	-0.450	-1.54	-0.715	-1.93
θ	1.237	4.95	0.730	1.72
ρ^a	0.594	7.46	0.405	2.19

Note: ^aThe significance of rho at the 5% level is measured using the 10% critical value, since the parameter is on the boundary of the admissible parameter space for rho if the null hypothesis (i.e. there is no unobserved heterogeneity) is true. See Arulampalam (2004).

For females, the only variable other than lagged employment status that exhibited significant correlation with employment prospects is children. Again, this may be interpreted as evidence that employers and employees place large weight on recent employment history in deciding whether to, respectively, hire or offer labour.

Harris (1996) found evidence that the probability of being employed was - for

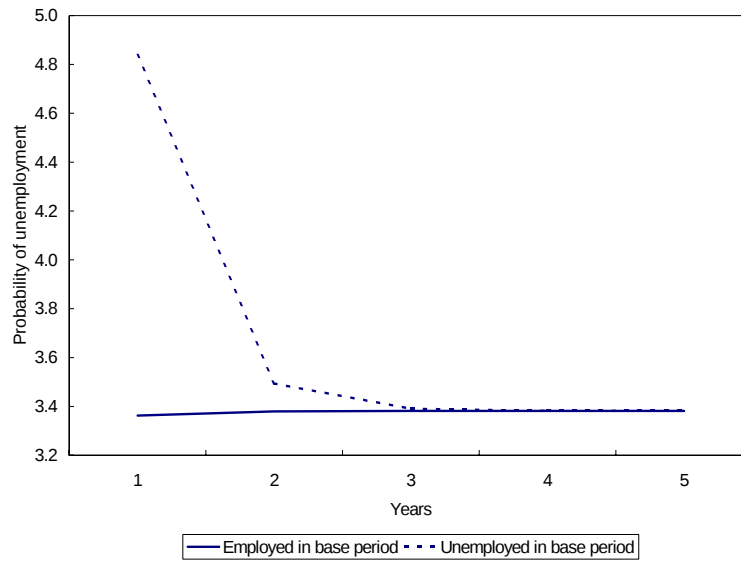


Figure 2.3: Probability of unemployment: low education females

both genders of low-education - significantly correlated with age, education, partner's employment status and existence of a disability. In addition, marital status and housing arrangements played a role for males, while females were affected by the number of children. By contrast, this chapter's results assign much lesser roles to many of these variables in explaining employment outcomes. This pattern is broadly similar to the results for the high-education groups.

Figures 2.3 and 2.4 present the probability of being unemployed depending on whether an individual is employed or unemployed in the base period. Interestingly, females who have obtained an education no further than to year 11 are *less* likely to be unemployed than their more highly educated counterparts. This holds regardless of whether they are employed or unemployed in the base period. A scarring effect is still evident, although the difference in probabilities (between being unemployed or employed in the base period) is only around 0.8 percentage points.

Low-education males appear the ones most likely to suffer from a spell of unemployment. There is almost a 9% probability that individuals who are unemployed in the base period will remain so in the subsequent year, although again, this effect falls substantially in the second year and is negligible by the fourth year. Nevertheless, once this scarring effect has worn off, the probability of being unemployed at any given time is still quite high

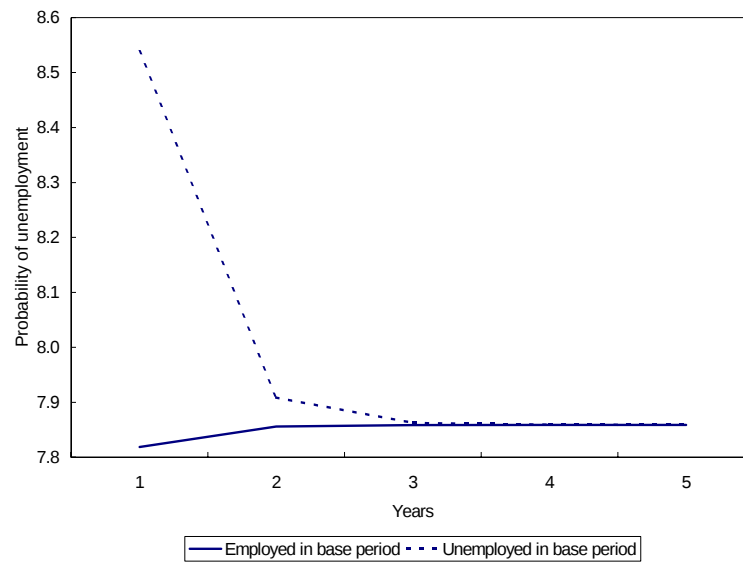


Figure 2.4: Probability of unemployment: low education males

(just over 8%).

These results are similar to Ellwood (1982) who reports that the scarring effect of unemployment on teenage males is not persistent. Individual heterogeneity was the driving force behind spells out of work, and the heaviest cost non-employment imposed was on earnings, primarily because it resulted in a lack of experience.¹³

¹³Unfortunately we are unable to test whether unemployment in the recent past has a greater scarring impact than unemployment in the more distant past because there are insufficient time periods in the data.

2.6 Conclusions

This Chapter uses a dynamic framework to analyse employment outcomes. By including a one-period lagged value of the dependent variable as an explanatory variable, it has been possible to incorporate the effect of employment history on current employment prospects. Importantly, this has been done after controlling for both observed and unobserved differences between individuals, such that the presence of true state dependence can be tested for.

Consistent estimation of the presence of state dependence depends upon the assumption that the specification of the stochastic elements of the model is correct, which we have not tested. However, the model is sufficiently general that it is not subject to problems of inconsistent estimation arising as a consequence of potential correlation between the unobserved effects (that is, unobserved heterogeneity) and the observed regressors; or, in the limiting case where such correlation approaches zero, truly exogenous individual effects. Thus, the possible sources of unobserved heterogeneity are wide. Differences in innate ability, motivation, or peer/family attitudes towards employment could all plausibly explain why individuals differ non-systematically.

The implications of such unobserved heterogeneity are that, in the absence of considerably more detailed data [for example, psychometric tests, which are controversial: see, for example, Heckman (1995)] omitted variable bias will arise in studies which fail to control for differences between respondents, to the extent that these differences are correlated with the measured regressors. It therefore is crucial to allow - in our model - for a process that generates unobserved heterogeneity.

The results indicate that prior employment status, as measured by the respondent's employment status one year ago, significantly predicts current employment outcomes across all demographic groups (see also Section 2.3.2). They also suggest that inclusion of prior employment status results in other variables, such as labour force experience, housing arrangements and marital status, being afforded a much smaller predictive role than would otherwise have been the case.

From a theoretical perspective, the study examines the issue of state dependence using a supply and demand framework. It is decided that supply and demand influences cannot be easily separated, and therefore only equilibrium outcomes should be modelled in the absence of information necessary to do so. These results are consistent with insider-outsider hypotheses of the labour market, suggesting some kind of partitioning of the labour

market with regard to prior employment status.

As noted in the introduction, it is not only the *overall* rate of unemployment which may be of interest to policy-makers, but also the *composition* of the stock of unemployed persons. It appears to be the case that those who become unemployed, for whatever reason, are likely to stay so, at least in the medium term. Therefore, significant action may be required to prevent unemployment becoming perpetually concentrated within relatively small groups. One example of policy intervention would be targetted wage subsidies, to prevent certain groups becoming unemployed in the first instance. To the extent that targetted wage subsidies will induce those previously unemployed to become employed, and to the extent that state dependence exists, then wage subsidies will not be necessary (on average) to prolong such individuals' employment. Thus, a subsidy for the (relatively short) period of time needed to induce transitions from unemployment to employment would be expected to have persistent effects on the level of youth employment, even if the subsidy were then discontinued. According to Arulampalam (2000, p.1) "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." Complementary actions could also be taken to attempt to provide advantage to those already unemployed, perhaps through schemes designed to aid formation of human capital. It would be highly desirable for future initiatives to analyse the likely interactions of such policies with the dynamic effects identified here.

Chapter 3

Towards a Dynamic Information-based Theory of Unemployment Persistence

3.1 Introduction

The results in Chapter 2 show there is evidence of both true state dependence (i.e. differences in employment outcomes caused structurally by individual employment history) and spurious state dependence (i.e. differences in employment outcomes caused by unobserved heterogeneity), in the context of the Australian youth labour market. However, the concepts of true and spurious state dependence do not provide a *theoretical* explanation for employment outcomes. Instead, they are simply different *empirical* specifications aimed at explaining typically-observed serial correlation (i.e. persistence) in individual employment observations. By contrast, to model the *causes* of persistence rather than simply their manifestations, it is necessary to explain the decision-theoretic choices which potentially lead to persistence.

This chapter constructs a decision-theoretic framework to examine the dynamic

behaviour of optimal employment decisions. The key aspect of this environment is that there exists uncertainty about workers' individual productivities. A representative firm is assumed to receive a 'noisy' signal of each worker's ability, where abilities may differ between workers, every period. The firm re-evaluates the information it has about individual workers' abilities at the start of every period, using the information contained in these signals and a Bayesian updating procedure, and then hires accordingly. We show under certain circumstances that persistent unemployment results, even for high-ability workers, due to the firm's unwillingness to hire from an unemployment pool of dubious quality.

The chapter is organised as follows. Section 3.2 reviews the relevant theoretical literature on unemployment persistence, and compares some of the approaches used with those put forward in this chapter. Section 3.3 develops a simple intertemporal model which predicts unemployment persistence among a workforce of heterogeneous abilities. Section 3.4 extends this model by relaxing a key restriction placed on the Section 3.3 model. Section 3.5 draws together conclusions from the chapter and outlines directions for future research.

3.2 Background

3.2.1 Traditional Explanations

Two explanations have commonly been used within the theoretical literature on persistent employment outcomes: depreciation of human capital, and intertemporally non-separable preferences. The former, as set out in Heckman (1991), proposes that the event of being unemployed may lead to a slowdown or reversal in the growth of human capital. The latter, described in Hotz *et al* (1988), allows preferences about each future period's consumption-leisure trade-off to depend upon the amount of leisure (i.e. unemployment) chosen in the present and all past periods; this implies that unemployed workers today will attach less weight to future consumption (relative to future leisure) than will their employed counterparts; so on average, those currently unemployed will seek fewer hours of future employment than those currently employed. Like the human capital explanation, this assumes complete information about workers' marginal products and reservation wages. However, unlike the human capital explanation, the source of persistence comes from individual workers changing their own utility values of leisure and so supplying fewer labour hours, rather than from firms demanding fewer labour hours due to the changing productivity values of individual workers.¹

Both explanations incorporate something about individual characteristics changing once a worker moves between the states of employment and unemployment. In the case of human capital depreciation, the relevant change affects the demand for their labour, while for intertemporally nonseparable preferences it is labour supply which changes.

¹Nonetheless, the explanation based on preferences is also capable of taking on a human capital interpretation: if we think of unemployed individuals accumulating leisure-specific human capital (e.g. greater efficiency in caring for children), then such individuals may become increasingly unwilling to re-enter employment.

3.2.2 Information-Based Explanations

In this chapter, we put forward a third explanation in which workers' characteristics do not change according to current state. Instead, we focus attention on the information available to the firm. In particular, we relax the assumption maintained by the other two explanations that complete information exists about workers' marginal products. If a firm cannot costlessly distinguish workers of different productivities, then it may use past unemployment records to sort workers. A worker who becomes unemployed, even though their marginal product may be relatively high, will find it difficult to return to employment since employers will infer them to be of low productivity by the fact they are unemployed.

To consider such an idea, it is natural to ask who the decision-makers are. In particular, is it only one firm that is trying to sort workers by their respective marginal products; and if not, how will workers and other firms react in the face of this? One approach, used by Montgomery (1999) is to consider the simplest case of no other firms and passive workers.² The firm hires from an unemployment pool; of this pool, it knows the composition in terms of worker ability, but does not know the ability of any particular individual in it before hiring. If the firm decides to hire from the unemployed pool, it observes a noisy (i.e. inaccurate) index of productivity for each worker and updates its beliefs about individual abilities based on such observations. It then decides whether to release any or all of its currently employed workers, optimally releasing those whom it has most reason to believe are of low ability. In the absence of any employment/re-employment frictions, the ability composition of the unemployed pool will worsen over time, which leads to a 'lemons' effect similar to that described by Akerlof (1970); namely, high ability unemployed workers will become increasingly disadvantaged through association with the increasing proportion of the unemployed who are of low ability. Under such circumstances, eventually the firm will simply stop rehiring from the unemployed pool; or, if there are exogenous retirements from employment, simply to use a stationary hiring rule. However, it is shown that in the presence of fixed hiring costs the firm will delay rehiring and firing until "sufficient" workers have retired to make rehiring profitable. Thus, endogenous employment cycles may result from the underlying information imperfection in this environment.³

²The assumption of no other firms may be alternatively viewed in the context of a 'representative firm' among many in the macroeconomy.

³The model including fixed hiring costs is similar in some ways to the literature of search models in employment since both analyse labour market frictions: see, for example, Zaretsky and Coughlin (1995), Mortensen (1989), Hosios (1990), Keller, Roberts and Stevens (2007). However, the focus is slightly different in a model where the key imperfection is the absence of information rather than the cost of acquiring

The model of Montgomery (1999) does not allow for any strategic interaction between decision-makers⁴. This is a limitation, since one of the major difficulties with such a dynamic model is that it does not explain, in a wider context, the absence of strategic downward pressure on wages exerted by high productivity unemployed individuals. Presumably, firms learn more quickly about employed workers than about unemployed workers, so why do not high ability unemployed workers strategically underbid their low ability companions with an eye to “proving” themselves over the course of time? There is empirical evidence [e.g. Ong and Mar (1992)] that wage profiles of re-employed workers rise steeply over time relative to those in continuous employment. This may be partly responsible for the relative dearth of such dynamic unemployment persistence models; indeed, we are not aware of any others within this class of fixed-wage models which do not incorporate strategic interaction.

A small number of frictionless, fixed wage, models incorporate both incomplete or imperfect information *and* strategic interaction; that is, treat both workers and firms as solving some optimisation problem. In Montgomery (1991), workers have heterogeneous abilities known to them but not to the firm. The crucial addition which permits strategic interaction is that each worker’s ability is also known to other workers. To utilise this information, the firm can choose to solicit information from existing workers about whom to hire; or, alternatively, can choose to simply hire from the outside labour market as a whole. Paraphrasing, the firm can choose between informal and formal hiring channels. In turn, each existing worker - if asked - can decide whether to refer a high ability worker, or simply to refer some member of their loosely defined social network. Although the model as it is written does incorporate wage determination to an extent, it is not difficult to see how this idea of social networks could be incorporated in a fixed wage setting. If workers are penalised in some way (perhaps through loss of reputation, or being moved forward in a queue to be laid off if aggregate conditions worsen) for referring workers of low ability, and analogously rewarded for referring those of high ability, a bilateral equilibrium between workers and firms can be established. In such an equilibrium, some workers will be continually unemployed because they are not in the right social network, *not* because their ability is low. As in Montgomery (1999), the fundamental labour market imperfection is the individual worker’s inability to signal credibly their true ability, and this can be exacerbated by the existence of social networks.

information.

⁴By contrast, Rodriguez-Planas (2009) incorporates the possibility that workers may strategically signal high productivity by choosing to be unemployed rather than to accept a low-paid job. In such a context, unemployment is not scarring but rather the reverse.

In Lazear (1999), information asymmetries arise in the context of the profitability of a particular enterprise, known to workers but not to a potential investor. This can be modified in our context by thinking of a worker as providing “enterprise-like” services over which they know the average profitability. It is shown that, if wages can be tied to output, then any truth-telling equilibrium where workers honestly report the profitability of the enterprise entails some linking of wages to profits. The motivation of this study is different from ours: it seeks to explain why profit-based pay can arise even in the absence of moral hazard concerns about the intensity of worker effort. However, the fundamental idea is that informational asymmetries make it necessary for firms to penalise workers for not reporting their information truthfully. Therefore, *ex ante* they commit to a policy which lays off workers (which may also brand them with a loss of reputation) if their future observed profitabilities are below some threshold level. Although this may not necessarily be *ex post* efficient, the firm must nonetheless credibly commit to this policy in order to induce favourable worker selection in the initial hiring stage. Thus, even though a worker may be of high average profitability, they may nonetheless be dismissed and branded indefinitely as unprofitable due to bad luck and the inability to signal their true average profitability. It may be seen that this fundamental inability underlies all of the fixed wage, non-search models involving information imperfections discussed here.⁵

⁵In a similar vein, MacLeod and Malcomson (1988) model the effects of imperfect information in *ranking* employees within a hierarchical structure under fixed wages.

3.2.3 Evaluation Criteria

In light of the discussion in Section 3.2.2 above, consider the following list of questions by which one can assess dynamic employment models.

1. Are wages and/or job characteristics endogenous?
2. What is the nature of job-matching technology?
3. Are both workers and firms fully-optimising agents?
4. Is there a role for competing firms to induce worker turnover?
5. What is the information structure for each of workers and firms?
6. If information is imperfect, what is the nature and scope of dynamic learning?

In Section 3.3 below, we develop a model which incorporates some of these considerations. The cause of state dependence in this model is, like most of those in Section 3.2.2 above, incomplete information about worker ability. Proceeding in a fixed-wage context, the model analyses the effects upon a firm's optimal hiring decisions of not being able to distinguish between two types of workers of different productivities. In spirit, our model has much in common with those in Section 2.2. However, there are crucial differences between our model and each of these. These differences are intended to focus attention upon the role that acquiring information about workers' productivities plays in determining individual employment outcomes.

3.3 Simple Model

We develop a baseline model incorporating incomplete information, which incorporates the fundamental characteristics of our theoretical approach.

Consider a single firm which hires from a labour force with fixed size where:

γ_t is the proportion of the labour force employed in period t

p_{nt} is the realised productivity of worker n employed in period t

$\bar{p}_t$ is the expected average productivity of workers employed in period t

In this model, the total size of the labour force is normalised to be equal to 1. Workers are intrinsically endowed with a time-invariant productivity characteristic θ , so heterogeneity is introduced by high productivity workers having $\theta = \theta_H$, while low productivity workers have $\theta = \theta_L$. Intrinsically, high productivity workers are assumed to have the ability to produce 1 unit of output per period, while low productivity workers can produce 0 units.⁶ However, the process linking intrinsic productivity to realised output is noisy, and it is only with probability q_j that any worker of type θ_j , where $j \in \{H, L\}$, will actually realise their intrinsic productivity level. With probability $1 - q_j$, a worker of type θ_j realises output at the *other* type's intrinsic level.

Proportion h of the labour force are of high productivity type. Thus, for these workers of type $\theta = \theta_H$,

$$E[p_{nt} \mid \theta = \theta_H] = q_H \cdot 1 + (1 - q_H) \cdot 0 = q_H \quad (3.1)$$

Similarly, for the proportion $(1 - h)$ of workers who are of type $\theta = \theta_L$,

$$E[p_{nt} \mid \theta = \theta_L] = q_L \cdot 0 + (1 - q_L) \cdot 1 = 1 - q_L \quad (3.2)$$

Since workers of type $\theta = \theta_H$ are assumed to have higher intrinsic productivity than those of type $\theta = \theta_L$, it is natural (and involves no loss of generality) to restrict all q_j values to be not less than one half.⁷

⁶Such as, for example, in preparing tenders to win a project, where the outcome is either a 'success' of some value or a "failure" worth nothing.

⁷If this were not the case, so that q_j were *less* than $\frac{1}{2}$ for some type j , then if one treated the inspection information as 'truthful', the estimate of θ_j would actually be *less* accurate than by ignoring the information; thus, it would be an unusual inspection process. However, even if this were the case, it would not present any problems since the information revealed by $q_j = \hat{q}_j < \frac{1}{2}$ is **equivalent** to that revealed by $q_j = 1 - \hat{q}_j > \frac{1}{2}$, for any $\hat{q}_j < \frac{1}{2}$. This can most easily be seen by considering $q_H = q_L = 0$, so that high productivity workers always produce nothing and low productivity workers always produce 1 unit. All information about a worker's type is then revealed by observing their output, so this is equivalent to the case where $q_H = q_L = 1$.

During each period t in this model, the firm observes an individual output for each worker for period t . Then at the end of period t , the firm decides which workers to rehire and which workers to release for its period $t + 1$ production requirements, basing this decision upon the individual workers' outputs observed during period t .

We assume the wage per worker paid by the firm is set exogenously at a level less than 1 unit of value, so that high productivity workers are profitable to the firm. It is also assumed the firm has a target workforce size $\gamma_t = \gamma^*$, $\forall t$. This may be thought of as due to a fixed coefficients Leontief-type production function where, additionally, the firm's capital stock is fixed. In Section 3.3.1 below, we assume the ability composition of the 'pool' of workers outside the firm is fixed over time. In Section 3.3.2 below, we examine the effects of different assumptions about the relationship between the firm's hiring decisions and the composition and size of the 'pool' of workers outside it. For example, if a sufficiently large pool of unemployed workers always exists, then the firm will be unconstrained in its hiring decisions; this guarantees that the firm will be able not only to achieve its target workforce size every period, but also to turn over as many workers as it wishes to while achieving this. The opposite case is that where the firm may wish to release more workers than are currently unemployed during some or all time periods, in which case such a firm will be constrained with respect to either its size or the extent of its turnover.

The model makes strong simplifying assumptions about technology and capital mobility. In this spirit, we also ignore any time discounting by the firm. These assumptions, although restrictive, allow us to observe the role that asymmetries in productivity information can play in determining employment outcomes.

The same equivalence applies for all other values of $q_j < \frac{1}{2}$; thus, as observed in the text, the restriction of $q_j \geq \frac{1}{2}, \forall j$ involves no loss of generality.

3.3.1 Exogenous Unemployment Composition

Consider the situation which the firm faces if it can inspect low productivity (θ_L) workers perfectly, but receives a noisy signal for high productivity (θ_H) workers. This special case is chosen to enable comparison of our results with those of Montgomery (1999), who uses this assumption in a broadly similar context; it also aids understanding of the processes we wish to model, in an even more simplified setting, which will be generalised somewhat in subsequent models considered. Further assume that, during period 1 and every subsequent period, the firm can employ at will from an outside pool of workers, where the proportion of θ_H workers is always equal to that proportion of θ_H workers which exists in the labour force as a whole h . To paraphrase, the composition of the outside pool of workers - which we might think of as being unemployed - is constant over time and unaffected by the firm's previous hiring and firing decisions. This is clearly unrealistic since, if the firm discharges workers in other than the exact proportions they are already represented in among the unemployed, the composition of the unemployed will necessarily change. However, this may be a useful approximation for situations where either labour turnover is very low, or where a very large pool of unemployed continues to exist over a long period of time. It is also useful for the purpose of assisting the reader to understand the workings of the most basic form of the model.

As these two extra assumptions lead to a special case of the model with $q_L = 1$ then, using (3.1), at the end of period 1 the firm has γ^* workers employed, with output per worker of

$$\bar{p}_1 = hq_H \quad (3.3)$$

From the firm's perspective, it is able to partition the γ^* workers into 2 distinct groups according to its observations of output during period 1. One group (which we call group A) has high realisations of output attributable to its members: all of them produce 1 unit of output, and all of these are of high intrinsic productivity type (i.e. of type θ_H).⁸ Another group, B, has low realisations of output, and all produce nothing. However, because of the underlying noisy signals of output, not all of this group are of low intrinsic productivity type θ_L ; some of group B are merely type θ_H workers that were 'unlucky' this period. This can be quantified by noting that, if the composition of types among employed workers is

⁸Since q_L is assumed to equal 1, type θ_L workers *always* produce nothing, so will never be included in Group A.

the same as that among the labour force as a whole:

$$\begin{aligned}\Pr[\theta = \theta_H \mid p_{nt} = 1] &= \frac{hq_H}{hq_H + (1-h)(0)} \\ &= 1\end{aligned}\tag{3.4}$$

$$\begin{aligned}\Pr[\theta = \theta_L \mid p_{nt} = 1] &= 1 - \frac{hq_H}{hq_H + (1-h)(0)} \\ &= 0\end{aligned}\tag{3.5}$$

$$\begin{aligned}\Pr[\theta = \theta_H \mid p_{nt} = 0] &= \frac{h(1-q_H)}{h(1-q_H) + (1-h)(1)} \\ &= \frac{h(1-q_H)}{1-hq_H}\end{aligned}\tag{3.6}$$

$$\begin{aligned}\Pr[\theta = \theta_L \mid p_{nt} = 0] &= 1 - \frac{h(1-q_H)}{h(1-q_H) + (1-h)(1)} \\ &= \frac{1-h}{1-hq_H}\end{aligned}\tag{3.7}$$

These are the individual type probabilities, conditioned upon the current period's output realisation, for any given worker.

Of course, if the composition of employed workers were different from that of the labour force as a whole, and this were known, then the firm would also take advantage of this additional information. We return to this shortly. However, in the meantime note that, for period 1, the proportion of each type in employment is the same as in the labour force as a whole.

Since it now knows the conditional probabilities over types, the firm is now able to calculate the expected future output of any *employed* worker based on their period 1

output realisation, as follows:

$$\begin{aligned}
E[p_{n2} \mid p_{n1} = 1] &= \Pr[\theta = \theta_H \mid p_{n1} = 1] \cdot E[p_{n2} \mid \theta = \theta_H] \\
&\quad + \Pr[\theta = \theta_L \mid p_{n1} = 1] \cdot E[p_{n2} \mid \theta = \theta_L] \\
&= 1 \cdot q_H + 0 \cdot (1 - 1) \\
&= q_H
\end{aligned} \tag{3.8}$$

$$\begin{aligned}
E[p_{n2} \mid p_{n1} = 0] &= \Pr[\theta = \theta_H \mid p_{n1} = 0] \cdot E[p_{n2} \mid \theta = \theta_H] \\
&\quad + \Pr[\theta = \theta_L \mid p_{n1} = 0] \cdot E[p_{n2} \mid \theta = \theta_L] \\
&= \frac{h(1 - q_H)}{1 - hq_H} \cdot q_H + \frac{1 - h}{1 - hq_H} \cdot (1 - 1) \\
&= \frac{h(1 - q_H)}{1 - hq_H} \cdot q_H
\end{aligned} \tag{3.9}$$

An information updating has thus occurred for employed workers, so that on average the firm has a more accurate idea of each worker's intrinsic productivity at the end of Period 1 than at its start.⁹ By contrast, the firm's information about the productivity of unemployed workers does not change during Period 1.

$$\begin{aligned}
E[p_{2n} \mid \text{unemployed in period 1}] &= hq_H + (1 - h)(1 - 1) \\
&= hq_H
\end{aligned} \tag{3.10}$$

Given the firm's updated information, what is its optimal Period 2 hiring decision? Clearly, this will depend on the expected future productivities - respectively - of high realisation workers (which we call Group A), low realisation workers (Group B) and the unemployed.¹⁰ This is addressed immediately below.

Proposition 1.

After Period 1, Group A workers have a higher expected future productivity than the currently unemployed; who, in turn, have a higher expected future productivity than Group B workers.

⁹Note that a special case of this model is that of complete information (ie. $q_H = q_L = 1$). In that case, $E[p_{n2} \mid p_{n1} = 0]$ simply equals zero as the firm infers worker type from observed output perfectly.

¹⁰Although it is tempting to use Groups "H" and "L" instead to denote current output *realisations*, this would be potentially confusing, since these are the same labels we use for *ability*. It is important to clearly distinguish between these two concepts, since the latter is a noisy signal of the first; this explains the non-mnemonic choice of notation.

Proof.

The proof proceeds in two parts. The left part of the inequality is satisfied by noting that

$$\begin{aligned}
 E[p_{2n} \mid p_{1n} = 1] - E[p_{2n} \mid \text{unemployed in period 1}] &= q_H - hq_H \\
 &= q_H(1 - h) \\
 &> 0
 \end{aligned}$$

Satisfaction of the right part of the inequality is along similar lines. We may note that

$$\begin{aligned}
 E[p_{2n} \mid \text{unemployed in period 1}] - E[p_{2n} \mid p_{1n} = 0] &= hq_H - \frac{h(1 - q_H)}{1 - hq_H} \cdot q_H \\
 &= hq_H \left[\frac{1 - hq_H - 1 + q_H}{1 - hq_H} \right] \\
 &= \frac{hq_H^2(1 - h)}{1 - hq_H} > 0
 \end{aligned}$$

which completes the proof.

■

Proposition 1 shows that, in the absence of turnover costs, the firm has a clear rationale for releasing all workers in Group B and replacing them with the same number from the Period 1 unemployed, while re-employing all workers in Group A. Indeed, such a plan could, in principle, be implemented at the end of any period, which brings us to Definitions 1 and 2.

Definition 1 A *hire-fire policy* in a given period is a plan by the firm to fire all workers with a low output realisation at the end that period.

Definition 2 A *repeated hire-fire policy* is a plan by the firm to use a hire-fire policy every period.

Under a hire-fire policy in period 1, released (i.e. "fired") workers will be all from Group B:

$$\begin{array}{ll} \gamma^* h (1 - q_H) & \text{workers of type } \theta_H \\ \gamma^* (1 - h) & \text{workers of type } \theta_L \end{array}$$

amounting to a total of $\gamma^* [h (1 - q_H) + (1 - h)]$ released workers. This measure of workers will be hired from the unemployed pool. The expected composition of the firm's period 2 workforce will therefore be

$$\begin{array}{ll} \gamma^* h q_H & \text{reemployed workers of type } \theta_H \\ 0 & \text{reemployed workers of type } \theta_L \\ \gamma^* h [h (1 - q_H) + (1 - h)] & \text{newly employed workers of type } \theta_H \\ \gamma^* (1 - h) [h (1 - q_H) + (1 - h)] & \text{newly employed workers of type } \theta_L \end{array}$$

so that $\gamma_1 = \gamma_2 = \gamma^*$ and the firm's target workforce size γ^* is achieved in both periods 1 and 2.

Now, the average productivities, respectively, of period 1 and 2 workers will be

$$\bar{p}_1 = h q_H \tag{3.11}$$

$$\begin{aligned} \bar{p}_2 &= \left[\frac{\gamma^* h q_H + \gamma^* h [h (1 - q_H) + (1 - h)]}{\gamma^*} \right] q_H \\ &= h q_H [q_H - h q_H + 1] \end{aligned} \tag{3.12}$$

which leads to the next inequality we can demonstrate.

Corollary 1.

If a hire-fire policy is used in period 1, the proportion of type θ_H workers in the firm's workforce is higher in Period 2 than in Period 1.

Proof.

Under a hire-fire policy, the proportion of type θ_H workers in the firm's workforce during Period 2 is

$$h q_H + h [h (1 - q_H) + (1 - h)] = h + q_H (1 - h)$$

which is higher than the proportion h of type θ_H workers which were in the firm's workforce during Period 1.

■

Corollary 1 is not surprising: the firm is simply using its additional information to raise average productivity by discarding workers who - on average - are less productive.

Of course, all discarded workers are not equally productive. Within our framework, $\gamma^*h(1 - q_H)$ workers of high intrinsic productivity type θ_H have been discarded purely due to an unlucky draw in period 1. This may be viewed as a detrimental, but not necessarily persistent, random shock to their productivity. If we want to use this framework to analyse persistence of unemployment outcomes, the logical question is: **How disadvantaged are these unemployed type θ_H workers, relative to type θ_L workers employed in period 2?** We now turn to this question.

To analyse this question, and others like it, it is necessary to more completely consider the evolution of the firm's labour force over time under a repeated hire-fire policy. For this purpose, we define two new variables.

Let α_t denote the expected proportion of type θ_H workers in employment with the firm during period t under a repeated hire-fire policy; and,

Let T_t denote the expected proportional turnover within the firm's workforce at the end of period t under a repeated hire-fire policy.

For example, in Period 1 we have $\alpha_1 = h$ and $T_1 = (1 - h) + h(1 - q_H)$. We are interested in the behaviour of α_t over time. The following proposition establishes this.

Proposition 2.

Under a repeated hire-fire policy beginning in Period 1, and for all $t \geq 1$:

$$\alpha_t = h \left[\frac{1 - [q_H(1 - h)]^t}{1 - q_H(1 - h)} \right]$$

Proof.

The proof is in 3 parts. In period 1, by assumption, the firm selects a representative sample of workers from the labour force with proportion of θ_H workers h ; this proves that $\alpha_1 = h$. Now, under a hire-fire policy, at the end of any period t the firm fires all workers with low output realisations; in a workforce with proportion of θ_H workers α_t turnover is

thus:

$$\begin{aligned} T_t &= \alpha_t (1 - q_H) + (1 - \alpha_t) \\ &= 1 - q_H \alpha_t \end{aligned}$$

If proportion T_{t-1} of the period $t - 1$ workforce were fired and hired, then proportion $(1 - T_{t-1})$ of the period $t - 1$ workforce must be reemployed during period t . All re-employed workers must be of type θ_H ; in addition, proportion h of those hired for period t must also be of type θ_H . Thus, for any period $t > 1$:

$$\begin{aligned} \alpha_t &= q_H \alpha_{t-1} + h T_{t-1} \\ &= q_H \alpha_{t-1} + h [1 - q_H \alpha_{t-1}] \\ &= h + \alpha_{t-1} [q_H (1 - h)] \end{aligned}$$

By repeated substitution of α_{t-1} and the observation that $\alpha_1 = h$, we have

$$\alpha_t = h + h q_H (1 - h) + h [q_H (1 - h)]^2 + \dots + h [q_H (1 - h)]^{t-2} + h [q_H (1 - h)]^{t-1}$$

which is a geometric series in t with initial term h and common ratio $q_H (1 - h)$; thus, for all $t \geq 1$, α_t can be written as

$$\alpha_t = h \left[\frac{1 - [q_H (1 - h)]^t}{1 - q_H (1 - h)} \right]$$

as stated.

■

Corollary 2.

For parameter values h and q_H the dynamic process for α_t under a repeated hire-

fire policy monotonically converges from below to

$$\frac{h}{1 - (1 - h) q_H}$$

Proof.

As $t \rightarrow \infty$,

$$\begin{aligned} \alpha_t &\rightarrow h \left[\frac{1 - 0}{1 - (1 - h) q_H} \right] \\ &\rightarrow \frac{h}{1 - (1 - h) q_H} \end{aligned}$$

which establishes convergence. Since $\frac{d\alpha_t}{dt} = -\frac{h}{1 - q_H(1 - h)} \ln [q_H (1 - h)] \exp [t \ln [q_H (1 - h)]]$, which is positive for all $t \geq 1$, then the convergence is monotonic and from below, as stated.

■

Proposition 2 and Corollary 2 show that a firm which follows a repeated hire-fire policy will increase the proportion of high quality workers employed every period. This is not a globally optimal policy: for example, if the firm keeps records of individual productivity outcomes for more than one period, it may be able to sort its workforce more effectively and raise average worker productivity further still. See Sections 3.4 and 3.5 below for further discussion of this point. However, to the extent that the firm cannot access or use the full history of productivity outcomes, a repeated hire-fire policy is the firm's optimal policy in the context of an outside unemployment pool which is fixed in composition over time.

3.3.2 Endogenous Unemployment Composition

As stated at the beginning of Section 3.3 above, in situations where the firm releases significant numbers of workers every period it is unrealistic to model newly hired workers as always being from a pool of exogenously fixed composition. Note that, although we refer to a single firm in the exposition, the same arguments apply to a labour market characterised by many identical firms, all following the same dynamic hiring strategy. To allow for more realistic dynamics, we turn to a slightly richer environment. The model is still “simple” in the sense we maintain the assumption of a noiseless process governing the output realisations of low ability workers (i.e. $q_L = 1$, so only high ability workers are affected by a noisy process). However, we now explicitly model the endogenous dynamics of the unemployed pool of workers under particular hiring policies, while assuming that the firm conditions its hiring decisions only upon known average labour force composition and the current period’s worker-specific outputs.

Continuing to assume that the firm’s initial draw of workers is in line with the overall composition of the labour force, both the firm and the unemployed pool in Period 1 have proportion h of type θ_H workers. Thus, we have $\alpha_1 = h$ and $\bar{p}_1 = hq_H$, as before. The firm optimally desires to release all Group B workers at the end of Period 1, as Proposition 1 from Section 3.3.1 still applies here. We now need to address the question of whether there are in fact enough unemployed workers to allow the firm to completely turn over all its Group B workers. As before, there will be measure $\gamma^*T_1 = \gamma^*(1 - hq_H)$ of Group B workers at the end of Period 1. Therefore, there will be some threshold value of firm size relative to the unemployed pool below which (and only below which) the firm will be able to make the full changes it desires. We therefore introduce the following definition.

Definition 3 *A **small firm in Period** t is a firm that is able to implement a repeated hire-fire policy for all periods up to and including Period t . A **large firm in period** t is a firm that is not a small firm in period t .*

Small Firm

Using Definition 3, a small firm in Period 1 must be one that can find at least measure $\gamma^* (1 - hq_H)$ of workers within the unemployed pool; since the actual measure in the unemployed pool at the end of Period 1 is equal to $1 - \gamma^*$, this is equivalent to stating that a small firm in Period 1 is one for which

$$\gamma^* \leq \frac{1}{2 - hq_H}$$

For general t , we can characterise a small firm according to the following condition.

Condition 1 *A firm is a small firm in period t if and only if*

$$\gamma^* \leq \tilde{\gamma}^* \equiv \frac{1}{1 + T_i}, \forall i \leq t \quad (3.13)$$

where Condition 1 follows from noting that, under a repeated hire-fire policy, the firm will seek to hire measure $\gamma^* T_t$ workers from an unemployed pool containing measure $1 - \gamma^*$, in each and every Period t . We may refer to this as the *small firm condition*.

For the purposes of this analysis, we first look at the case where the firm is a small firm in Period 1, while making no assumptions about later periods. We then examine the contrary case where the firm is a large firm in Period 1.

For a small firm at the end of Period 1, we know that the workers it hires from the unemployed pool have proportion h of type θ_H workers. Therefore, as in the earlier analysis where we additionally assumed this applies during all periods, the optimal composition of the firm's Period 2 workers is still exactly characterised by Proposition 2 from Section 3.3.1. Thus, we can still write:

$$\begin{aligned} \alpha_1 &= h \\ \alpha_2 &= h \left[\frac{1 - [q_H (1 - h)]^2}{1 - q_H (1 - h)} \right] \\ &= h [1 + q_H (1 - h)] \\ T_1 &= 1 - hq_H \end{aligned}$$

However, the measure of workers $\gamma^* (1 - hq_H)$ discharged by the firm at the end of

Period 1 does *not* have the same composition as the measure of workers already unemployed in Period 1. In particular, among these newly-discharged workers $\gamma^*h(1 - q_H)$ will be of type θ_H and the other $\gamma^*(1 - h)$ will be of type θ_L . Therefore, proportion $h \left[\frac{1 - q_H}{1 - hq_H} \right]$ of newly-fired workers are of type θ_H : this is necessarily a lower proportion than that among the Period 1 unemployed, h . Thus, the unemployed pool is *worsening* in terms of its average intrinsic productivity level; this is not surprising, since the firm is using the information gleaned from inspection of workers to send back to unemployment those who, on average, are of lower intrinsic productivity level. This will, as time goes on, discourage the firm from turning over its workforce more and more, and may eventually cause persistence.

To examine this formally, we now need a dynamic variable on the composition of the unemployed to go with those which describe, respectively, the composition of the employed (α_t) and turnover (T_t). We define this new variable immediately below.

Let β_t denote the average proportion of type θ_H workers in the unemployment pool during period t under a repeated hire-fire policy.

Our earlier analysis with constant unemployment composition can therefore be viewed as a special case of this analysis with $\gamma^* = 0$, so that $\beta_t = h, \forall t$.¹¹ Under this more general specification, however, we need to take into account how the firm's systematic reintroduction of (on average) less intrinsically productive workers into the unemployed pool will affect its composition. To evaluate this, we note that the population of high quality workers are divided into two groups: those who are employed by the firm and those who are unemployed. Proceeding to decompose the population of high quality workers in this way gives:

$$h = \alpha_t \gamma^* + \beta_t (1 - \gamma^*) \quad (3.14)$$

For any Period t , the composition of the firm's employed workers α_t depends upon both the previous period's employment composition α_{t-1} and the previous period's composition of the unemployed β_{t-1} :

$$\alpha_t = q_H \alpha_{t-1} + \beta_{t-1} T_{t-1} \quad (3.15)$$

Note that this expression reduces to the corresponding expression for α_t with constant

¹¹More meaningfully, as $\gamma^* \rightarrow 0, \beta_t \rightarrow h, \forall t$. (The firm does not actually exist if $\gamma^* = 0$).

unemployment composition, by setting $\beta_t = h$.

Using this expression and substituting for β_{t-1} gives:

$$\alpha_t = q_H \alpha_{t-1} + \frac{1}{1 - \gamma^*} [h - \alpha_{t-1} \gamma^*] [1 - q_H \alpha_{t-1}]$$

So the change in employment composition between Periods $t - 1$ and t can be written as a quadratic function of composition in Period $t - 1$:

$$\alpha_t - \alpha_{t-1} = f(\alpha_{t-1})$$

where

$$f(\alpha_{t-1}) = q_H \left(\frac{\gamma^*}{1 - \gamma^*} \right) \alpha_{t-1}^2 + \left(q_H - 1 - \frac{\gamma^*}{1 - \gamma^*} - \frac{h q_H}{1 - \gamma^*} \right) \alpha_{t-1} + \frac{h}{1 - \gamma^*} \quad (3.16)$$

We are interested in the behaviour of this function on the domain between $\alpha_{t-1} = 0$ and $\alpha_{t-1} = 1$. At the boundary points, $f(0) = \frac{h}{1 - \gamma^*} > 0$, while $f(1) = \frac{-(1-h)(1-q_H)}{1 - \gamma^*} < 0$. Furthermore, the slope of the function is

$$\frac{df(\alpha_{t-1})}{d\alpha_{t-1}} = 2q_H \left(\frac{\gamma^*}{1 - \gamma^*} \right) \alpha_{t-1} + \left(q_H - 1 - \frac{\gamma^*}{1 - \gamma^*} - \frac{h q_H}{1 - \gamma^*} \right) \quad (3.17)$$

so the turning point occurs where this slope is zero:

$$\begin{aligned} \alpha_{t-1}^{TURNINGPOINT} &= \frac{-\left(q_H - 1 - \frac{\gamma^*}{1 - \gamma^*} - \frac{h q_H}{1 - \gamma^*}\right)}{2q_H \left(\frac{\gamma^*}{1 - \gamma^*}\right)} \\ &= \frac{1 + q_H(\gamma^* + h - 1)}{2\gamma^* q_H} > 1 \end{aligned} \quad (3.18)$$

Using (3.17) and (3.18) together with the boundary values shows that, since the turning point occurs for a value of α_{t-1} greater than 1, therefore the function $f(\alpha_{t-1})$ is monotonically decreasing in α_{t-1} for all admissible values of α_{t-1} and crosses the horizontal axis in this range as shown by Figure 3.1 below.

As $f(\alpha_{t-1})$ describes the change in employment composition for a firm which uses

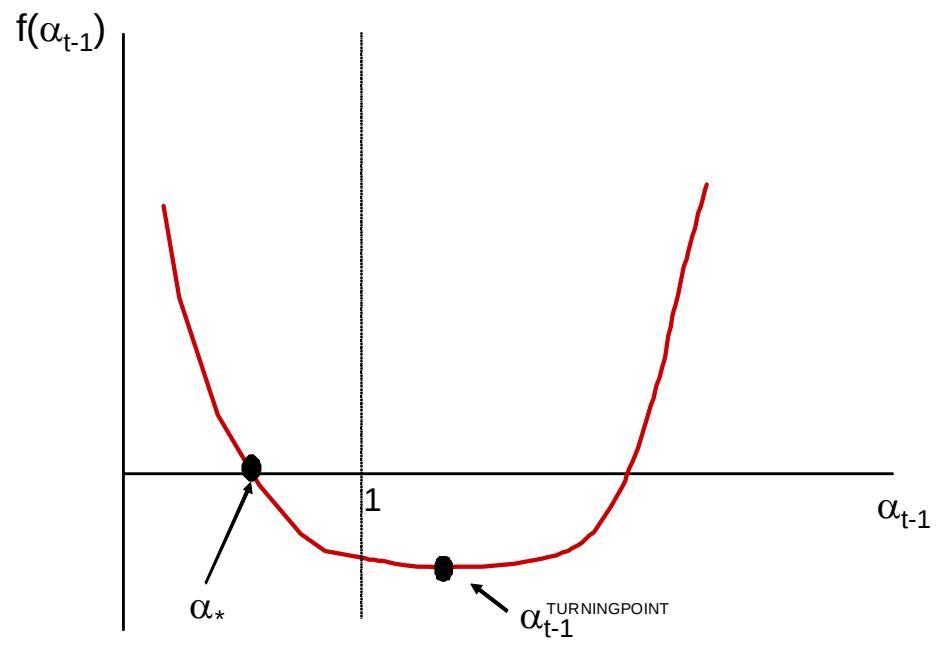


Figure 3.1: Composition of firm's labour force using a hire-fire policy

a repeated hire-fire policy, it is clear that there is some critical value of α_{t-1} beyond which employment composition will worsen (i.e. beyond which $f(\alpha_{t-1})$ becomes negative). We denote this critical value by α_* as shown in Figure 3.1 and note that α_* lies strictly in the range between 0 and 1.

Where $\alpha_{t-1} > \alpha_*$ this raises the possibility that the firm does not find it profitable to use a hire-fire policy at the end of period $t - 1$, since a hire-fire policy here worsens composition for period t . However, the optimal dynamic policy at period $t - 1$ does not only depend on whether a policy is optimal for period t , since the effects of the policy in subsequent periods $t + 1, t + 2, \dots$ also need to be considered. This issue is addressed further below. However, to the extent that the optimal dynamic policy where $\alpha_{t-1} > \alpha_*$ involves no further hiring in period $t - 1$, then this is a form of unemployment persistence since none of those unemployed in period $t - 1$ will be employed during period t .

The critical value α_* is characterised as follows:

$$\begin{aligned} f(\alpha_{t-1} \mid \alpha_{t-1} = \alpha_*) &= 0 \\ &= q_H \left(\frac{\gamma^*}{1 - \gamma^*} \right) \alpha_*^2 + \left(q_H - 1 - \frac{\gamma^*}{1 - \gamma^*} - \frac{hq_H}{1 - \gamma^*} \right) \alpha_* + \frac{h}{1 - \gamma^*} \end{aligned} \quad (3.19)$$

which is a quadratic expression in α_* and so has the two solutions

$$\alpha_* = \frac{- \left[q_H - 1 - \frac{\gamma^*}{1 - \gamma^*} - \frac{hq_H}{1 - \gamma^*} \right] \pm \sqrt{\left(q_H - 1 - \frac{\gamma^*}{1 - \gamma^*} - \frac{hq_H}{1 - \gamma^*} \right)^2 - 4q_H \left(\frac{\gamma^*}{1 - \gamma^*} \right) \left(\frac{h}{1 - \gamma^*} \right)}}{2q_H \left(\frac{\gamma^*}{1 - \gamma^*} \right)} \quad (3.20)$$

The expression in (3.20) is not amenable to further simplification, although it is clear that the negative root is the only relevant solution from the point of view of yielding admissible values of α_* between 0 and 1. Therefore, analysis of this solution is carried out below using extensive tabulation across different admissible parameter configurations instead of using a closed-form expression: see Appendix B for further details. Note that, for the special case with constant unemployment composition, (3.16) reduces to

$$f(\alpha_{t-1} \mid \gamma^* = 0) = h - \alpha_{t-1} + \alpha_{t-1} [(1 - h) q_H]$$

which is linear and decreasing in α_{t-1} and so (3.20) reduces to

$$\alpha_* = \frac{h}{1 - (1 - h)q_H}$$

which is never reached in practice since, from Corollary 2 in Section 3.3.1, $\alpha_t \rightarrow \frac{h}{1 - (1 - h)q_H}$ in the case where there is constant unemployment composition.

As shown in Example 3.1 below, if the three exogenous parameters in the model take values $\gamma^* = 0.20$, $h = 0.90$, $q_H = 0.70$ then this gives rise to the following dynamics.

t	α_t	T_t	β_t	$f(\alpha_{t-1})$	Small firm in t ?
1	0.9000	0.3700	0.9000		Yes
2	0.9630	0.3259	0.8843	0.0630	Yes
3	0.9623	0.3264	0.8844	-0.0007	Yes

Example 3.1 (using hire-fire policy, $t = 1$ to 3)

The firm's hiring needs account for 20% of the entire labour force in each period. In Period 1, the firm hires 90% high-ability workers and 10% low-ability workers. All 10% low-ability workers and $0.30 \times 90\% = 27\%$ high-ability workers produce nothing, so the firm turns over 37% of its workers at the end of Period 1.

In Period 2, its labour force composition improves so that it has 96.30% high-ability workers and 3.70% low-ability workers. As a result of this, the unemployment pool in Period 2 has worsened in composition of high-ability workers from 90% to 88.43%. Therefore, if the firm discharges the 32.59% ($= 3.70\% + 0.30 \times 96.30\%$) of its workers which produced nothing in Period 2, then it will worsen its Period 3 labour force composition. Instead the firm may choose to stop hiring from the unemployed pool and simply retain all of its Period 2 workers in Period 3.

If the firm stops hiring in Period 2, this implies that 70.74% ($= 90\% - 0.20 \times 96.30\%$) of the overall labour force are both high-ability and persistently unemployed, in the sense that the firm has stopped hiring from among the unemployed. Note that the results shown in the table assume the firm uses a hire-fire policy in each period, but if the firm stops hiring in Period 2 then the dynamics appear as follows.

t	α_t	T_t	β_t	$f(\alpha_{t-1})$	Small firm in t ?
1	0.9000	0.3700	0.9000		Yes
2	0.9630	0.0000	0.8843	0.0630	Yes
3	0.9630	0.0000	0.8843	0.0000	Yes

Example 3.1 (where hiring stops in Period $t = 2$)

The question remains: can a firm such as described in Example 3.1 do better by

continued use of a hire-fire policy beyond Period 1, or is it optimal for the firm to stop hiring in Period 2? This depends upon whether such a firm can recoup any productivity losses incurred by using a hire-fire policy in Period 2. An extended presentation of Example 3.1 is shown below out to Period 10, where we focus upon the behaviour in α_t and omit some of the remaining columns.

t	α_t	$f(\alpha_{t-1})$	Small firm in t ?
1	0.9000		Yes
2	0.9630	0.0630	Yes
3	0.9623	-0.0007	Yes
4	0.9623	0.0000	Yes
5	0.9623	0.0000	Yes
6	0.9623	0.0000	Yes
7	0.9623	0.0000	Yes
8	0.9623	0.0000	Yes
9	0.9623	0.0000	Yes
10	0.9623	0.0000	Yes

Example 3.1 (using hire-fire policy, $t = 1$ to 10)

For the firm in this example, it is clear that continuation of a hire-fire policy out to Period 10 is sub-optimal. Although this improves the firm's composition relative to that in Period 3, nonetheless the proportion of high-ability workers α_t converges immediately to 4 decimal places to 96.23% which is lower than the 96.30% which would have been achieved had the firm ceased hiring in Period 2. The firm never attains a value of α_t as high as 96.30% in *any* period beyond Period 2 by using a hire-fire policy. Therefore, the optimal policy for this firm is to cease hiring in Period 2. As explained above, this policy leads to persistent unemployment for 70.74% of the overall labour force who are high-ability.

The tabulations used for Example 3.1 were applied more widely to a large number of admissible parameter configurations (see Appendix B). This is done by using a table of values for different combinations of values of the three parameters γ^* , h and q . The table uses a step size of 0.05 for each parameter: γ^* goes from 0.05 to 0.95 inclusive (19 distinct values), h goes from 0.05 to 1.00 inclusive (20 distinct values) and q_H goes from 0.50 to 1.00 (11 distinct values). Therefore, the table has a total number of $19 \times 20 \times 11 = 4,180$ unique configurations. The table is described more extensively in Appendix B and, due to its large size, is provided as additional information accompanying this work on a portable medium or otherwise available upon request from the author.

Using this table, the following Propositions are then able to be demonstrated by exhaustive tabulation method.

Proposition 3.

A firm which is small in Period 1 always remains small in all subsequent periods.

Proof.

See Appendix B.

■

Example 3.1 provides an illustration of a firm to which Proposition 3 applies, since this firm can use a hire-fire policy in Period 1 (and thus is not constrained in its hiring decisions during Period 1, in the sense in which we use the word "constrained" within this chapter). We can see by inspection that it can, if it chooses, use a repeated hire-fire policy without constraint. Proposition 3 asserts that this characteristic is ubiquitous, in that it applies to all other firms who are not constrained in Period 1.

Proposition 4.

A firm which is small in Period 1 always improves its labour force composition in Period 2 by using a hire-fire policy in Period 1.

Proof.

Proof follows immediately from Corollary 1 in Section 3.3.1.

■

Proposition 5.

If a firm which is small in Period 1 improve its labour force composition in Period 3 by using a hire-fire policy in Period 2, then it always improves its labour force composition by using a hire-fire policy in all subsequent periods.

Proof.

See Appendix B.

■

See Example 3.2 below for an illustration of this. Whereas the firm in Example 3.1 did better to retain its Period 2 workforce into Period 3, by contrast a firm with $\gamma^* = 0.20, h = 0.70, q_H = 0.70$ has the following profile.

t	α_t	$f(\alpha_{t-1})$	Small firm in t ?
1	0.7000		Yes
2	0.8470	0.1470	Yes
3	0.8629	0.0159	Yes
4	0.8651	0.0022	Yes
5	0.8654	0.0003	Yes
6	0.8654	0.0000	Yes
7	0.8654	0.0000	Yes
8	0.8654	0.0000	Yes
9	0.8654	0.0000	Yes
10	0.8654	0.0000	Yes

Example 3.2 (using hire-fire policy, $t = 1$ to 10)

Proposition 5 shows that, in this stylised environment, there are two types of small firm: those such as the firm in Example 3.1 which optimally stops hiring after Period 2 and those such as the firm in Example 3.2 that continue to turnover and converge from below to an equilibrium. Crucially, Proposition 5 shows that there are *no* firms which can both improve their labour force composition by continuing to hire beyond Period 2 while also improving their composition by later cessation of hiring.

Proposition 6.

If a firm which is small in Period 1 worsens its labour force composition in Period 3 by using a hire-fire policy in Period 2, then it is always optimal for such a firm to retain all of its Period 2 labour force into Period 3 and into all subsequent periods.

Proof.

See Appendix B.

■

From these last four Propositions, it can be seen that every firm which is small in Period 1 (i.e. for all admissible values of γ^*, h and q_H) can be placed into one of two mutually exclusive categories:

1. Small in all periods, optimally hires and fires in all periods (by Propositions 3, 4 and 5). The firm in Example 3.2 falls into this category.
2. Small in all periods, optimally hires and fires in Period 1 then ceases hiring and retains its Period 2 labour force (by Propositions 3, 4 and 6). The firm in Example 3.1 falls into this second category.

It has been established in this section that a form of unemployment persistence occurs wherever the market environment (i.e. the values of γ^*, h and q_H) is such that the firm falls into the second category, and that this persistence does not occur wherever the

firm falls into the first category. This simplifies the analysis of persistence considerably since it means that, for any firm which is small in Period 1, it is only necessary to look at Period 2 to determine whether persistence occurs or not.

From these results and continuity of the dynamic system, the Weierstrass Extreme Value Theorem¹² tells us that, for persistence to occur, it is sufficient for there to be at least one configuration of parameter values which satisfies the condition $f(\alpha_{t-1}) < 0$ for $(t-1) = 2$. This brings us to the critical proposition which describes whether persistence occurs for a small firm.

Proposition 7.

For a firm which is small in Period 1, persistence occurs if and only if $\gamma^ > \hat{\gamma}^*$, where*

$$\hat{\gamma}^* = \frac{q_H(1-h)}{-2hq_H - hq_H^2 + h^2q_H^2 + 1 + q_H}$$

Proof.

We know from the discussion immediately above that persistence occurs if and only if $f(\alpha_{t-1}) < 0$ for $(t-1) = 2$. Substitution of $(t-1) = 2$ into (3.19) yields the following expression:

$$f(\alpha_2) = q_H \left(\frac{\gamma^*}{1-\gamma^*} \right) \alpha_2^2 + \left(q_H - 1 - \frac{\gamma^*}{1-\gamma^*} - \frac{hq_H}{1-\gamma^*} \right) \alpha_2 + \frac{h}{1-\gamma^*}$$

where

$$\alpha_2 = h[1 + q_H(1-h)]$$

The zero-point for this function occurs at:

$$\begin{aligned} \hat{\gamma}^* &\equiv \gamma^* : f(\alpha_{t-1} = 2) = 0 \\ &= \frac{q_H(1-h)}{-2hq_H - hq_H^2 + h^2q_H^2 + 1 + q_H} \end{aligned}$$

so this establishes that the threshold for persistence to occur is at $\gamma^ = \hat{\gamma}^*$*

¹²See, for example, Rudin (1976).

Differentiation then verifies that

$$\begin{aligned}\frac{\partial f(\alpha_{t-1}=2)}{\partial \gamma^*} &= q_H \alpha_2^2 \left[\frac{1}{(1-\gamma^*)^2} \right] + \alpha_2 (1 + h q_H) \left[\frac{-1}{(1-\gamma^*)^2} \right] + h \left[\frac{1}{(1-\gamma^*)^2} \right] \\ &= \left[\frac{1}{(1-\gamma^*)^2} \right] [q_H \alpha_2^2 - (1 + h q_H) \alpha_2 + h]\end{aligned}$$

which is negative $\forall h, q_H \in (0, 1)$. Therefore, it is for values of γ^* higher than $\hat{\gamma}^*$ that persistence occurs.

■

Proposition 7 reveals that persistence occurs only for "high" values¹³ of γ^* . This is because the larger is the firm relative to the unemployed pool, the more drastically the composition of the unemployed is affected by the firm's release of mostly low quality workers into unemployment after Period 1. Therefore, a larger firm faces a "worse" pool of unemployed after Period 2 than a smaller firm would, and for a sufficiently large firm¹⁴ this leads to persistence in employment composition.

Proposition 7 is not as sweeping as Proposition 2. In particular, it does not put forward a general prediction as to whether persistence will occur; instead, persistence is conditional upon parameter values. Although Proposition 7 is weaker in this way, it nonetheless provides a clear result. For any admissible configuration of the parameter values h, q_H and γ^* , a single condition can be evaluated to determine whether regular turnover is preferred over no turnover for a small firm. Persistence, if it occurs, takes place following the end of Period 2 when the firm makes its second turnover decision.

Table 3.1 illustrates this for a number of different parameter configurations at the end of Period 2. Recall that $\hat{\gamma}^*$ is the minimum size of the firm γ^* necessary for persistence to take place, and $\tilde{\gamma}^*$ is the maximum size of γ^* admissible for the firm to still be small. Therefore, $\hat{\gamma}^*$ and $\tilde{\gamma}^*$ form a range within which we persistence occur for a small firm. For example, in the first line, if γ^* is less than one third, a repeated hire-fire policy is optimal; if γ^* is between one third and one half, persistence occurs; and, if γ^* is greater than one half, the firm is large.

¹³Mathematically, this is clear even without differentiating γ^* . We already know that persistence is never satisfied for the *lowest* possible value of γ^* (i.e. $\gamma^* = 0$), since this is equivalent to the special case in Section 3.3.1 with constant composition. Therefore, as there is only one "persistence threshold" solution for γ^* , we know that it must be *high* values of γ^* which lead to persistence.

¹⁴Although not so large as to violate the small firm condition.

Table 3.1: Persistence thresholds (simple model)

Parameters		Simple Model	
h	q	$\hat{\gamma}^*$	$\tilde{\gamma}^*$
0.0	0.5	0.33	0.50
0.1	0.5	0.33	0.52
0.5	0.5	0.27	0.59
0.7	0.5	0.20	0.63
0.9	0.5	0.09	0.65
1.0	0.5	0.00	0.67
0.0	0.7	0.41	0.50
0.1	0.7	0.42	0.53
0.5	0.7	0.40	0.65
0.7	0.7	0.34	0.71
0.9	0.7	0.18	0.75
1.0	0.7	0.00	0.77
0.0	0.9	0.47	0.50
0.1	0.9	0.49	0.54
0.5	0.9	0.56	0.74
0.7	0.9	0.57	0.83
0.9	0.9	0.43	0.90
1.0	0.9	0.00	0.91
0.0	1.0	0.50	0.50
0.1	1.0	0.53	0.55
0.5	1.0	0.67	0.80
0.7	1.0	0.77	0.92
0.9	1.0	0.91	0.99
1.0	1.0	1.00	1.00

It is clear from inspection that the small firm threshold γ^* is monotonically increasing in each of the parameters h and q . In other words, if the firm has more able workers or better inspection technology then it optimally chooses a relatively small turnover, even under a hire-fire policy; therefore, the constraints imposed by the size of the unemployed pool do not bind so heavily. For example, if $h = 0.7$ and q increases from 0.7 to 0.9, then γ^* increases from 0.71 to 0.83, so a firm will be “small” over a greater range of its size parameter γ^* . However, the relationships between the lower bound for persistence $\hat{\gamma}^*$ and the parameters h and q are not so easily characterised. For all plausible values of h , $\hat{\gamma}^*$ is monotonically increasing in q . For low values of q , $\hat{\gamma}^*$ is monotonically *decreasing* in h , but this relationship in h is non-monotonic for higher values of q . For example, if $q = 0.5$ then $\hat{\gamma}^*$ falls as h increases; but, if $q = 0.9$ then $\hat{\gamma}^*$ moves up, and then back down, over an increasing range of h .

To understand these comparative statics, we put forward three issues, or effects, which the firm has to consider at the end of Period 2: namely, *how accurately* can the firm sort its Period 2 workers (i.e. a quality effect on the employed); *how much worse* was the average quality of those workers discharged after Period 1 than the initial quality of the unemployed (i.e. a quality effect on the unemployed); and, *how many* workers were discharged after Period 1 (i.e. a quantity effect on the unemployed). There are two quality effects and one quantity effect. The quality effect on the employed is simple, and is completely described by the parameter q . The higher is q , the greater is the quality effect on the employed and the smaller is the “ideal persistence region”¹⁵, since the firm can accurately sort Period 2 workers. The quality effect on the unemployed (also described by q) works in the other direction. The more accurately Period 1 workers are sorted, the worse is the composition of those workers sent into unemployment during Period 2, and the larger is the ideal persistence region (since the unemployed are less attractive). The quantity effect on the unemployed depends on the size of the firm relative to the unemployed pool *and* on the size of the firm’s Period 1 turnover. The larger is the flow of discharged workers into unemployment after Period 1, the more drastically average worker quality in the unemployed pool will be worsened (since by construction, workers discharged after Period 1 are, on average, worse than the Period 1 unemployed). Therefore, the higher the quantity effect on the unemployed, the larger is the ideal persistence region.

¹⁵The “ideal persistence region” refers to the region between $\hat{\gamma}^*$ and 1, rather than the *actual* persistence region between $\hat{\gamma}^*$ and $\tilde{\gamma}^*$. We concentrate on the former here since that is endogenous to the firm’s ideal decisions (which we wish to discuss) rather than endogenous to its labour force size constraints.

The firm has to weigh up two quality effects, working in different directions, and a quantity effect. However, an immediate simplification arises from noting that the quality effect on the employed always dominates that for the unemployed; this is equivalent to saying that the net effect of being able to sort more accurately is to make the ideal persistence region smaller.¹⁶ The effects therefore reduce to those which offset quality against quantity. If the quality effect is larger (smaller), the ideal persistence region is correspondingly smaller (larger).

Interpretation of the relationships, variously, between $\hat{\gamma}^*$ and q and between $\hat{\gamma}^*$ and h , becomes much clearer in light of the preceding paragraph's discussion. The first relationship may be interpreted as meaning that better inspection technology q gives the firm more to gain from turnover at the end of Period 2, thus making the ideal persistence region smaller by increasing $\hat{\gamma}^*$. This relationship is unambiguous, as it *only* involves quality effects. The second relationship is more complicated. More high ability workers in the firm's Period 1 workforce h will have two effects on the size of the ideal persistence region. The cohort of workers discharged into unemployment at the end of Period 1 has a better composition; this is a *smaller quality effect*, which will cause the persistence region to be larger. But then the size of the cohort of workers discharged is also smaller; this is a *smaller quantity effect*, which will cause the persistence region to be smaller. This is why the effect of changing h is ambiguous for $\hat{\gamma}^*$, since it has offsetting effects on quality and quantity.¹⁷

¹⁶Formally, this is achieved by taking $\frac{\partial \hat{\gamma}^*}{\partial q_H} = \frac{(-1+h)(-q_H^2 h + q_H^2 h^2 - 1)}{(-2q_H h - q_H^2 h + q_H^2 h^2 + 1 + q_H)^2} > 0$ since both expressions in parentheses in the numerator must be negative for all $1 \geq q_H \geq 0.5, 1 \geq h \geq 0$.

¹⁷In line with this, $\frac{\partial \hat{\gamma}^*}{\partial h} = q_H \frac{-2q_H^2 h + q_H^2 h^2 - 1 + q_H + q_H^2}{(-2q_H h - q_H^2 h + q_H^2 h^2 + 1 + q_H)^2}$, the sign of which is ambiguous. In general, however, quality effects tend to be more important when h is large or q is small, as is evident from Table 3.1.

Large Firm

We turn next to the case where Condition 1 does *not* hold, so that

$$\gamma^* > \tilde{\gamma}^* \equiv \frac{1}{1 + T_t}$$

in some period t . Under such a scenario, the firm is constrained by the small size of the unemployed pool, so is not able to implement turnover T_t . Instead it must make do with turnover Γ_t equal to the entire size of the unemployed pool:

$$\Gamma_t \equiv \frac{1 - \gamma^*}{\gamma^*}, \forall t \quad (3.21)$$

Note that, for a large firm, (3.14) still holds, so we may still write:

$$h = \alpha_t \gamma^* + \beta_t (1 - \gamma^*)$$

However, since the firm is now constrained in the extent to which it can turnover workers, the corresponding expression for the small firm (3.15) is now modified to become:

$$\begin{aligned} \alpha_t &= \frac{\gamma^* q_H \alpha_{t-1} + (1 - \gamma^*) \beta_{t-1} + \alpha_{t-1} \left(\frac{1 - q_H}{1 - q_H \alpha_{t-1}} \right) [\gamma^* (1 - q_H) \alpha_{t-1} - (1 - \gamma^*)]}{\gamma^*} \quad (3.22) \\ &= \alpha_{t-1} + \left(\frac{1 - \gamma^*}{\gamma^*} \right) \left[\beta_{t-1} - \frac{(1 - q_H) \alpha_{t-1}}{1 - q_H \alpha_{t-1}} \right] \end{aligned}$$

Again substituting for β_{t-1} from (3.15) and proceeding to write the change in employment composition as a function of prior composition gives:

$$\alpha_t - \alpha_{t-1} = g(\alpha_{t-1})$$

where

$$g(\alpha_{t-1}) = -\alpha_{t-1} + \frac{h}{\gamma^*} - \left(\frac{1 - \gamma^*}{\gamma^*} \right) (1 - q_H) \frac{\alpha_{t-1}}{1 - q_H \alpha_{t-1}} \quad (3.23)$$

Similar to our analysis for the small firm, we wish to evaluate the sign of $g(\alpha_{t-1})$ for each period, since if $g(\alpha_{t-1})$ is always non-negative then the firm will optimally use a repeated hire-fire policy. We illustrate this using Example 3.3 below, which uses parameter values $\gamma^* = 0.70, h = 0.60, q_H = 0.70$.

t	α_t	$f(\alpha_{t-1})$	T_t	Turnover (#)	Unemployed (#)	Constrained?
1	0.6000		58.00%	0.4060	0.3000	Yes
2	0.7241	0.1241	49.31%	0.3452	0.3000	Yes
3	0.6683	-0.0558	53.22%	0.3725	0.3000	Yes
4	0.6957	0.0273	51.30%	0.3591	0.3000	Yes
5	0.6828	-0.0129	52.20%	0.3654	0.3000	Yes
6	0.6890	0.0062	51.77%	0.3624	0.3000	Yes
7	0.6860	-0.0029	51.98%	0.3638	0.3000	Yes
8	0.6874	0.0014	51.88%	0.3632	0.3000	Yes
9	0.6868	-0.0007	51.93%	0.3635	0.3000	Yes
10	0.6871	0.0003				

Example 3.3 (using hire-fire policy, $t = 1$ to 10)

In Example 3.3, the firm's hiring needs account for 70% of the entire labour force in each period, so there is always 30% of workers (i.e. measure 0.3000) in the unemployed pool. We show both the firm's desired turnover T_t assuming the firm uses a repeated hire-fire policy and also the equivalent turnover measure measured in # of workers within the entire labour force (which has a total # of 1.000). The firm's turnover is constrained whenever this # of workers exceeds the maximum possible turnover of 0.3000 from the unemployed pool.¹⁸

In Period 1, the firm hires 60% high-ability workers and 40% low-ability workers. All 40% low-ability workers and $0.30 \times 60\% = 18\%$ high-ability workers produce nothing, so the firm desires to turn over 58% of its workers, or equivalently a total of $58\% \times 70\% = 40.6\%$ of the entire labour force, at the end of Period 1. However this is not possible since there are only 30% of the entire labour force in unemployment. Therefore, the firm's constrained turnover is 30% of the entire labour force, or equivalently 42.86% of the firm's labour force.

Therefore, the firm's Period 2 labour force composition improves so that it has 72.41% [$= 0.70 \times 60\% + 0.30 \times 60\% \times 15.14\% / 58\% + 42.86\% \times 60\%$] high-ability workers and 27.59% low-ability workers. As a result of this, the unemployment pool in Period 2 has worsened in composition of high-ability workers so that, in turn, the firm's Period 3

¹⁸Note that a firm which is large in Period 1 is, according to Definition 3, constrained in Period 1 in the sense of not being able to implement a hire-fire policy and, also according to Definition 3, large in all subsequent periods.

labour force composition will worsen to only 66.83% of high-ability workers if it uses its full constrained turnover at the end of Period 2. By contrast, if the firm stops hiring in Period 2, this implies that 9.3103% [= 30% x (18% / 58%)] of the overall labour force are both high-ability and persistently unemployed, in the sense that the firm has stopped hiring from among the unemployed.

One of the characteristics of Example 3.3 is that the firm's turnover is constrained in every period. However, this need not be the case. Example 3.4 below illustrates a case where the firm is constrained in some periods but not in others: we use parameter values $\gamma^* = 0.70, h = 0.60, q_H = 0.80$.

t	α_t	$f(\alpha_{t-1})$	T_t	Turnover (#)	Unemployed (#)	Constrained?
1	0.6000		52.00%	0.3640	0.3000	Yes
2	0.7582	0.1582	39.34%	0.2754	0.3000	No
3	0.6974	-0.0609	44.21%	0.3095	0.3000	Yes
4	0.7219	0.0246	42.25%	0.2957	0.3000	No
5	0.7108	-0.0111	43.13%	0.3019	0.3000	Yes
6	0.7159	0.0051	42.73%	0.2991	0.3000	No
7	0.7135	-0.0023	42.92%	0.3004	0.3000	Yes
8	0.7146	0.0011	42.83%	0.2998	0.3000	No
9	0.7141	-0.0005	42.87%	0.3001	0.3000	Yes
10	0.7144	0.0002				

Example 3.4 (using hire-fire policy, $t = 1$ to 10)

In Example 3.4, the firm is constrained in Period 1 but not in Period 2, since its labour force composition has improved in Period 2 and so its desired turnover using a hire-fire policy is sufficiently low that it can be met from the pool of unemployed. This pattern continues to oscillate during future periods in line with the firm's changing composition under the assumption that it uses a hire-fire policy.

Here, again, the pertinent question is: can large firms such as these do better by continued use of a hire-fire policy beyond Period 1, or is it optimal to stop hiring in Period 2? The answer is clear for the firms illustrated by both Examples 3.3 and 3.4, because the proportion of high-ability workers attained in Period 2 is never attained in any subsequent period using a hire-fire policy. In both examples, fairly rapid convergence in α_t takes place using a repeated hire-fire policy and this convergence is to a value lower than α_2 . Therefore each firm does best to stop hiring in Period 2.

Proposition 8 below extends these results to apply to all firms which are large in Period 1. To establish this, the tabulations used for Examples 3.3 and 3.4 were applied more widely to a large number of admissible parameter configurations, similar to that which was

done for the analysis of small firms.

Proposition 8.

It is always optimal for a firm which is large in Period 1 to use a hire-fire policy in Period 1, and then to retain all of its Period 2 labour force into Period 3 and into all subsequent periods.

Proof.

See Appendix B.

■

Proposition 8 establishes a clearer result than does Proposition 7, the corresponding result for small firms. Namely, if a firm is large in Period 1, then persistence always occurs: so the firms illustrated in Examples 3.3 and 3.4 are indeed typical firms within this framework. This result is one of true state dependence, since hypothetically two identical type θ_H workers may experience very different employment patterns over time (i.e. one always employed, the other always unemployed) wholly because of different random shocks to their respective outputs in Period 1. Thus, we can see that use of our simple framework has generated a clear result: persistence **sometimes** occurs in a “small firm” environment, and **always** occurs in a “large firm” environment.

3.3.3 Summary of Results

The simple model described in Section 3.3 allows us to draw several conclusions.

1. Comparison between different versions of the model demonstrate that changes to the composition of the unemployed may generate persistence. The key mechanism is that, as low-ability workers are discharged into unemployment, hiring from among the unemployed may be less productive to the firm if it cannot sort workers by intrinsic ability.
2. The existence of changes in unemployment composition is not sufficient to generate persistence. The analysis of small firms in Section 3.3.2 shows that the more accurately firms can sort their workers, the more productive is hiring from among the unemployed.
3. Persistence effects are strongest when aggregate unemployment levels are low. If this is the case (due perhaps to a large, nonsubstitutable capital stock), then the analysis of large firms in Section 3.3.2 shows that discharging a given number of low quality workers into a small unemployed pool will have a larger effect on the average quality of the unemployed. In such a case, hiring from among the unemployed may be less productive to the firm.

3.4 Extensions to the Simple Model

Using the simple model described in Section 3.3 above is advantageous to understanding how persistence may be generated in the presence of informational constraints. However, the model has several limitations. First, there is uncertainty about the intrinsic productivity of only one type of worker, the high type. This is quite restrictive. Second, the extent of persistence in the simple model is complete; that is, once persistence sets in, no worker among the unemployed pool will ever again find employment, even if they are of high intrinsic productivity type. Since complete persistence is not often observed empirically, this limits the model's predictive power. Third, although Bayesian updating occurs in the model, it is not pursued beyond the current period. All observations made by the firm about a worker's output realisations before the current period are disregarded, and each worker's characteristics are only assessed by aggregate characteristics and their current individual output realisation. It is likely an unnecessary and suboptimal strategy for the firm to restrict its information flow in this way. Fourth, workers are treated as passive agents unable to signal their individual intrinsic productivity types in any way. Thus, if persistence sets in, a high type has to accept being unemployed for the foreseeable future, rather than - say - acquiring costly education *a la* Spence (1973). Again, this may be an unrealistic restriction, both from the firm's and the worker's viewpoint, since gains from employment are wasted by the underlying uncertainty; thus, incentives exist to attempt, endogenously, to overcome the uncertainty.

In this section, the first limitation is addressed, leaving the last three for future research. In particular, we examine the implications of introducing uncertainty over both productivity types.

3.4.1 Uncertainty about Productivity of Both Worker Types

We return to the basic model with a single firm, where

γ_t is the proportion of the labour force employed in period t

p_{nt} is the realised productivity of worker n employed in period t

$\bar{p}_t$ is the expected average productivity of workers employed in period t

h is the proportion of high productivity workers in the labour force

q_j is successful inspection rate of individual productivity levels within currently employed workers of type j

We continue to assume a fixed proportion of the labour force in employment over time, so that $\gamma_t = \gamma^*$ for all t . However, we now allow $q_L < 1$, so that inspection of low intrinsic productivity types is noisy as well. For simplicity, we consider the case where $q_H = q_L = q$. Thus, for all workers of type $\theta = \theta_H$,

$$E[p_{nt} \mid \theta = \theta_H] = q \cdot 1 + (1 - q) \cdot 0 = q \quad (3.24)$$

Similarly, for workers of type $\theta = \theta_L$,

$$E[p_{in} \mid \theta = \theta_L] = q \cdot 0 + (1 - q) \cdot 1 = 1 - q \quad (3.25)$$

These may be compared with the analogous expressions in (3.1) and (3.2). Again, it is natural to restrict q to be greater than $\frac{1}{2}$, and all other assumptions are identical to those used in the simple model. Then, at the end of period 1, the firm has measure γ^* workers employed, with output per worker of

$$\bar{p}_1 = hq + (1 - h)(1 - q) \quad (3.26)$$

The firm is able to partition the γ^* workers into 2 distinct groups according to its observations of output during period 1. One group (A) has high *realisations* of output attributable to its members; the other (B) has low realisations. However, because of the underlying noisy signals of output, neither group is homogeneous in terms of intrinsic productivity

types. Indeed, probabilities of type, conditioned on output, are:

$$\Pr[\theta = \theta_H \mid p_{nt} = 1] = \frac{hq}{hq + (1-h)(1-q)} \quad (3.27)$$

$$\begin{aligned} \Pr[\theta = \theta_L \mid p_{nt} = 1] &= 1 - \frac{hq}{hq + (1-h)(1-q)} \\ &= \frac{(1-h)(1-q)}{hq + (1-h)(1-q)} \end{aligned} \quad (3.28)$$

$$\Pr[\theta = \theta_H \mid p_{nt} = 0] = \frac{h(1-q)}{h(1-q) + (1-h)q} \quad (3.29)$$

$$\begin{aligned} \Pr[\theta = \theta_L \mid p_{nt} = 0] &= 1 - \frac{h(1-q)}{h(1-q) + (1-h)q} \\ &= \frac{(1-h)q}{h(1-q) + (1-h)q} \end{aligned} \quad (3.30)$$

Small Firm

Assume the firm is small in the sense of Condition 2, and consider the implications of a hire-fire policy in this setting. Using the same definitions of α_t, β_t and T_t as for the simple model, we have the following results:

$$\begin{aligned} \alpha_1 &= h \\ T_1 &= (1-h)q + h(1-q) \\ &= q + h - 2hq \\ \alpha_2 &= q\alpha_1 + \beta_1 T_1 \\ &= qh + \beta_1(q + h - 2hq) \\ T_2 &= (1-\alpha_2)q + \alpha_2(1-q) \\ &= q + [qh + \beta_1(q + h - 2hq)] - 2q[qh + \beta_1(q + h - 2hq)] \end{aligned}$$

Under an assumption of constant composition among the unemployed (as was used in the first subcase of the simple model), $\beta_t = h$ for all t . Under this assumption it is straightforward to show that qualitatively similar results to those in the constant unemployment composition subcase of the simple model will be obtained. Namely, there will never be persistence in employment outcomes, since the firm can always improve its composition by

using a repeated hire-fire policy. We therefore omit further discussion of this possibility, as it is more interesting (and realistic) to consider the case where the composition of the unemployed pool depends on the firm's hiring decisions.¹⁹ Thus, we obtain

$$\begin{aligned}\beta_1 &= h \\ \beta_2 &= \beta_1 \left[\frac{(1 - \gamma^*) - \gamma^* T_{t-1}}{1 - \gamma^*} \right] + \frac{\alpha_1 (1 - q) \gamma^*}{1 - \gamma^*}\end{aligned}$$

which is similar to the analogous expression for the simple model. Thus, the only effect imposed by setting $q_H = q_L = q$ arises in the turnover function T_t , where turnover under a hire-fire policy is now smaller due to not being able, information-wise, to discharge all workers of type θ_L (although, under a hire-fire policy, it can certainly discharge all workers in Group B).

What are the firm's optimal hiring decisions in this setting? To answer this we use the same solution method as in Section 3.3, which decomposes the population of high quality workers into two groups: those who are employed by the firm and those who are unemployed. Equation (3.14) still holds so we can write

$$h = \alpha_t \gamma^* + \beta_t (1 - \gamma^*) \quad (3.31)$$

However, the expression for the firm's desired turnover using a repeated hire-fire policy is now

$$T_t = q + (1 - 2q) \alpha_t \quad (3.32)$$

and we also know that the result of using a hire-fire policy, in terms of effect upon α_t , is

$$\alpha_t = q \alpha_{t-1} + \beta_{t-1} T_{t-1} \quad (3.33)$$

Putting (3.31), (3.32) and (3.33) together gives the key difference equation:

$$\alpha_t - \alpha_{t-1} = h(\alpha_{t-1})$$

¹⁹The relevant comparison is with the model in Section 3.3.2, above.

where

$$h(\alpha_{t-1}) = \frac{\gamma^*}{1-\gamma^*} (2q-1) \alpha_{t-1}^2 + \left[q-1 + \frac{1}{1-\gamma^*} (h-2hq-q\gamma^*) \right] \alpha_{t-1} + \frac{hq}{1-\gamma^*} \quad (3.34)$$

We use the extensive tabulation method described in Section 3.3.2 to characterise the firm's optimal behaviour. Indeed, it turns out that Propositions 3, 4, 5 and 6 from Section 3.3.2 are also valid within this environment (see Appendix B). We can also demonstrate Proposition 9, which is the analogue of Proposition 7, in this environment.

Proposition 9.

For a firm which is small in Period 1, persistence occurs if and only if $\gamma^ > \hat{\gamma}^*$, where*

$$\hat{\gamma}^* = \frac{h-2hq+q}{h^2-4qh^2+4h^2q^2-4hq^2+h+2q}$$

Proof.

Proof proceeds in identical fashion as for Proposition 7. We know from the discussion immediately above that persistence occurs if and only if $h(\alpha_{t-1}) < 0$ for $(t-1) = 2$. Substitution of $(t-1) = 2$ into (3.34) yields the following expression:

$$h(\alpha_2) = \frac{\gamma^*}{1-\gamma^*} (2q-1) \alpha_2^2 + \left[q-1 + \frac{1}{1-\gamma^*} (h-2hq-q\gamma^*) \right] \alpha_2 + \frac{hq}{1-\gamma^*}$$

where

$$\alpha_2 = hq + \beta_1 (q + h - 2hq)$$

The zero-point for this function occurs at:

$$\begin{aligned} \hat{\gamma}^* &\equiv \gamma^* : f(\alpha_{t-1} = 2) = 0 \\ &= \frac{h-2hq+q}{h^2-4qh^2+4h^2q^2-4hq^2+h+2q} \end{aligned}$$

so this establishes that the threshold for persistence to occur is at $\gamma^ = \hat{\gamma}^*$*

Differentiation verifies the direction of the inequality.

■

Table 3.2: Persistence thresholds (small firm) for Simple Model and Extended Model

h	q	Simple Model		Extended Model 1	
		$\hat{\gamma}^*$	$\tilde{\gamma}^*$	$\hat{\gamma}^*$	$\tilde{\gamma}^*$
0.0	0.5	0.33	0.50	0.37	0.50
0.1	0.5	0.33	0.52	0.37	0.52
0.5	0.5	0.27	0.59	0.32	0.59
0.7	0.5	0.20	0.63	0.27	0.62
0.9	0.5	0.09	0.65	0.17	0.64
1.0	0.5	0.00	0.67	0.10	0.66
0.0	0.7	0.41	0.50	0.45	0.50
0.1	0.7	0.42	0.53	0.46	0.53
0.5	0.7	0.40	0.65	0.45	0.64
0.7	0.7	0.34	0.71	0.41	0.69
0.9	0.7	0.18	0.75	0.29	0.73
1.0	0.7	0.00	0.77	0.16	0.75
0.0	0.9	0.47	0.50	0.51	0.50
0.1	0.9	0.49	0.54	0.53	0.54
0.5	0.9	0.56	0.74	0.62	0.72
0.7	0.9	0.57	0.83	0.64	0.80
0.9	0.9	0.43	0.90	0.57	0.86
1.0	0.9	0.00	0.91	0.36	0.87
0.0	1.0	0.50	0.50	0.54	0.50
0.1	1.0	0.53	0.55	0.57	0.55
0.5	1.0	0.67	0.80	0.71	0.77
0.7	1.0	0.77	0.92	0.82	0.87
0.9	1.0	0.91	0.99	0.96	0.94
1.0	1.0	1.00	1.00	1.00	0.95

Table 3.2 compares the results obtained here with those from the simple model for a small firm. Both the firm's lower bound for persistence $\hat{\gamma}^*$ and the small firm constraint $\tilde{\gamma}^*$ are different; in the case of $\tilde{\gamma}^*$, this is due to the fact that the turnover function T_t takes smaller values in this scenario (although Condition 1 still applies here). In any case, as shown in Table 3.2, the actual differences in $\tilde{\gamma}^*$ between the two small firm models are only slight. For example, in the first line of Table 2, $\tilde{\gamma}^* = 0.50$ to 2 decimal places for both the Simple model and the Extended model. However, the differences in $\hat{\gamma}^*$ tend to be much greater; the Extended model, which allows for uncertainty about both types, consistently calculates *higher* values for $\hat{\gamma}^*$ than does the Simple model. Thus, persistence takes place over a *smaller* region of parameter configurations when we allow for uncertainty about both types. This is partly due to the smaller quantity effects in the model. Since some of the type θ_L workers can “masquerade” as type θ_H , the firm will discharge less workers (i.e. lower T_t) every period under a hire-fire policy. Therefore, the effect of such a policy, which worsens the composition of the unemployed, is less drastic than if all type θ_L workers had been discharged. Therefore, the firm has more incentive to continually rehire, which is demonstrated by the results in Table 3.2.²⁰

Large Firm

Condition 1 from Section 3.3 still describes whether a firm is small in any period, so a large firm in Period 1 satisfies:

$$\gamma^* > \tilde{\gamma}^* \equiv \frac{1}{1 + T_t}$$

In this environment, (3.32) is the firm's desired turnover, with the firm constrained to a maximum turnover of $\frac{1-\gamma^*}{\gamma^*}$ if large in the sense of Condition 1. Therefore, we can write

$$\alpha_t = q\alpha_{t-1} + \beta_{t-1} \left(\frac{1 - \gamma^*}{\gamma^*} \right) \quad (3.35)$$

²⁰Note, however, that quality effects are also present, and are less clear than in the simple model. In particular, the quality effect on the unemployed pool no longer always dominates the quality effect on the employed; specifically, the sign of $\frac{\partial \hat{\gamma}^*}{\partial q} = \frac{h[2h^2 - 8qh^2 + 8q^2h^2 - 12q^2h - h + 8qh + 4q^2 - 1]}{(h^2 - 4qh^2 + 4q^2h^2 - 4q^2h + h + 2q)^2}$ is now ambiguous. Therefore, it is possible that, for some parameter configurations, the quality effect could overturn the quantity effect which would lead to persistence arising over a larger region. However, the results in Table 3.2 suggest that the quantity effects are strong enough to consistently generate a *smaller* persistence region than does the simple model.

and, putting (3.31) and (3.35) together gives the difference equation for a large firm:

$$\alpha_t - \alpha_{t-1} = j(\alpha_{t-1})$$

where

$$j(\alpha_{t-1}) = (q-2)\alpha_{t-1} + \frac{h}{\gamma^*} \quad (3.36)$$

Appendix B shows that Proposition 8 also applies to the system described by (3.36), so again we conclude that persistence always occurs from Period 2 onwards for a large firm.

3.5 Conclusion

We group all of the theoretical results together for ease of comparison. In all, we have considered six variations of the same model. The key issue examined was that of unemployment persistence: will it always, sometimes or never occur under different underlying assumptions about the decision-making environment?

Table 3.3: Summary of Models' Predictions

DOES PERSISTENCE RESULT?		
	Small Firm	Large Firm
Simple Model, exogenous composition	Never	Never
Simple Model, endogenous composition	Sometimes	Always
Extended Model	Sometimes	Always

Table 3.3 summarises these findings, where small and large are in the sense of Condition 1. Recall again that these labels are not literally a description of any one firm's size (unless the relevant market is served by only 1 firm); but, instead a characterisation of the size of the employed pool relative to the labour force.

The natural baseline model is the simple model with endogenous composition and noisy inspections over only one worker type. This model gives us strong results: in particular, that persistence always occurs if the unemployed pool is sufficiently small (i.e. if $\gamma^* > \tilde{\gamma}^*$ so that the firm is large). For this model, “sufficiently small” would easily apply to most modern economies. For example, from Table 3.1 we can see that, if the overall proportion of high productivity workers h is 0.5 and inspection success q is 0.7, then the large firm condition is satisfied (and hence we predict complete persistence) for any $\gamma^* > 0.65$; that is, for any unemployment rate less than 35%.

The extended model we use suggests persistence occurs less widely than does the simple model. The extended model, which incorporates the possibility of uncertainty in inspecting both worker types, compresses the size of the “persistence region” from below. Since the firm cannot fire all of its low realisation workers, it will not be able to discharge a drastically large proportion of low quality workers into unemployment in Period 2; thus, it will have more incentives to hire from unemployment in future periods. This reduces the possibility of persistence to an extent. However, we still obtain from the extended model the result that the firm, if large, will never rehire after Period 2. Since the lower bound for a large firm is not much higher than for the simple model (e.g. for the example above, where $h = 0.5$ and $q = 0.7$, $\tilde{\gamma}^* = 0.68$ as compared with $\tilde{\gamma}^* = 0.65$), the extended model

accordingly makes some of the same strong predictions about persistence as does Simple 2.

In Section 3.2.3 above, a list of questions for describing those theoretical models which explain persistence was set out. Having explained our model, we turn again to answer these questions:

1. Are wages and/or job characteristics endogenous?

No.

2. What is the nature of job-matching technology?

Frictionless: neither search mechanisms nor hiring costs are imposed on workers and firms.

3. Are both workers and firms fully-optimising agents?

Firms optimise, while workers are all assumed ready to supply labour at the fixed wage.

4. Is there a role for competing firms to induce worker turnover?

No.

5. What is the information structure for each of workers and firms?

Imperfect information exists for firms, and information is not relevant for workers.

6. If information is imperfect, what is the nature and scope of dynamic learning?

Firms learn by Bayesian updating about each worker. However, firms do not use repeated observations on any worker to infer productivity (i.e. Bayesian updating is first-order).

It is also instructive to compare the predictions made here with those of Montgomery (1999), since that work is closer to this Chapter's than any other in the literature. Both this Chapter and Montgomery's use a basic framework of discrete-time Bayesian updating of information about workers' abilities, under fixed wages. However there are four key differences in Montgomery's work. First, the arbitrary assumption that worker types are equally represented in the labour force is made; this corresponds to imposing the restriction of $h = \frac{1}{2}$ in our model. Second, the assumption is maintained throughout that only high ability workers are observed with some error; whereas, this assumption is relaxed in our extended model. Third, Bayesian updating is not merely first-order but takes into

account a worker's entire employment history with the firm. Finally, exogenous hiring costs exist in Montgomery's work; and, indeed, are crucial to explaining the ultimate result of cycles in employment patterns. The first two differences show that, in these respects, our model is more general than Montgomery's approach, while our model is a special case of Montgomery's in relation to the last two differences. Therefore, although the two models slightly diverge with respect to the directions they extend the existing theory in, the two approaches can nonetheless be viewed as complementary.

Two conclusions emerge from this Chapter:

1. Theoretical explanations of persistence should, ideally, be robust against optimal actions being taken by those who are disadvantaged by its effects.
2. The existence of imperfect information can, under certain limited circumstances, theoretically explain unemployment persistence.

However, each of these conclusions immediately raise fresh questions. Conclusion 1) should make us pause to consider how optimal actions (especially by workers) can they be incorporated into a model such as ours. Could some sort of costly signalling be introduced, so that workers (and, in particular, high-ability workers) are not simply the passive recipients of an imperfect sorting mechanism?

Conclusion 2) raises the question of how far this model can be extended in its current form. If we acknowledge that, ultimately, an explanation for persistence which relies on a fixed-wage firm with fixed-coefficients technology is not a complete explanation, then developing a more general approach will be necessary. However, already the basic theoretical model is used is already intractable to the extent that results from both our extended models rely heavily on exhaustive tabulations. Can this be improved?

We suggest it is imperative to investigate refinements which enable the model to make less sweeping predictions about the extent of persistence. For example, in our model the unemployed workers initially have the same average characteristics as the incumbents in the labour force. However, Pryor and Schaffer (1999) among others suggest that in fact younger workers (i.e. likely entrants) typically have more skills than those they replace in the workforce. Therefore, in our context, this may be an effect which actually causes the composition of the unemployed pool to *improve*, which would offset against the endogenous worsening of the unemployed pool's composition in the models presented here.

Finally, this study's results should be of particular use to policy-makers in the UK environment. As alluded to in Chapter 1, it is not only the *overall* rate of unemployment

which is of concern, but also the *composition* of unemployment. It appears to be the case that those who become unemployed, for whatever reason, are likely to stay so. Therefore, significant action may be required to prevent unemployment becoming perpetually concentrated within relatively small groups. Such action could, plausibly, take the form of targetted wage subsidies, to prevent certain groups becoming unemployed in the first instance. Complementary actions could also be taken to attempt to provide advantage to those already unemployed, perhaps through schemes designed to aid formation of human capital. It is desirable that future initiatives analyse the likely interactions of such policies with dynamic effects such as those identified in this Chapter.

Chapter 4

Measuring Unemployment

Persistence for the UK

4.1 Introduction

This Chapter discusses the use of several panel data techniques to perform analysis on binary data. Discussion is organised around the estimation, and presentation of, results from an empirical investigation into the correlation over time (i.e. serial correlation) among individual employment outcomes for men. In the relevant literature, stability in this context is referred to - variously - as persistence, state dependence or hysteresis. These phrases all carry the connotation that employment outcomes display positive serial correlation. The Chapter's results - based on the UK British Household Panel Survey (BHPS) over the period 1991 to 2000 - are consistent with those predicted by Chapter 3.

The Chapter is organised as follows. Section 4.2 reviews the extant empirical literature on causes of persistence. Section 4.3 discusses methodological and estimation-related issues. Section 4.4 discusses the data set used. Results are presented in Section 4.5. Robustness analysis for the results is conducted in Section 4.6. Section 4.7 concludes the Chapter.

4.2 Background

For at least 30 years, there has existed a substantial literature which focusses on explaining the empirical regularity that individual employment outcomes display positive serial correlation. Two competing explanations have traditionally been put forward to explain this. The first is that unemployed workers seeking paid employment are relatively disadvantaged by the mere experience of being unemployed; this is referred to as “true state dependence”, “structural state dependence” or simply “state dependence”. The second is that workers possess unobserved characteristics¹ which are temporally persistent, heterogeneous across individuals *and* important for determining individual labour supply and/or demand; this is referred to as “unobserved heterogeneity” or, as described by James Heckman, “spurious state dependence”.²

Each explanation, taken by itself, can explain serial correlation in individual employment outcomes. However, although the potential for competing explanations has never been seriously challenged, disagreement in the literature has arisen on the issue of which of the two is a more important explanation. Phelps (1972) argued that true state dependence exists while, by contrast, Cripps and Tarling (1974) argued that what appears to be state dependence is, in fact, spurious state dependence due to the existence of unobserved heterogeneity. Since then, a number of international studies have focussed strongly upon this issue. Chapter 2 above examines the issue in the context of the Australian youth labour market, and Buddelmeyer *et al* (2010) for the adult Australian labour market. Arulampalam (2000, 2004), Arulampalam *et al* (2000), Cappellari *et al* (2010), Henley (2004) and Narendranathan and Elias (1993) have each examined the issue in the context of the UK, Hyslop (1999) and Chay and Hyslop (2000) for the US, Flaig *et al* (1993) and Muhleisen and Zimmermann (1994) for the former West Germany and Raaum and Røed (2006) for Norway. The approach frequently adopted in such studies (as indeed adopted in Chapter 2 of this work) is to postulate a probability distribution for unobserved heterogeneity, and then to control for this when estimating any parameters signifying true state dependence.³

¹“Unobserved” refers to the *researcher’s* ability to measure these characteristics. Participants in the labour market may, in this context, be able to perfectly observe such characteristics, which is how the word “unobserved” should be interpreted.

²Heckman (1981a,b) actually lists 3 conditions which may result in what he terms “spurious state dependence”. Although most of the debate has occurred over the last 30 years, it should be noted that it has been over half a century since Silcock (1954) raised the possibility of such dynamic labour market relationships.

³An alternative approach is to use non-parametric or semi-parametric estimation procedures [see Manski (1987) for theory and applications in a state dependent binary choice environment]. The advantage of these approaches is they allow us to relax distributional assumptions about observed and unobserved heterogeneity.

In order to control for unobserved heterogeneity, panel data is usually required so as to identify the same individuals over time. Empirical results from such studies indicate that, for many countries, evidence of *both* state dependence and unobserved heterogeneity exists. Discussion of empirical results from such studies typically is focussed upon identifying significant differences between parameters which measure state dependence, in different environments. Environments may differ according to geography (e.g. cross-country studies), time (e.g. studies covering different periods for the same country) or observed characteristics of typical respondents (e.g. studies which estimate parameters for both highly-educated and minimally-educated groups).

Two other bodies of literature have examined related, but distinct, issues. The question of whether future wage outcomes are predicted by unemployment spells has been examined by Arulampalam *et al* (2001). There is also a substantial body of literature on duration dependence [e.g. Van Den Berg and Van Ours (1994), Bradley *et al* (2003), Clasen *et al* (2004), Burgess and Turon (2005)] which looks at whether unemployment outcomes depend upon the length of individuals' unemployment spells.

Empirically determining which explanation, if either, is valid may be crucial to assessing the potential effectiveness of many policy interventions. One relevant question is: can short-term policies have long-term outcomes? For example, consider a policy of targetted temporary wage subsidies aimed at assisting the re-employment prospects of the long term unemployed. To the extent that such subsidies will induce those previously unemployed to become employed, and to the extent that true state dependence is driving persistence in employment outcomes, then wage subsidies will not be necessary (on average) to *prolong* such individuals' employment. Thus, a subsidy for the relatively short period of time needed to induce transitions from unemployment to employment would be expected to have persistent effects on the employment for those targetted, even if the subsidy were later discontinued. On the other hand, if spurious state dependence is the dominant force driving persistence then since, by construction, unobserved heterogeneity is temporally persistent, the subsidy would only be effective in prolonging employment for as long as it is continued. This is because, although the subsidy may counterbalance those unobserved characteristics that make unemployment likely, it cannot set in train a dynamic process whereby the characteristics themselves are altered.

Thus, only in an environment where true state dependence exists would this temporary policy intervention of wage subsidies be likely to have permanent employment effects. The key point here is that policy interventions are usually costly; therefore, they may be

deemed more effective the longer their effects persist, since the policy maker may not wish to prolong the intervention indefinitely. By this criterion, policies to alleviate unemployment will be more effective where true state dependence, rather than spurious state dependence, primarily determines employment outcomes.⁴

⁴For example, Arulampalam (2000, p.1) asserts simply that "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."

4.3 Estimation Methodology

This section discusses various methods which may be used to estimate several model specifications, all of which are compatible with observations collected in a panel structure. It is intended to serve as a comprehensive collection of existing techniques, which we set out in detail as we are not aware of another comparable collection elsewhere in the literature. However, it is focussed around the specific problem we are interested in; namely, to estimate the extent to which state dependence in employment outcomes exists, given the presence of unobserved heterogeneity across individuals. Thus, this section also is motivated by wishing to set out clearly the reasons why we place more credence upon some of our empirical results than upon others.

4.3.1 Empirical Model

Consider the general model

$$\begin{aligned} y_{it} &= \begin{cases} 1 & \text{if } y_{it}^* \geq 0 \\ 0 & \text{if } y_{it}^* < 0 \end{cases} \\ y_{it}^* &= \alpha + x'_{it}\beta + \delta y_{i,t-1} + \eta_i + \varepsilon_{it} \end{aligned} \quad (4.1)$$

$i = 1, \dots, N$ and $t = 2, \dots, T$

where:

y_{it} is individual i 's employment status in period t , such that $y_{it} = 1$ if the individual is employed, and zero otherwise

x_{it} is a $k \times 1$ vector of individual i 's measurable characteristics in period t

α is a universal effect which is constant (i.e. time-invariant) across all individuals and time periods.

β is a $k \times 1$ vector of coefficients weighting x'_{it}

δ is the (scalar) coefficient of $y_{i,t-1}$

η_i is an unobserved term which varies across individuals, but is assumed to be time-invariant.

ε_{it} is an unobserved term which varies across both individuals and time periods.

Strictly speaking, both η_i and ε_{it} are measures of unobserved heterogeneity since

they both vary across individuals at any given time. However, we adopt the standard practice of referring *only* to η_i (the time-invariant “individual effect”) when discussing unobserved heterogeneity. This model may be interpreted as each individual having some commonly specified employability index $y'_{it} \equiv y_{it}^* - \varepsilon_{it}$ which may depend upon characteristics observed and unobserved, and upon last period’s employment status. However, the time-varying random component ε_{it} also affects whether each individual is currently employed.

This practice is, in fact, a very limited way of allowing for unobserved heterogeneity. In particular, since η_i is assumed to be constant over time this immediately rules out any idea that unobserved differences between individuals may evolve over time.⁵

Note also that this specification makes a second assumption about the nature of unobserved heterogeneity: namely, that it is *additive*. In a linear setting, this means that only the intercept of someone’s employability index is affected by heterogeneity (i.e. we generalise the model in Chapter 2 by adding a constant α to the unobserved heterogeneity term η_i), while the slope coefficients β and δ are completely independent of η_i . Again, this is a strong assumption and rules out any interaction between unobserved and observed characteristics.⁶ While recognising the limitations of these assumptions, we nonetheless adopt them for the purposes of this study and discuss various specifications which may be viewed as alternative versions of this model.

The specification in Equation (4.1) is a conditional model. It is referred to as dynamic, since it encompasses the possibility that lagged employment status $y_{i,t-1}$ may have some contemporary effect on the probability distribution of current employment status y_{it} . A *static model*, however, is simply a special case of the dynamic model with the additional restriction that $\delta = 0$ ⁷, which means it is an unconditional model. Standard econometric arguments relating to variable inclusion apply here. If the “true model” is the dynamic one (i.e. $\delta \neq 0$), then parameter estimation for the static model will likely be inconsistent due to standard omitted variable bias. In any event, it is necessary for us to estimate δ since primarily we are interested in state dependence: namely, dynamic effects. The parameter of interest, in this context, is δ as the larger is δ then the greater are the effects of state dependence (i.e. the effects of $y_{i,t-1}$ upon the probability distribution of y_{it}) and

⁵The assumption might be suspect if heterogeneity represented, say, individual motivation to work, since motivation could conceivably be affected by spells out of employment as per Hotz *et al* (1988).

⁶For example, again supposing that motivation forms part of unobserved heterogeneity, this may magnify by a scalar factor the likely marginal effects of education on employment status, which would represent an increased coefficient on years of education. As explained, the additive error-components decomposition rules such effects out.

⁷If a static specification were used then the time index would run from 1 to T , rather than from 2 to T .

if $\delta = 0$ then no state dependence exists. We are not primarily interested in estimating the parameter vector of observed coefficients β . However, insofar as we also seek to condition our estimates of state dependence upon observed characteristics, we find it useful to estimate this parameter vector consistently; in this context, we are also interested in estimating α .

An important issue concerns how to choose the distribution function for ε_{it} , $F[c] \equiv \Pr(\varepsilon_{it} \leq c)$. In practice, this amounts to a choice between a linear and non-linear specification.

Among non-linear specifications, two commonly-used models within the literature are the logit model:

$$F_{LOGIT}[c] = \frac{\exp[c]}{1 + \exp[c]} \quad (4.2)$$

and the probit model:

$$F_{PROBIT}[c] = \int_{-\infty}^c \phi(t) dt \quad (4.3)$$

where $\phi(t) \equiv \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ is the probability density function for the standard normal distribution. It is standard for either model to additionally assume that the ε_{it} terms are independently distributed across both individuals and time periods, so that $E[\varepsilon_{it}\varepsilon_{js}] = 0, \forall i \neq j, t \neq s$. This is because independence is the most simplifying, although also most restrictive, assumption.

These two models can be contrasted and compared with the linear probability model (LPM), where the uniform distribution is used:

$$F_{LPM}(c) = c \quad (4.4)$$

Consider the two non-linear models first. The two distribution functions F_{LOGIT} and F_{PROBIT} are similar in shape and both, like F_{LPM} , are monotonically increasing. However, in contrast to F_{LPM} , both F_{LOGIT} and F_{PROBIT} have the desirable property of being bounded by zero and unity:

$$\lim_{c \rightarrow \infty} F[c] = 1, \quad \lim_{c \rightarrow -\infty} F[c] = 0$$

which naturally accommodates interpretation of F_{LOGIT} and F_{PROBIT} as standard prob-

ability measures. The asymptotic behaviour also implies that the marginal effect of a unit change in the index upon the probability of being employed is smallest at the extremes. This means that a unit change in one of the explanatory variables would be most likely to have its smallest effect upon $\Pr(y_{it} = 1)$ where individual i is already virtually certain of being, say, employed [ie. as the upper limit of $\Pr(y_{it} = 1)$ is approached].⁸

Under the linear probability model (LPM), $F_{LPM}(c)$ is monotonic in y'_{it} which is necessary for any model we wish to specify, since y'_{it} is interpreted as an employability index, and so higher values of y'_{it} should *increase* the probability of an individual being employed. Note that although y'_{it} can be interpreted within the LPM as the probability of being employed, yet it is not bounded to lie between 0 and 1 as a standard probability would. Also, the marginal effect of a unit change in the employability index y'_{it} on the probability of being employed is the same for all values of y'_{it} . Finally, the two point distribution implies that the variance of ε_{it} is $\text{var}(\varepsilon_{it}) = [y'_{it} \cdot (1 - y'_{it})]$, which is necessarily heteroskedastic since y'_{it} is a function of the explanatory variables y_{t-1} and x_{it} .

The two non-linear choices for the functional form of F possess some theoretical advantages.⁹ Indeed, most of the empirical labour market literature over the last thirty years, with the exception of Stewart (2007), has avoided the LPM specification when modelling transitions between employment states, due to these three problems of unbounded probabilities, constant marginal effects and heteroskedastic disturbances. However, the LPM may be useful to consider in two distinct circumstances. First, for reasons of tractability, it may be necessary to impose further unpalatable restrictions upon the nonlinear distribution functions, where this would not have been necessary in a linear framework. These may include either imposing the restriction of homoskedasticity upon the disturbance terms ε_{it} , or restricting past, present and future values of measurable characteristics x_{it} to be strictly exogenous¹⁰ to ε_{it} : we highlight cases in Section 4.3.3 below where the LPM allows us to test and potentially relax these restrictions while the nonlinear models do not. Second, there may be reason to believe that a unit change in one of the explanatory variables has a marginal effect of similar magnitude for all values of the employability index $g(\dots)$, over the *observed* range of measurable characteristics x_{it} . In the first circumstance, then the LPM may be used as an approximation; in the second circumstance, as a more accurate

⁸The same argument applies to the lower limit [ie. where individual i is almost certain to be unemployed].

⁹At least 4 other models also possess these properties, but are not as tractable for estimation purposes. See Amemiya (1981) for a more specific comparison of these two, and other, discrete choice models.

¹⁰Variable x_{it} is said to be strictly exogenous if it is uncorrelated with past, present and future shocks, so that $E[x_{it}\varepsilon_{is}] = 0 \quad \forall s, t$

specification.

Put another way, if changes in $g(\dots)$ affect y_{it} in a severely nonlinear manner, then the LPM will yield severely biased and misleading results due to its underlying misspecification. Yet if the underlying nonlinearities are modest, while there is reason to believe the nonlinear models are misspecified in another dimension (such as in the assumptions they make about the distributions of ε_{it} or η_i), then linear methods may effectively avoid introducing other potentially larger biases of nonlinear estimation.¹¹ Essentially, then, our judgment on the LPM involves a trade-off between its greater analytic flexibility (especially in a dynamic setting) and the unattractiveness of some of its assumptions.

¹¹Stewart (2007, p.513) notes that versions of the linear probability model "handle unobserved heterogeneity in a less restrictive way" (than the dynamic random-effects probit model).

4.3.2 Estimation of Nonlinear Specifications

Using the Logit specification in (4.2) or the Probit specification in (4.3), the model (4.1) contains two potential sources of random variation: time-varying and time-invariant. Each time-varying component ε_{it} is taken to be a stochastic draw from distribution $F[\cdot]$, as described in Section 4.3.1 above. To allow for the time-invariant individual effects η_i , it is necessary to use estimation techniques which explicitly recognise these, rather than potentially introducing bias by using conventional 'pooled' logit or probit models (which effectively impose the restriction $\eta_i = 0, \forall i$). In particular, it may be necessary to specify the distribution from which the η_i terms arise. On this issue, the literature contains two alternative approaches: it is possible either to specify individual effects as "fixed", or as "random." Under a specification of fixed effects, no originating relationship between any given two individuals' η_i terms is postulated. Thus, no restrictions are placed upon the distribution of η_i and all η_i terms are treated as fixed constants. Under a specification of random effects, however, there exists a joint distribution between x_{it} and η_i which must be specified, and the η_i term for each individual i is initially drawn from a common marginal distribution. Thus, an originating relationship necessarily exists between the η_i terms, and between η_i and x_{it} , under a random effects specification.

Estimation of the Fixed Effects Specification

Where individual effects η_i are taken to be fixed constants, then it is not possible to identify their effects separately from those of any observed time-invariant explanatory variables (such as year of birth). Such variables must therefore be excluded from analysis in order to estimate the other parameters of the model (as would be the case for estimation of a static fixed effects specification). This limits estimation capabilities somewhat; however, far more serious problems exist due to the dynamic nature of the specification, whether linear or nonlinear. For a dynamic *linear* model, a differencing procedure can be used to eliminate η_i in each of the $N(T-1)$ equations. So, for example, the LPM can be rewritten as:

$$y_{it} - \bar{y}_{i,t} = \delta (y_{i,t-1} - \bar{y}_{i,t-1}) + (x'_{i,t} - \bar{x}'_{i,t}) \beta + (\varepsilon_{it} - \bar{\varepsilon}_{i,t})$$

where the 'bar' symbol denotes individual average (e.g. $\bar{y}_{i,t} \equiv \frac{1}{T-1} \sum_{t=2}^T y_{it}$, $\bar{y}_{i,t-1} \equiv \frac{1}{T-1} \sum_{t=2}^T y_{i,t-1}$, etc). However, correlation necessarily exists between the transformed ex-

planatory variables and the transformed stochastic errors, since $\bar{y}_{i,t-1}$ is a function of lagged ε_i terms. Therefore, if standard linear estimation methods are applied to this transformed model, then δ is estimated with bias.¹² A similar issue arises in relation to dynamic *non-linear* models and this bias shrinks only as $T \rightarrow \infty$, not as $N \rightarrow \infty$. See Heckman (1981b) for further discussion of this.

Chamberlain Estimator In the absence of any explanatory variables other than the lagged dependent variable, Chamberlain (1985) has shown, for $T \geq 4$, that a method exists to consistently estimate the state dependence parameter δ from a dynamic logit model with either fixed or random effects. Chamberlain's model also specifies the probability of employment in the initial period. In its simplest form, where $T = 4$, the full model is:

$$\begin{aligned} g_0(\eta_i) &= \Pr(y_{i1} = 1 \mid \eta_i) \\ g_A(y_{i,t-1}, \eta_i) &\equiv \delta y_{i,t-1} + \eta_i \\ y_{i1} &= I[g_0(\eta_i) + \varepsilon_{i1}] \\ y_{it} &= I[g_A(y_{i,t-1}, \eta_i) + \varepsilon_{it}] \text{ for } t = 2, \dots, 4 \\ \Pr(\varepsilon_{it} \leq c) &= \frac{\exp[c]}{1 + \exp[c]} \text{ for } t = 1, \dots, 4 \end{aligned}$$

It follows immediately by Bayesian induction that

$$\begin{aligned} \Pr[(y_{i1} = \theta_{i1} \cap y_{i2} = 0 \cap y_{i3} = 1 \cap y_{i4} = \theta_{i4}) \mid (\eta_i, y_{i2} + y_{i3} = 1)] &= \frac{1}{1 + \exp[\delta(\theta_{i1} - \theta_{i4})]} \\ \Pr[(y_{i1} = \theta_{i1} \cap y_{i2} = 1 \cap y_{i3} = 0 \cap y_{i4} = \theta_{i4}) \mid (\eta_i, y_{i2} + y_{i3} = 1)] &= \frac{\exp[\delta(\theta_{i1} - \theta_{i4})]}{1 + \exp[\delta(\theta_{i1} - \theta_{i4})]} \end{aligned}$$

where $\theta_{i1}, \theta_{i4} \in \{0, 1\}$. These probabilities are independent of the individual effect η_i , and hence avoid the construction of estimation bias. In Chamberlain's model, estimation of δ proceeds by conditioning upon data from those individuals who experienced changes in employment status between periods 2 and 3. The Chamberlain conditional maximum likelihood estimator of δ is found by maximising, with respect to δ , the log likelihood

¹²See Nickell (1981) for a derivation of the size of bias in estimation of this dynamic linear model.

function corresponding to the specification:

$$\log L_{FE} = \sum_{i=1}^N I[y_{i2} + y_{i3} = 1] \cdot \{y_{i2} [\delta(y_{i1} - y_{i4})] - \log[1 + \exp[\delta(y_{i1} - y_{i4})]]\} \quad (4.5)$$

Chamberlain (1985) demonstrates that this estimator is consistent as $N \rightarrow \infty$. The estimator can easily be extended to $T > 4$ using the analogous principle, that estimation proceeds conditional on the employment histories of those individuals changing employment status in intermediate periods, which are a set of sufficient statistics for δ . However, Chamberlain's estimator is not capable of consistently estimating δ in the presence of additional explanatory variables x_{it} . This is because the corresponding probability statements, where $g(\dots) = \delta y_{i,t-1} + x'_{it}\beta + \eta_i$, are no longer independent of η_i .

Honore and Kyriazidou Estimator To extend Chamberlain's model, Honore and Kyriazidou (2000) consider estimation of the dynamic logit in the presence of additional explanatory variables which are strictly exogenous. Their model, for $T = 4$, is:

$$\begin{aligned} g_0(\eta_i) &= \Pr(y_{i1} = 1 \mid \eta_i) \\ g_H(y_{i,t-1}, \eta_i) &\equiv \delta y_{i,t-1} + x'_{it}\beta + \eta_i \\ y_{i1} &= I[g_0(\eta_i) + \varepsilon_{i1}] \\ y_{it} &= I[g_H(y_{i,t-1}, \eta_i) + \varepsilon_{it}] \text{ for } t = 2, \dots, 4 \\ \Pr(\varepsilon_{it} \leq c) &= \frac{\exp[c]}{1 + \exp[c]} \text{ for } t = 1, \dots, 4 \end{aligned}$$

Now define events V and W as follows:

$$\begin{aligned} V &\equiv \{y_{i1} = \theta_{i1} \cap y_{i2} = 0 \cap y_{i3} = 1 \cap y_{i4} = \theta_{i4}\} \\ W &\equiv \{y_{i1} = \theta_{i1} \cap y_{i2} = 1 \cap y_{i3} = 0 \cap y_{i4} = \theta_{i4}\} \end{aligned}$$

Honore and Kyriazidou (2000) demonstrate that, if $x_{i3} = x_{i4}$, then:

$$\begin{aligned}\Pr(V \mid (\eta_i, V \cup W, x_{i3} = x_{i4})) &= \frac{1}{1 + \exp[(x_{i2} - x_{i3})\beta + \delta(\theta_{i1} - \theta_{i4})]} \\ \Pr(W \mid (\eta_i, V \cup W, x_{i3} = x_{i4})) &= \frac{\exp[(x_{i2} - x_{i3})\beta + \delta(\theta_{i1} - \theta_{i4})]}{1 + \exp[(x_{i2} - x_{i3})\beta + \delta(\theta_{i1} - \theta_{i4})]}\end{aligned}$$

both of which are independent of η_i . The underlying principle is to identify δ from the employment data of those individuals whose exogenous characteristics are unchanged in the final two periods 3 and 4. Further, β is identified from these same individuals, by looking at changes in their exogenous characteristics in the middle periods 2 and 3. However, this strict principle may be impossible to apply, especially if data on observed characteristics is continuous (for then there are no individuals with time-stationary characteristics in the final two periods). To deal with this eventuality, Honore and Kyriazidou propose an estimator which can use data from individuals whose characteristics in the final two periods are not necessarily identical. This estimator is the argument in (β, δ) which maximises the function:

$$HK_4 \equiv \sum_{i=1}^N I[y_{i2} + y_{i3} = 1] \cdot K\left[\frac{x_{i3} - x_{i4}}{\sigma_N}\right] \cdot \log\left[\frac{\exp[(x_{i2} - x_{i3})'\beta + \delta(y_{i1} - y_{i4})]^{y_{i2}}}{1 + \exp[(x_{i2} - x_{i3})'\beta + \delta(y_{i1} - y_{i4})]}\right] \quad (4.6)$$

where $K[\cdot]$ is a kernel weighting function which weights individuals' data in inverse proportion to individual differences between x_{i3} and x_{i4} , and σ_N is a bandwidth that approaches zero as $N \rightarrow \infty$. For $T > 4$, the objective function to be maximised becomes:

$$\begin{aligned}HK_T \equiv & \sum_{i=1}^N \sum_{2 \leq s < t \leq T} I[y_{is} + y_{it} = 1] \cdot K\left[\frac{x_{i,s+1} - x_{i,t+1}}{\sigma_N}\right] \times \\ & \log\left[\frac{\exp[(x_{is} - x_{it})'\beta + \delta(y_{i,s-1} - y_{i,t+1}) + \delta(y_{i,s+1} - y_{i,t-1}) \cdot I(t - s \geq 2)]^{y_{is}}}{1 + \exp[(x_{is} - x_{it})'\beta + \delta(y_{i,s-1} - y_{i,t+1}) + \delta(y_{i,s+1} - y_{i,t-1}) \cdot I(t - s \geq 2)]^{y_{is}}}\right]\end{aligned} \quad (4.7)$$

and the estimator $\Theta_{HK} \equiv (\hat{\beta}_{HK}, \hat{\delta}_{HK})$ is the argument in (β, δ) which maximises HK_T . Standard errors for $(\hat{\beta}_{HK}, \hat{\delta}_{HK})$ are not capable of being estimated analytically, and must be obtained using the bootstrap resampling procedure due to Efron [see Efron (1979), Efron and Tibshirani (1993)]. The estimator also suffers from the considerable drawback that it is incapable of generating predicted employment probabilities; neither is it capable of estimating coefficients on time-invariant characteristics (including a constant term, or time

dummies used as explanatory variables), since it requires that $x_{is} - x_{it}$ have support in a neighbourhood of 0 for any $t \neq s$ (Hsiao, 2003). However, a key advantage to using this procedure lies in the estimator's robustness to various relationships between individual effects, initial conditions and explanatory variables. In addition, it possesses the highly desirable characteristics of consistency and asymptotic normality as $N \rightarrow \infty$. If $\Pr(x_{ir} = x_{i,r+1}) > 0$ for some value of r (due to, for example, discrete explanatory variables or controlled experiments), and if $x_{i,r-1} - x_{ir}$ has nonzero variation conditional upon $x_{ir} = x_{i,r+1}$, then Θ_{HK} converges to the true vector (β, δ) at the usual rate $\sqrt{N}$. However, if in practice the explanatory variables used are continuous and uncontrolled for in repeated experiments, then Θ_{HK} converges at the potentially much slower rate $\sqrt{N \cdot (\sigma_N)^k}$, where k is the number of explanatory variables included.

Estimation of the Random Effects Specification

Where individual effects η_i are taken to arise from a stochastic process, it is necessary to model more than the marginal distribution of η_i . Additionally, to avoid plausible sources of bias due to omitted variables, the relationships between each individual effect and the explanatory variables x_{it} and initial conditions (x_{i1}, y_{i1}) are modelled.

First, for reasons of tractability, it is standard to assume that the marginal distribution of η_i is normal (i.e. Gaussian) and this is adopted in all models here. Second, it is possible, for reasons of simplification, to assume that η_i is orthogonal to all explanatory variables x_{it} so that $\text{cov}(\eta_i, x_{it}) = 0 \forall i, t$. However, it is by no means necessary to adopt such a restrictive specification, and some of the models in this section employ a considerably more general specification. Third, several specifications of initial conditions are examined in this section, starting with the most restrictive and ending with the most general.

In addition to specifying the distribution of η_i , it is necessary to specify the distribution of the transitory disturbances ε_{it} (as is also necessary for the fixed effects models above). Since the logit specification $F_{LOGIT}[c] = \frac{\exp[c]}{1+\exp[c]}$ is particularly tractable for fixed effects models, we also adopt it for random effects models, so as to allow clear comparison of results.

Exogenous Initial Conditions The simplest, and most restrictive, specification of initial conditions assumes that there is no relationship between initial conditions and either individual effects or explanatory variables. If we also assume normally distributed individ-

ual effects, logistically distributed transitory disturbances, and individual effects orthogonal to all explanatory variables, then this amounts to placing the following restrictions on the model:

$$\begin{aligned}
 \eta_i &\sim iid N(0, \sigma_\eta^2) \\
 F_{LOGIT}[c] &= \frac{\exp[c]}{1 + \exp[c]} \\
 cov(\eta_i, x_{it}) &= 0 \quad \forall i, t \\
 cov(\eta_i, \varepsilon_{it}) &= 0 \quad \forall i, t \\
 \Pr(y_{i1} | x_{i1}, \eta_i) &= \Pr(y_{i1}) \quad \forall i
 \end{aligned}$$

Under this specification, the likelihood function for the full sample is:

$$L_{EX} = \prod_{i=1}^N \int \prod_{t=2}^T \left(\frac{\exp[\alpha + x'_{it}\beta + \delta y_{i,t-1} + \eta_i]}{1 + \exp[\alpha + x'_{it}\beta + \delta y_{i,t-1} + \eta_i]} \right) d\phi(\eta_i, \sigma_\eta^2) \quad (4.8)$$

where $\phi(\eta_i) = \frac{1}{\sqrt{2\pi\sigma_\eta^2}} \exp\left(-\frac{\eta_i^2}{2\sigma_\eta^2}\right)$ is the probability density function for the normal distribution, imposed on the individual effects η_i . The estimator $\Theta_{EX} \equiv \arg \max(L_{EX}) = (\hat{\alpha}_{EX}, \hat{\beta}_{EX}, \hat{\delta}_{EX}, \hat{\sigma}_{\eta EX}^2)$ estimates the corresponding parameters consistently as $N \rightarrow \infty$ by MLE, if the assumptions are valid. In particular, however, it should be noted that if either observed differences between individuals (x_{it}) or unobserved differences (η_i) play a role in determining the initial period's employment status, then Θ_{EX} does not estimate δ (i.e. state dependence in the employment-generating process) consistently.

Equilibrium Initial Conditions Alternatively, we consider adopting the assumption that the employment generating process is in dynamic equilibrium (i.e. steady state) at the beginning of the sample period. Under this assumption, the initial period observed is no different from any other period in terms of any individual's employment probability during it, as the employment process - by construction - has been proceeding before the initial

period. Therefore, the following implicit relationship holds:

$$\begin{aligned} p_i &\equiv \Pr(y_{i,1} = 1 \mid x_{i,1}, \eta_i) \\ &= \frac{p_i^-}{1 + p_i^- - p_i^+} \end{aligned}$$

where $p_i^+ \equiv \Pr(y_{it} = 1 \mid y_{i,t-1} = 1, x_{i,1}, \eta_i), \forall t$ and $p_i^- \equiv \Pr(y_{it} = 1 \mid y_{i,t-1} = 0, x_{i,1}, \eta_i), \forall t$ are the staying and moving probabilities to employment, conditional upon the previous period's employment outcome, so that p_i^+ and p_i^- are constants which do not vary across time periods. If we again assume normally distributed individual effects, logistically distributed transitory disturbances, and no relationship between individual effects and explanatory variables, then this amounts to placing the following restrictions on the model:

$$\begin{aligned} \eta_i &\sim iid N(0, \sigma_\eta^2) \\ F_{LOGIT}[c] &= \frac{\exp[c]}{1 + \exp[c]} \\ cov(\eta_i, x_{it}) &= 0 \quad \forall i, t \\ cov(\eta_i, \varepsilon_{it}) &= 0 \quad \forall i, t \\ p_i^+ &\equiv \frac{\exp[x'_{it}\beta + \delta + \eta_i]}{1 + \exp[x'_{it}\beta + \delta + \eta_i]}, \forall i, t \\ p_i^- &\equiv \frac{\exp[x'_{it}\beta + \eta_i]}{1 + \exp[x'_{it}\beta + \eta_i]}, \forall i, t \\ \Pr(y_{i1} \mid x_{i1}, \eta_i) &\equiv p_i = \frac{p_i^-}{1 + p_i^- - p_i^+} \end{aligned}$$

Under this specification, the likelihood function for the full sample is:

$$L_{EQ} = \prod_{i=1}^N \int \prod_{t=2}^T \left(\frac{\exp[\alpha + x'_{it}\beta + \delta y_{i,t-1} + \eta_i]}{1 + \exp[\alpha + x'_{it}\beta + \delta y_{i,t-1} + \eta_i]} \right) \cdot p_i^{y_{i1}} (1 - p_i)^{1-y_{i1}} d\phi(\eta_i, \sigma_\eta^2) \quad (4.9)$$

and the estimator $\Theta_{EQ} \equiv \arg \max (L_{EQ}) = (\hat{\alpha}_{EQ}, \hat{\beta}_{EQ}, \hat{\delta}_{EQ}, \hat{\sigma}_{\eta EQ}^2)$ estimates the corresponding parameters consistently as $N \rightarrow \infty$ by MLE, if the assumptions are valid. In this case, we note that the process is only likely to satisfy the restriction of being in dy-

namic equilibrium if the explanatory variables used either are themselves stationary, or exert negligible joint influence on employment status.

Reduced Form Initial Conditions A considerably less restrictive specification than either of the previous two uses Heckman's (1981b) reduced form approach, which models initial conditions as depending upon both observed and unobserved individual characteristics. Under this specification,

$$y_{i1} = I \left[\alpha_1 + x'_{i1}\beta_1 + \gamma_1\eta_i + \varepsilon_{i1} \right]$$

where $I[\cdot]$ denotes an indicator function which takes the value 1 if its argument is strictly positive and 0 otherwise.

Under this specification, employment status in the initial period y_{i1} is assumed to be a function of x_{i1} and η_i , where the initial period's coefficient vector on explanatory variables, β_1 , is permitted to differ from that in all other periods, β , and the individual effect is permitted to vary from other periods' individual effects by multiplicative factor γ_1 . If $\gamma_1 = 0$ then this case corresponds to the specification with exogenous initial conditions discussed earlier in this section. Initial employment state y_{i1} is also subject to a transitory disturbance term ε_{i1} , which is assumed to follow the same logistic distribution as the other periods' disturbance terms. Again assuming normally distributed individual effects, logistically distributed transitory disturbances, and no relationship between individual effects and explanatory variables, the reduced form specification amounts to the following restrictions on the model:

$$\begin{aligned} \eta_i &\sim iid N(0, \sigma_\eta^2) \\ F_{LOGIT}[c] &= \frac{\exp[c]}{1 + \exp[c]} \\ cov(\eta_i, x_{it}) &= 0 \quad \forall i, t \\ cov(\eta_i, \varepsilon_{it}) &= 0 \quad \forall i, t \\ \Pr(y_{i1} | x_{i1}, \eta_i) &= \frac{\exp[\alpha + x'_{i1}\beta_1 + \gamma_1\eta_i]}{1 + \exp[\alpha + x'_{i1}\beta_1 + \gamma_1\eta_i]} \quad \forall i \end{aligned}$$

Under this specification, the likelihood function for the full sample is:

$$L_{RF} = \prod_{i=1}^N \int \prod_{t=2}^T \left(\frac{\exp [\alpha + x'_{it}\beta + \delta y_{i,t-1} + \eta_i]}{1 + \exp [\alpha + x'_{it}\beta + \delta y_{i,t-1} + \eta_i]} \right) \cdot \left(\frac{\exp [\alpha + x'_{i1}\beta_1 + \gamma_1\eta_i]}{1 + \exp [\alpha + x'_{i1}\beta_1 + \gamma_1\eta_i]} \right) d\phi(\eta_i, \sigma_\eta^2) \quad (4.10)$$

from which we can form the estimator

$\Theta_{RF} \equiv \arg \max(L_{RF}) = \left(\hat{\alpha}_{RF}, \hat{\beta}_{RF}, \hat{\delta}_{RF}, \hat{\alpha}_{1RF}, \hat{\beta}_{1RF}, \hat{\gamma}_{1RF}, \hat{\sigma}_{\eta RF}^2 \right)$ which is again a consistent estimator for the corresponding parameters as $N \rightarrow \infty$, if the assumptions are valid.

Restricted Correlation and Reduced Form Initial Conditions The previous three models, although providing more or less general specifications for initial conditions, all assume the individual effects η_i are orthogonal to all explanatory variables x_{it} . To relax this assumption, we consider the possibility that individual effects may be described by a stochastic process which takes as its arguments observed time-varying characteristics (Chamberlain (1984)). Assuming the form of the relationship to be linear, then:

$$\begin{aligned} \eta_i &= x'_{i1}\pi + x'_{i2}\pi + \dots + x'_{iT}\pi + \zeta_i \\ &= \frac{1}{T} (x'_{i1} + x'_{i2} + \dots + x'_{iT}) \cdot (T\pi) + \zeta_i \end{aligned}$$

where π is a $(k_v \times 1)$ parameter vector, k_v is the number of time-varying explanatory variables included and ζ_i is a random term assumed to be normally distributed and independent of all other disturbance terms. Continuing to use the most general specification of initial conditions (i.e. the reduced form) in the same manner as above, this amounts to placing the following restrictions on the model:

$$\begin{aligned} F_{LOGIT}[c] &= \frac{\exp[c]}{1 + \exp[c]} \\ \eta_i &= \sum_{t=1}^T x'_{it}\pi + \zeta_i \quad \forall i \\ \zeta_i &\sim iid \ N(0, \sigma_\zeta^2) \\ cov(\eta_i, \varepsilon_{it}) &= 0 \quad \forall i, t \\ \Pr(y_{i1} | x_{i1}, \eta_i) &= \frac{\exp[\alpha + x'_{i1}\beta_1 + \gamma_1\eta_i]}{1 + \exp[\alpha + x'_{i1}\beta_1 + \gamma_1\eta_i]} \quad \forall i \end{aligned}$$

Under this specification, the likelihood function for the full sample is:

$$L_{RES} = \prod_{i=1}^N \int \prod_{t=2}^T \left(\frac{\exp \left[\alpha + x'_{it} \beta + \delta y_{i,t-1} + \sum_{i=1}^T x'_{it} \pi + \zeta_i \right]}{1 + \exp \left[\alpha + x'_{it} \beta + \delta y_{i,t-1} + \sum_{i=1}^T x'_{it} \pi + \zeta_i \right]} \right) \cdot \left(\frac{\exp \left[\alpha + x'_{i1} \beta_1 + \gamma_1 \left(\sum_{i=1}^T x'_{it} \pi + \zeta_i \right) \right]}{1 + \exp \left[\alpha + x'_{i1} \beta_1 + \gamma_1 \left(\sum_{i=1}^T x'_{it} \pi + \zeta_i \right) \right]} \right) d\phi(\zeta_i, \sigma_\zeta^2) \quad (4.11)$$

from which we can form the estimator

$$\Theta_{RES} \equiv \arg \max (L_{RES}) = \left(\hat{\alpha}_{RES}, \hat{\beta}_{RES}, \hat{\delta}_{RES}, \hat{\alpha}_{1RES}, \hat{\beta}_{1RES}, \hat{\gamma}_{1RES}, \hat{\pi}_{RES}, \hat{\sigma}_{\eta_{RES}}^2 \right)$$

which is again consistent for the corresponding parameters as $N \rightarrow \infty$, if the assumptions are valid. Θ_{RES} allows for "restricted" correlation between η_i and x_{it} , as it assumes that characteristics exert the same influence over the individual effect in *all* time periods, which effectively imposes the restriction that π is constant over time. The substance of this restriction is that the time invariant η_i varies across individuals only with the time-invariant average value of the x_{it} characteristics, and not with the time path of the x_{it} characteristics.

Unrestricted Correlation and Reduced Form Initial Conditions Finally, we relax the restriction that the coefficient vector π is constant over time, in the equation relating individual effects to observed characteristics. The least restrictive linear relationship is:

$$\eta_i = x'_{i1} \pi_1 + x'_{i2} \pi_2 + \dots + x'_{iT} \pi_T + \zeta_i$$

where each π_t is a $(k_v \times 1)$ parameter vector, k_v is the number of time-varying explanatory variables included and ζ_i is again a random term assumed to be normally distributed and independent of all other disturbance terms. Continuing to use the reduced form specification

of initial conditions, this amounts to placing the following restrictions on the model:

$$\begin{aligned}
F_{LOGIT}[c] &= \frac{\exp[c]}{1 + \exp[c]} \\
\eta_i &= \sum_{t=1}^T x'_{it}\pi_t + \zeta_i \quad \forall i \\
\zeta_i &\sim iid \ N(0, \sigma_\zeta^2) \\
cov(\eta_i, \varepsilon_{it}) &= 0 \quad \forall i, t \\
\Pr(y_{i1} | x_{i1}, \eta_i) &= \frac{\exp[\alpha + x'_{i1}\beta_1 + \gamma_1\eta_i]}{1 + \exp[\alpha + x'_{i1}\beta_1 + \gamma_1\eta_i]} \quad \forall i
\end{aligned}$$

Under this final specification, the likelihood function for the full sample is:

$$\begin{aligned}
L_{UN} &= \prod_{i=1}^N \int \prod_{t=2}^T \left(\frac{\exp[\alpha + x'_{it}\beta + \delta y_{i,t-1} + \sum_{t=1}^T x'_{it}\pi_t + \zeta_i]}{1 + \exp[\alpha + x'_{it}\beta + \delta y_{i,t-1} + \sum_{t=1}^T x'_{it}\pi_t + \zeta_i]} \right) \cdot \quad (4.12) \\
&\quad \left(\frac{\exp[\alpha + x'_{i1}\beta_1 + \gamma_1(\sum_{t=1}^T x'_{it}\pi_t + \zeta_i)]}{1 + \exp[\alpha + x'_{i1}\beta_1 + \gamma_1(\sum_{t=1}^T x'_{it}\pi_t + \zeta_i)]} \right) d\phi(\zeta_i, \sigma_\zeta^2)
\end{aligned}$$

from which we can form the estimator

$$\begin{aligned}
\Theta_{UN} &\equiv \arg \max (L_{UN}) \\
&= (\hat{\alpha}_{UN}, \hat{\beta}_{UN}, \hat{\delta}_{UN}, \hat{\alpha}_{1UN}, \hat{\beta}_{1UN}, \hat{\gamma}_{1UN}, \hat{\pi}_{1UN}, \dots, \hat{\pi}_{T,UN}, \hat{\sigma}_{\eta UN}^2) \text{ which is} \\
&\text{again consistent for the corresponding parameters as } N \rightarrow \infty, \text{ if the assumptions are valid.}
\end{aligned}$$

Figure 4.1 illustrates a taxonomy of the nonlinear models used. The fixed effects specification HK is the most general. Among the five random effects specifications, UN is the most general and the other four are special cases of UN. Specifically, RF is a special case of RES and both EQ and EX are special cases of RF.

In choosing between different specifications, our primary goal is to avoid choosing any which make invalid assumptions, as to do otherwise may introduce bias and inconsistency into our parameter estimation. However, we should also endeavour to choose the most efficient specification (i.e. that with the minimum sampling variance from among the group of valid specifications). The fundamental issue is which criteria should be used to decide whether the restrictions used by any given specification are valid or invalid. We propose to use the test statistic due to Hausman (1978), which is a function of the param-

Taxonomy of Random Effects Models

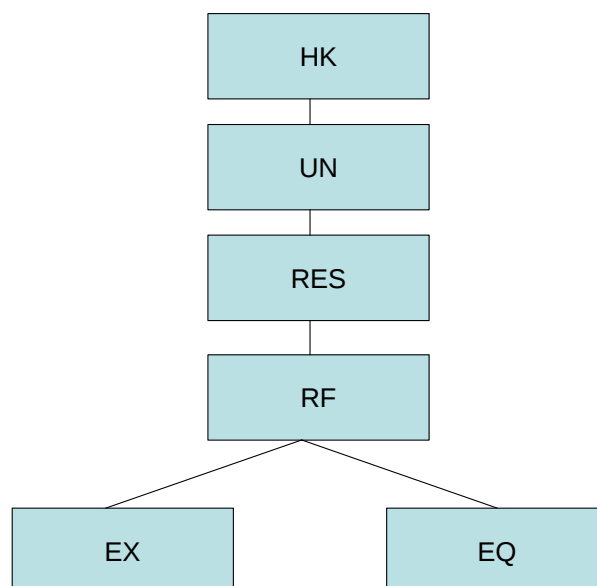


Figure 4.1: Taxonomy of Random Effects Models

eter estimators (i.e. random variables defined according to each respective estimation rule) from, respectively, the efficient specification¹³ and a less efficient specification. This test statistic takes the form:

$$H \equiv (\widehat{\varphi}_R - \widehat{\varphi}_U)' \widehat{V}^{-1} (\widehat{\varphi}_R - \widehat{\varphi}_U) \quad (4.13)$$

where $\widehat{\varphi}_R$ and $\widehat{\varphi}_U$ are the most restrictive and less restrictive parameter estimators respectively, and $\widehat{V} \equiv \widehat{var}(\widehat{\varphi}_U - \widehat{\varphi}_R)$. Hausman (1978, pp.1253-1254) has proven, under the null hypothesis that both specifications are valid, that $\widehat{V} = \widehat{var}(\widehat{\varphi}_U) - \widehat{var}(\widehat{\varphi}_R)$ and that H follows an asymptotic chi-square distribution with degrees of freedom equal to the number of parameters estimated when no misspecification is present. In this context, misspecification is defined by Hausman as a failure of the orthogonality assumption $E(\varepsilon | X) = 0$ for one or more variables X and error term ε .

Hausman's H statistic can be used to test the random effects specification against the fixed effects specification, using these principles. Hausman (1978, pp.1261-1264) proposed to use this test statistic in the context of two linear estimation methods, one which is used to estimate parameters from a random effects specification and the other to estimate parameters from a fixed effects specification. If the random effects specification is the correct specification, then the first estimation method is asymptotically efficient while the second estimation method is unbiased and consistent but not efficient. Hausman notes that "when the econometrician finds his estimates (from the fixed effects estimator) to be unsatisfactory, this evidence is a finding against his specification, not his choice of estimator" (Hausman 1978, p.1263).

In the context of this work, since the estimator Θ_{HK} used to estimate parameters from the fixed effects specification refrains from making assumptions about orthogonality of observed variables with respect to the individual effects, it represents a less restrictive specification than any of those using random effects. Therefore, Hausman's test may be used to assess whether any of the random effects specifications is valid. To do this, we may consider any of the parameter estimators mentioned in this section in the context of random effects, which we generically denote here as Θ_{RE} , and construct the null hypothesis that Θ_{RE} is efficient. We then propose to compare Θ_{RE} against Θ_{HK} and use Hausman's H statistic in (4.13) to test the null hypothesis that the more restrictive random effects specification is valid and efficient¹⁴. If there is no significant evidence to reject the null

¹³The efficient specification is defined as that which attains the asymptotic Cramer-Rao bound: see Hausman (1978, p.1253).

¹⁴Note that if a specification is efficient then it is also valid.

hypothesis, then we may prefer to use one of the random effects specifications so that we may use a more precise estimation method; otherwise, we may prefer to use the fixed effects specification to avoid inconsistent parameter estimation.

4.3.3 Estimation of Linear Specifications

As discussed in Section 4.3.1 above, the LPM in (4.4) is a special case of the fundamental model in (4.1) where the time-varying errors ε_{it} follow the uniform probability distribution. However, we will refer to it using the simpler equivalent form:

$$y_{it} = a + gy_{i,t-1} + x'_{it}b + \eta_i + \varepsilon_{it} \quad (i = 1, \dots, N) \quad (t = 2, \dots, T) \quad (4.14)$$

where we use different notation from that used for the nonlinear estimators, to emphasise the very different distributional assumptions made.

For the same reasons discussed in the context of the nonlinear estimators, static methods such as pooled ordinary least squares (OLS), fixed effects and random effects which effectively assume that $g = 0$ are not consistent estimators for either a, b or g for fixed T .¹⁵ Note that the dynamic model (4.14) is conditional upon $y_{i,t-1}$ while, by contrast, the model with the restriction that $g = 0$ is an unconditional model. OLS estimates parameters in the dynamic model inconsistently, since even if η_i is uncorrelated with x_{it} it will still be correlated with the dependent variable $y_{i,t-1}$ in the equation for period $t - 1$, and so η_i will also be correlated with the lagged dependent variable $y_{i,t-1}$ in the equation for period t . As a result, the composite error term $(\eta_i + \varepsilon_{it})$ will be correlated with the explanatory variables, which gives rise to inconsistency of parameter estimation. Taking first differences of the data from individual means (as performed by static fixed and random effects) can deal with this issue, but as the first difference transformations themselves introduce other correlations between transformed errors and transformed explanatory variables (as discussed previously), then this merely substitutes one source of inconsistency for another.

Consistent estimators using principles of Instrumental Variable Estimation for (the above) have been proposed by Anderson and Hsiao (1981), Arellano and Bond (1991) and Blundell and Bond (1998). See also Bond (2002) for a summary. These models, which can be estimated using the "DPD for Ox" module within the Ox programming language developed by Doornik *et al* (2002)¹⁶, or using Stata's routines *xtabond* and *xtabond2*, share

¹⁵Calculations of the asymptotic fixed T biases associated with these three conventional estimators (in the context of a dynamic linear model) can be found in, respectively, Hsiao (2003, p.74), Nickell (1981) and Trognon (1978).

¹⁶The DPD for Ox module, in particular, provides great flexibility of specification and a wide choice of reported diagnostic statistics.

a common form. First write the $(T - 1)$ equations for individual i in the stacked form:

$$\mathbf{y}_i = \mathbf{W}_i \boldsymbol{\mu} + \boldsymbol{\iota}_i \eta_i + \boldsymbol{\varepsilon}_i$$

where $\boldsymbol{\mu} \equiv (a \ g \ b)$, $\mathbf{W}_i \equiv (\boldsymbol{\iota}_i \ y_{i,t-1} \ x'_{it})$, $\boldsymbol{\varepsilon}_i$ is individual i 's vector of stacked time-varying errors and $\boldsymbol{\iota}_i$ is a $(T - 1)$ vector of ones. Using this notation, then various linear estimators of $\boldsymbol{\mu}$ can be calculated, all of the general form:

$$\hat{\boldsymbol{\mu}} = [(\sum_i \mathbf{W}_i^{*'} \mathbf{Z}_i) \mathbf{A}_N (\sum_i \mathbf{Z}_i' \mathbf{W}_i^*)]^{-1} (\sum_i \mathbf{W}_i^{*'} \mathbf{Z}_i) \mathbf{A}_N (\sum_i \mathbf{Z}_i' \mathbf{y}_i^*) \quad (4.15)$$

where $\mathbf{A}_N \equiv (\frac{1}{N} \sum_i \mathbf{Z}_i' \mathbf{H} \mathbf{Z}_i)^{-1}$ and $\mathbf{W}_i^*$ and $\mathbf{y}_i^*$ are the same transformation of $\mathbf{W}_i$ and $\mathbf{y}_i$ which could be, for example, the identity transformation (i.e. levels) y_{it} , first differences $y_{it} - y_{i,t-1}$, deviations from individual means $y_{it} - \bar{y}_i$ or combinations of these. $\mathbf{Z}_i$ is a matrix of instrumental variables and $\mathbf{H}$ is a matrix which weights observations across the N individuals.

Difference GMM Estimators

The dynamic model in levels contains the individual effect η_i in the composite error term $(\eta_i + \varepsilon_{it})$, which is necessarily correlated with the lagged dependent variable $y_{i,t-1}$. Therefore, the model's parameters can only be estimated consistently in levels (i.e. using the identity transformation) if a matrix of instruments uncorrelated with η_i can be found. Where there are no such instruments available, then the first difference transformation can be used to eliminate η_i from the transformed error term:

$$y_{it} - y_{i,t-1} = g(y_{i,t-1} - y_{i,t-2}) + (x'_{it} - x'_{i,t-1})b + (\varepsilon_{it} - \varepsilon_{i,t-1})$$

which eliminates bias arising from η_i but does not, by itself, allow consistent OLS estimation of g and b , because $\varepsilon_{i,t-1}$ is correlated with $y_{i,t-1}$.

However, under the assumption that ε_{it} does not exhibit serial correlation, levels of the dependent variable lagged by two or more periods are not correlated with $\varepsilon_{i,t-1}$. Under this maintained assumption, this implies that valid instruments for $(y_{i,t-1} - y_{i,t-2})$ include $y_{i,t-2}, y_{i,t-3}, \dots, y_{i1}$. Therefore, Instrumental Variables (IV) estimation can be undertaken

to potentially generate a consistent estimator of g and b . Arellano and Bond (1991) propose expansion of the set of instruments to include, *inter alia*, current and lagged levels (or first difference transformations) of any strictly exogenous x_{it} variables (i.e. $x_{it}, x_{i,t-1}, x_{i,t-2}$ and so on).¹⁷ By contrast, if some or all of the x_{it} variables are not strictly exogenous but are nonetheless predetermined, then only lagged levels or first differences of $(x_{i,t-1}, x_{i,t-2}, \dots)$ are valid instruments.¹⁸

Estimators formed in this way are termed "Difference Generalised Method of Moments (GMM) estimators", since they exploit the following $(T-2)(T-1)/2$ moment restrictions:

$$\begin{aligned} E[(\varepsilon_{it} - \varepsilon_{i,i-1}) y_{i,t-2}] &= 0 \\ E[(\varepsilon_{it} - \varepsilon_{i,i-1}) y_{i,t-3}] &= 0 \\ \dots\dots\dots &= 0 \\ E[(\varepsilon_{it} - \varepsilon_{i,i-1}) y_{i1}] &= 0 \end{aligned} \tag{4.16}$$

for $t = 3, \dots, T$ which can be applied to the model (4.15) in conjunction with the first difference transformation. In this case, the transformations used are $\mathbf{y}_{iDGM}^* = (\Delta y_{i3} \Delta y_{i4} \dots \Delta y_{iT})'$, $\mathbf{W}_{iDGM}^* = (\Delta \mathbf{w}'_{i3} \Delta \mathbf{w}'_{i4} \dots \Delta \mathbf{w}'_{iT})'$, $\Delta \mathbf{w}_{it} = (\Delta y_{i,t-1} \Delta x'_{it})$, where $\Delta(\cdot)_t \equiv (\cdot)_t - (\cdot)_{t-1}$. The corresponding valid instrument set is

$$\mathbf{Z}_{iDGM} = \begin{pmatrix} y_{i1} & x_{i1} & x_{i2} & 0 & 0 & 0 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & y_{i1} & y_{i2} & x_{i1} & x_{i2} & x_{i3} & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & y_{i1} & \dots & y_{i,T-2} & x_{i1} & \dots & x_{i,T-1} \end{pmatrix}$$

¹⁷Variable x_{it} is said to be strictly exogenous if it is uncorrelated with past, present and future shocks, so that $E[x_{it}\varepsilon_{is}] = 0 \quad \forall s, t$. See also footnote 10 above.

¹⁸Variable x_{it} is said to be predetermined if it is uncorrelated with present and future shocks, so that $E[x_{it}\varepsilon_{is}] = 0 \quad \forall s \geq t$. This allows for feedback from past shocks on to the current value of the explanatory variables.

if all of the x_{it} are predetermined and

$$\mathbf{Z}_{iDGMM} = \begin{pmatrix} y_{i1} & x_{i1} & x_{i2} & x_{i3} & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & y_{i1} & y_{i2} & x_{i1} & x_{i2} & x_{i3} & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & y_{i1} & \cdots & y_{i,T-2} & x_{i1} & \cdots & x_{i,T-1} \end{pmatrix}$$

if all of the x_{it} are strictly exogenous.¹⁹

The choice of the weighting matrix $\mathbf{H}$ depends upon whether the parameter vector $\boldsymbol{\mu}$ is to be estimated via one or two steps.

One-Step Estimation Under one-step GMM estimation of $\boldsymbol{\mu}$, the optimal weighting matrix is the inverse of the covariance of the orthogonality conditions (4.16), which is the following $(T - 2)$ square matrix with 2 in the main diagonal, -1 in the first subdiagonals and 0 otherwise:

$$\mathbf{H}_1^{DGMM} \equiv \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & & 0 \\ 0 & -1 & 2 & & \vdots \\ \vdots & & & & -1 \\ 0 & 0 & \cdots & -1 & 2 \end{pmatrix}$$

Then, using the general form in (4.15), the one-step estimator $\hat{\boldsymbol{\mu}}_1^{DGMM}$ is calculated as follows:

$$\begin{aligned} \mathbf{A}_{1N} &\equiv (\sum_i \mathbf{Z}'_{iDGMM} \mathbf{H}_1^{DGMM} \mathbf{Z}_{iDGMM})^{-1} \\ \mathbf{M}_1 &\equiv (\sum_i \mathbf{W}_{iDGMM}^* \mathbf{Z}_{iDGMM}) \mathbf{A}_{1N} (\sum_i \mathbf{Z}'_{iDGMM} \mathbf{W}_{iDGMM}^*) \\ \hat{\boldsymbol{\mu}}_1^{DGMM} &= \mathbf{M}_1^{-1} (\sum_i \mathbf{W}_{iDGMM}^* \mathbf{Z}_{iDGMM}) \mathbf{A}_{1N} (\sum_i \mathbf{Z}'_{iDGMM} \mathbf{y}_{iDGMM}^*) \end{aligned}$$

¹⁹Dummy explanatory variables for each year can also be accommodated: for details see Doornik *et al* (2002).

It is possible to estimate heteroskedasticity-robust standard errors for use with these estimators, using the estimation procedure due to White (1980) as detailed in Appendix C.2. As none of the nonlinear estimation procedures we have considered in Section 4.3.2 above enable heteroskedasticity-robust standard errors to be estimated, this points to a considerable advantage of using the linear model. See also the discussion in Section 4.3.1, above.

Two-Step Estimation Under two-step estimation of $\boldsymbol{\mu}$, the weighting matrix is the vector product of the estimated residuals from one-step estimation discussed above, so that $\mathbf{H}_{2i}^{DGMM} \equiv \widehat{\varepsilon}_i \widehat{\varepsilon}_i'$. We can use the same general form to specify the two-step estimator $\widehat{\boldsymbol{\mu}}_2^{DGMM}$ as follows:

$$\begin{aligned} \mathbf{A}_{2N} &\equiv \left(\sum_i \mathbf{Z}'_{iDGMM} \mathbf{H}_{2i}^{DGMM} \mathbf{Z}_{iDGMM} \right)^{-1} \\ \mathbf{M}_2 &\equiv \left(\sum_i \mathbf{W}_{iDGMM}^* \mathbf{Z}_{iDGMM} \right) \mathbf{A}_{2N} \left(\sum_i \mathbf{Z}'_{iDGMM} \mathbf{W}_{iDGMM}^* \right) \\ \widehat{\boldsymbol{\mu}}_2^{DGMM} &= \mathbf{M}_2^{-1} \left(\sum_i \mathbf{W}_{iDGMM}^* \mathbf{Z}_{iDGMM} \right) \mathbf{A}_{2N} \left(\sum_i \mathbf{Z}'_{iDGMM} \mathbf{y}_{iDGMM} \right) \end{aligned}$$

Estimation of standard errors is also discussed in Appendix C.2.

Diagnostic Tests The consistency of the estimators $\widehat{\boldsymbol{\mu}}_1^{DGMM}$ and $\widehat{\boldsymbol{\mu}}_2^{DGMM}$ for parameter estimation of $\boldsymbol{\mu}$ rests crucially on the validity of several identifying restrictions. One of these is the assumption of no serial correlation in the *original* shocks ε_{it} . This assumption is equivalent to assuming that there is no second-order serial correlation in terms of the transformed error terms $\varepsilon_{it} - \varepsilon_{i,t-1}$. To test this assumption, Arellano and Bond (1991) develop test statistics (“ m_1 ” and “ m_2 ”) for first and second order serial correlation, respectively, in the transformed residuals.

Additionally, the set of overidentifying restrictions used to estimate the parameters can be tested, using the Sargan test of overidentifying restrictions described in Hansen (1982). The statistic

$$\begin{aligned} S_1 &\equiv \left(\sum_i \widehat{\varepsilon}_i' \mathbf{Z}_{iDGMM} \right) \mathbf{A}_{1N} \left(\sum_i \mathbf{Z}'_{iDGMM} \widehat{\varepsilon}_i \right) \widehat{\sigma}_{DGMM}^{-2} \\ S_2 &\equiv \left(\sum_i \widehat{\varepsilon}_i' \mathbf{Z}_{iDGMM} \right) \mathbf{A}_{2N} \left(\sum_i \mathbf{Z}'_{iDGMM} \widehat{\varepsilon}_i \right) \end{aligned} \quad (4.17)$$

(depending on whether one-step or two-step estimation is used) follows an asymptotic chi square distribution if the null hypothesis of valid instruments is correct. The degrees of

freedom for these tests are equal to the number of overidentifying restrictions. Full details of both procedures are contained in Appendix C.2.

Additionally, to the extent that a given specification exploits additional restrictions (and so uses additional instruments) compared to a less restrictive specification, validity of the additional restrictions can be tested using the Sargan Difference test. Define the statistic

$$D = S_R - S_U \quad (4.18)$$

where S_R and S_U are the Sargan instrument validity statistics for the more restrictive and less restrictive sets of results, respectively. Under the null hypothesis that the additional restrictions imposed are valid, the Sargan Difference test statistic D is distributed according to the Chi-Square distribution $\chi^2_{(d_R - d_U)}$ where d_R and d_U are the degrees of freedom under which S_R and S_U are evaluated, respectively.

System GMM Estimators

Although the Difference GMM estimator is consistent for μ as $N \rightarrow \infty$, unfortunately it does not always have desirable properties in finite samples. Simulation results in Alonso-Borrego and Arellano(1996) have shown it to be both imprecise and severely biased, in environments where all of the available instruments are no more than weakly correlated with the explanatory variables in first differences ($\Delta y_{i,t-1}, \Delta x_{it}$) which they are used as instruments for. Where instruments are weak, Staiger and Stock (1997) have shown that estimation of the underlying parameters is imprecise; further, even small correlations between a weak instrument set and error terms will cause parameter estimation to be severely biased and inconsistent. Blundell and Bond (1998) show that these twin problems are particularly severe in two cases: first, where the dynamic process is close to non-stationary in the sense that δ is close to unity; and second, where the relative variance of the individual effects $\frac{\text{var}(\eta_i)}{\text{var}(\varepsilon_{it})}$ is very high.

To improve upon the performance of the Difference GMM estimator in practice, it is necessary to specify additional restrictions upon the underlying dynamic process. Blundell and Bond (1998) propose the following $(T - 2)$ additional moment restrictions based upon

the underlying assumption that the initial levels $y_{i,1}$ are mean-stationary:²⁰

$$\begin{aligned}
 E[(y_{i,T-1} - y_{i,T-2})(\eta_i + \varepsilon_{i,T})] &= 0 \\
 E[(y_{i,T-2} - y_{i,T-3})(\eta_i + \varepsilon_{i,T-1})] &= 0 \\
 &\dots\dots\dots = 0 \\
 E[(y_{i,2} - y_{i,1})(\eta_i + \varepsilon_{i,3})] &= 0
 \end{aligned} \tag{4.19}$$

in addition to those restrictions specified for the Difference GMM estimator. Under these extra assumptions, $\Delta y_{i,T-1} \equiv (y_{i,T-1} - y_{i,T-2})$ is a valid instrument for the level of $y_{i,T-1}$; $\Delta y_{i,T-2}$ a valid instrument for the level of $y_{i,T-2}$ and so on. Estimators formed in this way are termed "System GMM estimators", since they exploit restrictions on both levels *and* first differences of the error term $\eta_i + \varepsilon_{i,t}$. The valid instrument set $\mathbf{Z}_{iSGMM}$ for this estimator is therefore:

$$\mathbf{Z}_{iSGMM} = \begin{pmatrix} \mathbf{Z}_{iDGMM} & 0 & \cdots & 0 \\ 0 & \Delta y_{i2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \Delta y_{i,T-1} \end{pmatrix}$$

Estimation proceeds in analogous manner to the Difference GMM estimator, and both one-step and two-step estimators can be calculated similarly.²¹ Details, including those concerning estimation of standard errors and diagnostic tests, are contained in Appendix C.2.

²⁰Specifically it requires that $E\left[\left(y_{i,1} - \left(\frac{\eta_i}{1-\alpha}\right)\right)\eta_i\right] = 0$ for $i = 1, \dots, N$. See Bond (2002, p.153).

²¹Further note that, if any of the explanatory variables are uncorrelated with η_i , then lagged levels of such variables can be used as instruments for the y_{it} in levels to improve accuracy of parameter estimation.

4.3.4 Summary of Models to be Estimated

For the nonlinear models we estimate parameters for the employment state equation in (4.1).

Six different Logit specifications are used (as described in Section 4.3.2 above):

- Honore and Kyriazidou's Fixed Effects specification
- Uncorrelated Random Effects specification²² with Exogenous Initial Conditions
- Uncorrelated Random Effects specification with Equilibrium Initial Conditions
- Uncorrelated Random Effects specification with Reduced Form Initial Conditions
- Correlated Random Effects specification with Restricted Correlation and Reduced Form Initial Conditions
- Correlated Random Effects specification with Unrestricted Correlation and Reduced Form Initial Conditions

For the linear models we estimate parameters for the employment state equation in (4.14).

Two main LPM specifications are used, corresponding to moment restrictions (as described in Section 4.3.3 above):

- Difference GMM moment restrictions only
- Difference GMM moment restrictions and additional System GMM moment restrictions

Within each of these two main specifications, several alternative sub-specifications are considered, corresponding to additional independence restrictions between either the levels or time differences of x'_{it} , and the composite errors $\eta_i + \varepsilon_{is}$ (where $s = 1, \dots, T$).

²²Uncorrelated in the sense that it is assumed all observed explanatory variables are orthogonal to temporally-persistent unobserved differences between individuals: $E[x_{it}\eta_i] = 0 \quad \forall i, t$

4.4 Sample Selection Methodology

Since the objective of this Chapter is to examine persistence in employment status, and particularly to distinguish true from spurious state dependence (see Section 4.2, above), it is necessary to use repeated observations of employment status for individuals participating in the labour market. It is preferable to also use information on the observed characteristics of those individuals, to construct a rich set of covariates which control for as much heterogeneity as possible when estimating state dependence effects.

The British Household Panel Survey (BHPS) is the source of individual level data used. The BHPS has been conducted annually since 1991 in England and Scotland south of the Caledonian canal, on a randomly selected sample of households which are repeatedly observed. It is a high quality panel survey, both in terms of the breadth and detail of questions asked in the survey, and in terms of the attention given to resampling all adult respondents in each household year after year, which in some cases involves detailed work finding respondents who have changed address. We recall from Chapter 2 that our modelling strategy for the Australian youth labour market was influenced by concerns about endogenous attrition within the Australian Longitudinal Survey, so lack of attrition within the BHPS is an important indicator of data quality for our purposes. For a general description of the BHPS, see Taylor (1994).

As individual employment effects are potentially long-lived, it is desirable to use a long panel (i.e. a large number of observations per respondent) in order to cover a large portion of each respondent's working life. The first ten waves of the BHPS (i.e. years 1991-2000) are used, which is a relatively long panel by comparison with other studies which have examined the issue of labour market persistence. For example, Arulampalam (2004) uses the first seven waves of the BHPS to analyse the UK labour market, and Chapter 2 of this work uses four waves of the Australian Longitudinal Survey. The 1991-2000 period also captures diverse macroeconomic conditions over a business cycle, including the UK recession at the beginning of the 1990s and the subsequent period of high economic growth towards the end of the 1990s.

Although information about individuals, including employment information, is continuous, we approximate individual information over the period 1991-2000 by using ten annual observations per individual. The BHPS contains additional data on work patterns between annual sampling points (i.e. intra-year) where respondents are asked to complete work histories which are up to one year old at time of sampling. These work histories allow

construction of more than one sampling point per year. However, Paull (1996, 2002) notes the work histories are potentially subject to inaccuracies caused by recall bias, so the intra-year work data are not used and we use the more reliably measured one sampling point per year to construct the data set used in this Chapter, which is also broadly consistent with the definition of employment status used in Chapter 2 above.

4.4.1 Variable Selection

Employment

A crucial issue in our empirical methodology concerns how we define the variable which measures employment status. In theory, there is nothing to prevent this variable taking on a range of values to distinguish between many different “employment states”: wage-earning; salary-earning; self-employed; unemployed but searching; unemployed but not searching; and retired, being some obvious examples. However, in practice there is a trade-off between the statistical degrees of freedom lost from estimating a multi-state model and the benefits of more accurate characterisation. A multi-state model can also, for some functional forms including the Logit model, lead to “grossly false predictions of the outcomes” as pointed out by Hsiao (2003, p.192) if we assume erroneously that irrelevant alternatives are independent of each other. As a result, we follow standard practice among labour market studies of employment outcomes by using a binary variable (rather than a multi-state variable) for employment state.

The first state is “employment”, where we include self-employed workers as well as wage and salary earners to avoid attrition from workers who move between self-employment and earning wages or salaries. The question then is: what should the complementary state be? If “non-employment” is defined too narrowly, then many observations which do not fit either state will have to be discarded, and we risk possible endogenous selection bias infecting our results if the assignment of workers to the states we omit is not random.

For example, consider a situation where most workers who lose employment are sufficiently discouraged that they do not search for replacement jobs and thus never regain employment. If we define “non-employment” only as unemployed workers searching for jobs, then - by our definition - such workers will have missing values for employment state after losing employment. Therefore, we will seldom see anyone move permanently from employment to non-employment, and so may conclude erroneously that the extent of

persistence is small.

On the other hand, defining “non-employment” to include all those not employed has drawbacks too. The reason we use information to select sub-groups of individuals for consideration is, theoretically, because we have prior reason to believe they are engaged in a labour market process that others are not. Thus, we risk conflating disparate groups if we ignore differences between those not currently employed. For example, including long-standing retirees among the “non-employed” would - by such a definition - provide evidence for persistence. Yet, this would be spurious evidence as, theoretically, the unemployment scarring phenomenon which we have in mind has nothing to do with the decisions of those long retired. It can therefore be seen that deciding upon a wider or narrower definition of “non-employment” involves an inevitable trade-off between the potential selection bias associated with very narrow definitions, and the lack of precision which is likely to result from very wide ones.

In this study, we wish to capture the idea that unemployment may be the result of informational constraints upon firms, but is in no real sense a “voluntary” decision by workers. Therefore, we measure “employment” to include those who identify themselves as *currently* self-employed, in paid employment or on maternity leave, while “non-employment” includes only those who identify themselves as being *currently* unemployed or on a government training scheme. Thus, employment status is measured by self-classification.²³ Retirees, those looking after family or home, full time students and the long term sick or disabled are all omitted from consideration.

Other Variables

Explanatory variables measuring age in 1991, years of education, marital status, number of children, general health, housing ownership, domicile region of residence and time of observation are included.

It is important to control for age in order to capture skill formation and domestic responsibilities among individuals, each of which may have a bearing on employment propensity. The relationship between age and the probability of being employed is likely to be nonlinear, with a maximum employment probability at some intermediate age. Theoretical explanations incorporate human capital accumulation and learning about optimal

²³Variables *wJBSTAT* are used. The directions to each respondent were “Please look at this card and tell me which best describes your current situation” (the card specifies up to 10 options, including ‘Unemployed’). See Appendix B.1 for further details.

skills to employment matching, which may explain lower employment rates among younger workers. By contrast, greater domestic responsibilities and an increased opportunity cost of unemployment at higher wages may explain higher employment rates among middle-aged workers, while inflexibility of skill formation and demand-side sorting of employees following learning by employers may explain lower employment rates among older workers. Therefore, since age effects upon employment may plausibly be nonlinear, both linear and quadratic age variables are included.

Effects of health and years of education are consistent with theories of heterogeneous productivities across workers, which may explain supply-side effects upon employment propensity through healthier/more educated individuals having higher opportunity costs of unemployment; alternatively, it is also consistent with demand-side employer sorting effects according to health and/or education.

Effects due to housing ownership, marital status and number of children may be consistent with supply-side effects to optimise income by substitution between leisure and consumption. For example, one's spouse may provide a source of extra income and reduce employment propensity, or the financial responsibility associated with house ownership may increase employment propensity. Similarly, having more children could lead to a substitution away from leisure to provide for the children, or alternatively could lead to a substitution towards leisure in the event that social support dependent on having children provides an additional source of income. Further, in the circumstance that income from social support is greater than income available from wages, this may result in unemployment being financially beneficial compared to employment, in the presence of a large number of children.²⁴

Regional differences in availability and different types of employment may cause employment probabilities to vary, and to this extent it is desirable to control for region of residence. Regional differences are measured using a binary indicator variable, which diagnostic analysis (unreported) shows is a robust measure of region in the sense that results do not materially change if a wider set of regional dummy variables is used. Finally, since it is likely that macroeconomic conditions will affect the probability of employment, time effects are controlled for. As noted in Section 4.3.2 above, the estimator Θ_{HK} which we use to estimate parameters for the HK Fixed Effects model (4.7) is not capable of estimating coefficients on time dummies as explanatory variables; therefore, in lieu of using

²⁴Note that, using our definition of employment status, it is possible for a respondent to classify themselves as "unemployed", without having to satisfy the International Labour Organisation definition of 'unemployed' meaning available for and seeking employment.

time dummies, a quadratic time trend in years is included. See Appendix C.1 for more details of the construction of these variables.

4.4.2 Shape of Data

The BHPS contains data from approximately 11 thousand households, which are sampled repeatedly at a yearly interval. Observations on individuals aged 16 years and above within these households (i.e. all individuals, not just heads of households) are also recorded in the BHPS.

Since we wish to examine individuals who are continually in the labour market, a 10 year balanced panel structure is adopted. That is, any individual for which responses were not recorded in all 10 waves for all variables described in Section 4.4.1 above is completely omitted from the main sample considered. In particular, we do not seek to model the individual decision of whether to participate in the labour market. Since labour force participation rates are typically lower for women than for men, we choose to only consider male respondents on the grounds that any selection biases from not explicitly modelling participation are likely to be less severe for men.²⁵ The selection criteria for the main sample used is that respondents must be male, and have recorded values for employment status and the 8 explanatory variables described in Section 4.4.1 above, over all 10 BHPS waves during the period 1991-2000. Using this selection rule, we construct a balanced longitudinal sample of 1285 men.

Using a balanced panel carries some potential costs in terms of endogenous attrition. In particular, we are mindful of the possibility that the likelihood of continuous observation may depend on individual heterogeneity, both observed and unobserved. Under such circumstances, our sample selection methodology raises the possibility of bias and inconsistency in any parameters we estimate. For example, if individuals who leave a survey are also likely to have unobserved characteristics which lead to unemployment, then *not* taking this into account will result in our using a sample upon which even the most robust estimation methods will systematically overestimate the true extent of state dependence. We cannot escape this problem entirely even by using an unbalanced panel; for, even then, endogenous selection bias still arises in terms of the attrition of missing values. However, by excluding all individuals with any missing values this effectively creates even more “missing

²⁵Diagnostic analysis (unreported) on the BHPS showed that repeated participation rates of men in the labour force over 1991-2000 were indeed substantially higher than those of women.

values”; thus, to the extent that endogenous selection bias already exists, our methodology exacerbates it. For a further discussion on this issue, see Verbeek and Nijman (1996).

Since we are interested in examining transitions between states from wave to wave, it is clearly not possible - for our purposes - to use any observation on an individual from a wave where that individual does not have an observation in the wave immediately prior. Nonetheless, it would still be possible to maintain the unbalanced panel structure while simply omitting non-consecutive observations (e.g. if an individual responded in the first 6 waves but not the last 4, then we could simply consider the first 6 observations for them).

Empirical evidence reported by Arulampalam (1998, p.15) on the BHPS from 1991-1995 suggests that the problem of endogenous attrition is not so large as to seriously threaten the credibility of results obtained using the balanced panel structure. In other words, we do not think the determinants of whether an individual is continuously observed are strongly correlated with that individual’s observed and unobserved characteristics, and so the restriction of using a balanced panel is likely innocuous. However, we also conduct sensitivity analysis in Section 4.6 below to assess how robust our results are to the choice of a 10 year balanced panel structure.

4.4.3 Sample Characteristics

Table 4.1 presents the distribution of employment outcomes, as well as descriptive statistics (with some statistics scaled for estimation purposes) over demographic variables of interest. In addition to the full sample, 5 sub-samples have been created to try to capture transitional regularities among respondents. Columns (2) and (3) are composed of those who - respectively - always are employed, or never are. Column (4) is composed of those who make one transition from employment; that is, who are employed in 1991, and then during some year become unemployed without *subsequently* then becoming re-employed. Analogously, column (5) is composed of those who make one transition to employment; that is, who are unemployed in 1991, and then during some year become employed, then *subsequently* remaining employed until 2000. Column (6) is composed of all other respondents; that is, those who make more than one transition between employment states over the 10 waves.

Several features are apparent from Table 4.1. In the aggregate, employment is typical, and only 27 respondents (2% of the sample of 1285 respondents) are unemployed for more than 4 of the 10 waves. Indeed, 1049 respondents (81%) in our sample are employed at

Table 4.1: Sample Characteristics for BHPS 1991-2000

	Full Sample (1)	Employed 10 years (2)	Employed 0 years (3)	Single Transition from Work (4)	Single Transition to Work (5)	Multiple Transitions (6)
Means and Standard Deviations						
Employment (1=Employed)	0.96 (0.20)	1.00 (0.00)	0.00 (0.00)	0.72 (0.45)	0.81 (0.40)	0.79 (0.41)
Age/100 (1991)	0.36 (0.10)	0.36 (0.10)	0.32 (0.13)	0.41 (0.08)	0.31 (0.10)	0.34 (0.11)
Marital (1=Cohabiting)	0.79 (0.41)	0.81 (0.39)	0.74 (0.44)	0.65 (0.48)	0.77 (0.42)	0.69 (0.46)
Children/10	0.09 (0.11)	0.09 (0.11)	0.19 (0.12)	0.11 (0.16)	0.11 (0.12)	0.08 (0.11)
Years of Education/100	0.12 (0.03)	0.12 (0.02)	0.10 (0.01)	0.11 (0.02)	0.11 (0.02)	0.12 (0.02)
Health (1 = Good Health)	0.97 (0.16)	0.98 (0.16)	0.96 (0.20)	0.99 (0.10)	0.98 (0.12)	0.95 (0.21)
Housing (1 = Own House)	0.83 (0.38)	0.87 (0.34)	0.06 (0.24)	0.71 (0.46)	0.62 (0.49)	0.70 (0.46)
Region (1 = South England)	0.45 (0.50)	0.45 (0.50)	0.20 (0.40)	0.55 (0.50)	0.47 (0.50)	0.45 (0.50)
Number of years worked (column percentages)						
zero	0.4	—	100	—	—	—
one	0.2	—	—	9.1	1.7	—
two	0.2	—	—	9.1	—	1.2
three	0.2	—	—	—	1.7	1.2
four	0.5	—	—	9.1	3.5	2.5
five	0.6	—	—	—	—	4.9
six	1.2	—	—	—	5.2	7.4
seven	1.8	—	—	—	8.6	11.1
eight	3.7	—	—	—	20.7	21.6
nine	9.6	—	—	72.7	58.6	50.0
ten	81.6	100	—	—	—	—
Sample Size	1285	1049	5	11	58	162

Notes:

Standard deviations are in parentheses.

Variable definitions are contained in Appendix C.1.

all 10 waves; our definition of unemployment obviously, for the most part, omits consideration of those who voluntarily absent themselves from the labour force.²⁶ The respondents who are employed at all 10 waves have relatively high house ownership rates (87% cf. 83% in the full sample), and relatively high cohabitation rates (81% cf. 79% in the full sample).

Multiple transitions within the labour force are made by 162 respondents (13%). These respondents have relatively low house ownership rates (70% cf. 83% in the full sample), report relatively poor health (95% report good health cf. 97% in the full sample) and are relatively young (average age 34 years cf. 36 years in the full sample).

A single transition to work is made by 58 respondents (5%). These respondents have relatively low house ownership rates (62% cf. 83% in the full sample), relatively low cohabitation rates (77% cf. 79% in the full sample) and are relatively young (average age 31 years cf. 36 years in the full sample). There are 5 respondents who are unemployed at all 10 waves, and these men on average have 2 less years of education (10 years cf. 12 years) than the rest of the sample.

It is noteworthy that respondents in columns (4), (5) and (6) of Table 4.1, who have in common that they each make at least one transition between employment states, have employment rates which are much lower than the 96% employment rate for the aggregate sample. This suggests that compositional differences within the labour force play a substantial role in determining employment status, which is a key motivation for our analysis. Indeed, respondents who experience transition within the labour force exhibit striking compositional differences from the rest of the sample in several respects. Respondents who make one or more transitions, on average, have fewer years of education, much lower house ownership rates and are less healthy, compared with those in employment at all 10 waves. From the lower part of Table 4.1, we can also see that 50% of the 162 respondents who make multiple transitions nonetheless are in employment at 9 of the 10 waves; in other words, they are employed in 1991, and make exactly one transition to unemployment and one transition back to employment. In view of the strong compositional differences notwithstanding, it seems that experiencing even one wave of unemployment is associated with also having some characteristics (e.g. health, house ownership) different from the typical characteristics of the majority (81%) of the men in our sample who are employed at all 10 waves.

Almost all respondents have been formally educated for at least 10 years, which is

²⁶It is important to note that neither these 1049 men, nor the 5 men who were unemployed during every wave, provide any evidence about whether the cause of persistence in employment status is state dependence cf. unobserved heterogeneity. Therefore, it is the remaining 231 men whose employment histories are crucial for the purposes of this study.

the legal requirement in the UK. Over two thirds of respondents are married or cohabiting (measured by a binary variable which takes the value 1 if the respondent is in cohabitation and zero otherwise), and a wide cross-section of working ages is represented.

Overall, there are significant compositional differences within our sample. Those men within the sample who are employed during a large number of waves are relatively old, and likely to be cohabiting and to own their house. This is potentially important for estimating employment persistence, to the extent that such characteristics - and perhaps also unobserved characteristics - are linked with employment outcomes.

Table 4.2: Probability of Labour Force Transitions

Wave	All	1	2	3	4	5	6	7	8	9	10
Unemp(macro) (%)	9.3	9.3	11.6	12.5	11.5	10.2	9.7	8.2	6.9	6.8	6.2
Unemp(BHPS) (%)	3.9	7.0	6.7	5.4	5.3	4.3	4.2	2.8	2.4	2.2	2.2
Pr [emp (t) emp (t-1)]	0.98	—	0.97	0.98	0.97	0.98	0.98	0.99	0.99	0.99	0.99
Pr [un (t) emp (t-1)]	0.02	—	0.03	0.02	0.03	0.02	0.02	0.01	0.01	0.01	0.01
Pr [emp (t) un (t-1)]	0.50	—	0.52	0.53	0.49	0.54	0.39	0.56	0.47	0.45	0.52
Pr [un (t) un (t-1)]	0.50	—	0.48	0.47	0.51	0.46	0.61	0.44	0.53	0.55	0.48

Notes:

Wave numbers refer to BHPS waves, between 1991 (Wave 1) and 2000 (Wave 10).

Source of macroeconomic unemployment data is the Labour Force Study.

Table 4.2 provides more detail on observed employment/unemployment patterns, grouped by survey and wave. The first part of Table 4.2 compares historical unemployment incidence over the sample period 1991-2000. The first line shows macroeconomic unemployment figures among UK men, ranging from 6.2% in 2000 to 12.5% in 1993. The second line shows the corresponding unemployment figures from our BHPS sample, which are considerably lower than those for the male population as a whole, and display somewhat different time series behaviour compared with the male population as a whole over the ten year period. This is because our BHPS sample focusses on those men continually in the labour force. For both our sample and for the male population as a whole, unemployment rates fall during the second half of the sample period.

The other parts of Table 4.2 set out transition proportions within and between employment states, in Markov matrix format. Conditional on being employed in the previous

year, the vast majority of respondents (98% for the full sample) again are employed in the current year. However, for those respondents unemployed in the previous year, only 39% to 56% (50% for the full sample) of these are able to regain employment a year later. Thus, we see strong regularities of unemployment persistence. However, inspection of Table 4.2 tells us nothing about whether it is state dependence or unobserved heterogeneity which is generating these regularities.

Among other labour market studies, Arulampalam *et al* (2000) use the first 5 waves of the BHPS to examine employment patterns, excluding self-employment, for men only. Their panel is unbalanced in the sense that some respondents are not surveyed during all waves, so it has approximately double the number of respondents compared with ours²⁷. They estimate slightly higher overall male unemployment levels than we do in 1991 to 1994, and a slightly lower unemployment level in 1995. They estimate *employment* transition proportions which are very similar to ours; that is, 97% to 98% of employed men at any wave remain employed at the next wave. However, their estimated *unemployment* transition proportions point to some differences in measured persistence compared to ours: 49% to 53% of unemployed men at any wave remain unemployed at the next wave for their sample (1991-1995), compared with 46% to 51% for ours (first 5 waves:1991-1995) and 46% to 61% for ours (all 10 waves:1991-2000). Nonetheless, these differences are sufficiently small, especially over the comparable period 1991-1995, to reassure us that our sample is indeed representative of that which could have been obtained using other than a balanced panel.

²⁷See Table 1 of Arulampalam *et al* (2000, p.28).

4.5 Estimation Results

The primary empirical focus of this Chapter is to assess evidence from the UK labour market on the existence of state dependence and individual-specific heterogeneity. In particular, the central hypothesis of interest is whether non-zero state dependence effects exist, after controlling for observed and unobserved heterogeneity.

As identified in Section 4.3 above, many estimators are likely to be inadequate for these purposes. Among the dynamic specifications considered, we have seen that there are many plausible sources of correlation between the error term and either the observed characteristics or lagged employment status variables; under such circumstances, estimators of state dependence will be inconsistent.

The results presented in detail below can be summarized as follows: all estimation methods provide support for the hypothesis that non-zero state dependence effects exist. In other words, even after controlling for observed and unobserved differences between individuals, observing that an individual is employed (unemployed) this year predicts a significantly higher probability that the same individual will be employed (unemployed) one year later. This is a result which appears to be beyond reasonable doubt, and is robust to all the specifications considered.

However, it is not so obvious to discern the magnitude of the state dependence effect. Although it is clear that present employment predicts a higher probability of future employment probability, it is less clear how to measure the probabilistic advantage - in terms of future employment - conferred by present employment. Much of the discussion which follows focusses on how to measure this reliably.

There is also considerable evidence that differences between individuals, both observed and unobserved, are powerful predictors of differences in employment outcomes. The evidence suggests strongly that some relevant differences between individuals are difficult or impossible to control for using observed explanatory variables such as those included here. Therefore, it is important to model unobserved heterogeneity, for otherwise there is a risk that it may be misleadingly conflated with the state dependence effect.

4.5.1 Results from Nonlinear Specifications

As described in Section 4.3.4 above, six Logit specifications have been used to model employment status according to the specification in Equation (5). Five of these adopt a random effects framework, and all of these five assume that permanent unobserved differences between individuals η_i are independent of transient unobserved differences between individuals ε_{it} . Therefore they are each, in this respect, more restrictive than the final method considered, that of Honore and Kyriazidou (2000) which uses fixed effects, since this last method makes no corresponding assumption and so allows for unrestricted correlation between the η_i and x_{it} terms.

Discussion of Results from all Nonlinear Specifications

Table 3 compares results from the specification which uses fixed effects:

- Honore and Kyriazidou's Fixed Effects specification in (4.7) (FE-HK) [column vi]

with results from each of five methods which use random effects:

- Uncorrelated Random Effects specification with Exogenous Initial Conditions in (4.8) (RE-EX) [column i]
- Uncorrelated Random Effects specification with Equilibrium Initial Conditions in (4.9) (RE-EQ) [column ii]
- Uncorrelated Random Effects specification with Reduced Form Initial Conditions in (4.10) (RE-RF) [column iii]
- Correlated Random Effects specification with Restricted Correlation and Reduced Form Initial Conditions in (4.11) (CRE-R) [column iv]
- Correlated Random Effects specification with Unrestricted Correlation and Reduced Form Initial Conditions in (4.12) (CRE-U) [column v]

The last three specifications use age, age squared, marital status, number of children, years of education, general health, housing ownership and domicile region of residence in the equations for initial conditions. For further details of the estimation methods and variables used, see Section 4.3.2 above and Appendix C.2.

Table 4.3: Results from Dynamic Logit Models (Main sample: N=1285, T=1991-2000)

	(RE-EX) (i)	(RE-EQ) (ii)	(RE-RF) (iii)	(CRE-R) (iv)	(CRE-UN) (v)	(FE-HK) (vi)
y_{t-1}	2.781 (.000)	2.413 (.000)	2.361 (.000)	2.322 (.000)	2.318 (.000)	2.079 (.000)
Constant	-3.729 (.000)	-1.920 (.038)	-3.315 (.002)	-3.207 (.044)	-0.598 (.742)	—
Age	5.220 (.308)	3.140 (.486)	5.255 (.358)	5.245 (.362)	-7.523 (.312)	—
Age ²	-0.578 (.416)	-0.463 (.464)	-0.564 (.480)	-0.671 (.400)	0.799 (.430)	—
Marital	0.331 (.048)	0.323 (.010)	0.358 (.042)	-0.230 (.392)	-0.217 (.540)	-0.135 (.695)
Children	-0.789 (.278)	-0.684 (.302)	-0.736 (.342)	-0.052 (.966)	0.774 (.614)	3.152 (.070)
Education	7.970 (.024)	4.048 (.200)	7.993 (.044)	8.098 (.658)	19.74 (.584)	21.21 (.288)
Health	1.201 (.000)	1.168 (.000)	1.310 (.000)	1.242 (.000)	1.347 (.000)	1.311 (.011)
Housing	1.240 (.000)	0.997 (.000)	1.255 (.000)	0.857 (.008)	0.952 (.034)	1.186 (.000)
Region	0.130 (.422)	0.173 (.220)	0.163 (.034)	1.773 (.006)	1.680 (.114)	2.360 (.974)
Time	2.790 (.036)	1.960 (.000)	2.792 (.034)	2.069 (.132)	2.557 (.106)	2.524 (.221)
Time ²	-0.909 (.420)	-0.281 (.644)	-0.822 (.466)	-0.090 (.936)	-0.469 (.724)	-0.508 (.803)
VAR(η_i)	1.479 (.000)	1.597 (.000)	2.187 (.000)	2.333 (.000)	1.738 (.002)	—
γ_1	—	—	0.692 (.000)	0.684 (.000)	0.791 (.001)	—
Log(L) x N	-1244.1	-1555.5	-1514.3	-1503.5	-1468.6	—
HausmanBase	RE-RF	RE-RF	CRE-R	CRE-UN	N/A	RE-RF
HausmanStat	20.40	1.399	0.588	0.000	N/A	5.96
Hausman-p	0.000	0.236	0.443	0.983	N/A	0.014

Notes:

p-values based on the Normal distribution are in parentheses.

HausmanBase is the less restrictive model used for a Hausman test on the coefficient of y_{t-1} against the model in each respective column, except for the final column where HausmanBase is the more restrictive model.

HausmanStat is the value of Hausman's test statistic and Hausman-p is the associated p-value (1 degree of freedom).

A total of 178 men are used by the FE-HK estimator to identify parameters.

The state dependence effect is directly measured by the coefficient on the lagged (employment status) dependent variable y_{t-1} . Note that since the employment status is defined as equal to one when an individual is employed, a positive coefficient for a given variable x means that the probability of employment increases (i.e. that the probability of unemployment decreases) for an increase in the value of x . Results from all six specifications identify a state dependence effect which is significantly different from zero.

Among the observed characteristics included, effects of age in 1991 upon employment probabilities are insignificant although there is some weak evidence of a quadratic relationship between employment status and age (i.e. men of middle age more employable than either younger or older men). However, this evidence is quite weak. It should be noted that these results say nothing about whether there are age effects upon the likelihood of continuing employment in a particular job, but instead are pertinent to the likelihood of continuing to be employed in a job generally. To this extent, there is no clear profile of age against employment status.²⁸

The orthogonality restriction used by RE-RF that all observed explanatory variables x_{it} are uncorrelated with the unobserved characteristics η_i , if valid (see below), allows us to identify significant effects from marital status upon employment status. This effect is insignificant when CRE-R is used for estimation, but using this orthogonality restriction allows us to estimate a significant positive coefficient for cohabitation (i.e. "Marital" = 1) upon being employed. To the extent that cohabitation involves greater financial responsibility, this is consistent with a stronger preference for being employed [e.g. Schoeni (1995)], given cohabitation.

There is little evidence that number of children has any effect on employment status, which is shown by no model being able to identify a non-zero coefficient (for number of children upon employment status) at the 5% significance level. By contrast, the orthogonality restriction used by RE-RF, if valid, allows us to identify a significant positive coefficient for extra years of education upon being employed. However, the significant coefficients on marital status and education disappear when we allow for any form of correlation between explanatory variables and individual effects. This is consistent with two competing interpretations. First, human capital formation through education may lead to a higher

²⁸As discussed in Section 4.3.2 above, since FE-HK exploits only first differences of explanatory variables, coefficients on time-invariant variables (including any constant term) cannot be identified separately. By contrast, RE-RF is capable of identifying these coefficients; also, since it is estimated using maximum likelihood rather than kernel weighting, both the estimated variance of η_i (which measures the extent of unobserved heterogeneity) and the maximised log-likelihood value (scaled by $N = 1285$) can be reported.

opportunity cost of being unemployed, and so increases the likelihood that an individual with more years of education will be employed; paraphrasing, it is consistent with structural causation from education to employment probability. Second, it is consistent with correlation between marital status and unobserved heterogeneity, correlation between education and unobserved heterogeneity, or both. Under this type of explanation, more employable individuals (i.e. men with high values of η_i) are also more likely to be married, and tend to have more years of education, but there need be no structural causation from marital status or education to employment probability.

There is clear evidence of a positive coefficient from being in good health (i.e. "Health" = 1) upon being employed. These results are strong, both in the sense of generating very significant coefficient estimates (i.e. low p-values) for any given estimation method, and in the sense of being robust to allowing for correlation between explanatory variables and individual effects. This is consistent with healthy individuals searching for employment with greater intensity, and/or employers attributing a signal of lower productivity to - and therefore being less willing to hire - individuals in poor health.

There is also clear evidence of a positive coefficient from owning one's house (i.e. "Housing" = 1) upon being employed. This is consistent with home ownership imposing large adjustment costs upon moving to a different area, which reduces the net expected benefits to an individual of becoming unemployed to search for a new job. However, this result is also consistent with home ownership being an endogenous manifestation of an individual's present and future expectations of being employed [e.g. Booth *et al* (1999)]. Therefore, the direction of causality between home ownership and current employment status is potentially ambiguous.

Both RE-RF and CRE-R estimate a significant positive coefficient of residing in the south of England (i.e. "Region" = 1) upon being employed. Additional robustness analysis (unreported) was carried out on several more granular specifications of region (e.g. dividing into all 20 reported regions, looking at Scotland separately to the North of England, etc.) none of which led to estimates of the main parameter of interest δ changing significantly.

A positive coefficient on the included time trend is identified at a 5% significance level using RE-RF's orthogonality restriction. Unemployment levels later in the 1990s were, on average, lower than at the start of the decade, which reflects macroeconomic fluctuations (see also Table 4.2).

The variance of individual effects η_i is estimated as significantly different from zero ($p=0.000$) for all specifications²⁹. Since this coefficient measures unobserved heterogeneity between individuals, this is therefore evidence that true state dependence is not wholly responsible for the transitions shown in Table 4.2. In other words, these results show that both true state dependence and spurious state dependence exist, and controlling for unobserved heterogeneity is necessary to consistently estimate the effects of true state dependence.

Overall, the primary finding from Table 4.3 is that a clearly positive (i.e. greater than zero) structural state dependence effect in employment status exists, even after initial conditions and individual differences – both observed and unobserved – have been controlled for. This is the key finding, and is consistent with earlier studies and with the results in Chapter 2.

Among the secondary findings, it is clear that good health, owning one's place of residence, and living in the south of England are each positively correlated with being employed. Evidence for effects of other observed characteristics is not so robust to different specifications, although there is some evidence that cohabitation and/or more education are positively correlated with being employed.

Selection of Preferred Nonlinear Specifications

We now consider, in the context of the main sample of 1285 men, which of the six specifications presented in Table 4.3 are most likely to lead to consistent parameter estimation. As described in Section 4.3.2 above, Hausman's (1978) test for model misspecification takes the form shown in (4.13):

$$H \equiv (\widehat{\varphi}_R - \widehat{\varphi}_U)' \widehat{V}^{-1} (\widehat{\varphi}_R - \widehat{\varphi}_U)$$

where $\widehat{\varphi}_R$ and $\widehat{\varphi}_U$ are the restricted and unrestricted parameter estimators respectively, and $\widehat{V} = \widehat{var}(\widehat{\varphi}_U) - \widehat{var}(\widehat{\varphi}_R)$. Under the null hypothesis that the restricted model is valid, H follows an asymptotic chi-square distribution with degrees of freedom equal to the number of columns in each parameter vector. In this section, we discuss the results from Hausman tests on the coefficient δ on the lagged employment variable y_{t-1} which measures the extent

²⁹Note that significance of the variance of η_i is measured using a non-standard likelihood ratio test statistic. See Arulampalam (2004) or Lawless (1987).

of estimated persistence in employment status. Therefore, each Hausman test has degrees of freedom equal to 1, to allow us to focus upon this main parameter of interest. The Hausman test statistics are reported in Table 4.3.

The Hausman tests provide evidence against Random Effects where initial conditions are assumed to be exogenous, when compared to Random Effects with reduced form initial conditions (HausmanStat in Column i = 20.40, $p = 0.000$). The coefficient on y_{t-1} is significantly different when estimated with exogenous initial conditions (Column i), compared to when estimated with reduced form initial conditions (Column iii). This is interpreted as evidence against exogenous initial conditions, since the restrictions imposed by reduced form initial conditions are a strict subset of the restrictions imposed by exogenous initial conditions (i.e. there are fewer restrictions placed upon the reduced form initial conditions model so it is the more general model). We therefore reject the null hypothesis that y_{i1} the initial period's employment outcome is orthogonal to η_i the individual effect.
30

By comparison, there is less evidence against the restriction that initial conditions are in dynamic equilibrium. Comparing the coefficient on y_{t-1} when estimated with equilibrium initial conditions (Column ii), with that estimated under reduced form initial conditions (Column iii), the Hausman test is unable to detect any difference at conventional significance levels (HausmanStat in Column ii = 1.399, $p = 0.236$). Again, the restrictions imposed by reduced form initial conditions are a strict subset of the restrictions imposed by equilibrium initial conditions, so evidence against the assumption of equilibrium initial conditions is not strong based on a Hausman test of the coefficient on y_{t-1} . There is not sufficient evidence here to reject the null hypothesis that the process is in dynamic equilibrium in the initial period.

The restriction of orthogonality between observed characteristics and unobserved individual effects is not rejected using a Hausman test of the coefficient on y_{t-1} , comparing the RE-RF specification against the CRE-R specification (which also specifies reduced form initial conditions). The Hausman test in Column iii finds no significant difference ($p = 0.443$) between the coefficient on y_{t-1} when estimated with standard Random Effects under reduced form initial conditions (Column iii) [which imposes the orthogonality restriction], and (restricted) Correlated Random Effects under reduced form initial conditions (Column

³⁰As further evidence against the validity of exogenous initial conditions, the null hypothesis that $\gamma_1 = 0$ (see discussion of the Reduced Form Initial Conditions model in Section 4.3.2, above) is rejected using an unreported t-test ($p = 0.000$)

iv) [which relaxes the orthogonality condition to the extent that unobserved individual effects are specified to have a linear relationship with the means across waves of each observed characteristic]. There is not sufficient evidence here to reject the null hypothesis that observed characteristics x_{it} for $t \geq 2$ are orthogonal to η_i the individual effect.

Similarly, relaxing the orthogonality restriction even further to allow for a relationship between unobserved individual effects and each period's individual observed characteristics does not lead to a significantly different value of the estimated coefficient on y_{t-1} . The Hausman test in Column iv finds no significant difference ($p = 0.983$) between the coefficient on y_{t-1} when estimated under restricted Correlated Random Effects with reduced form initial conditions (Column iv) [which only allows unobserved individual effects to be related to the means of observed characteristics], and unrestricted Correlated Random Effects with reduced form initial conditions (Column v) [which allows time-varying correlation between unobserved individual effects and observed characteristics]. There is not sufficient evidence here to reject the null hypothesis that π , the coefficient in a linear relationship between η_i and x_{it} , is constant over time.

More general still is the Honore and Kyriazidou Fixed Effects Model (Column vi). However, a Hausman test of the coefficient on y_{t-1} when estimated by this method against the coefficient on y_{t-1} when estimated under the most general of the other models CRE-UN (Column v) could not be carried out, as the estimated standard error when estimating y_{t-1} under FE-HK was less than the corresponding standard error under CRE-UN. However, a Hausman test of the coefficient on y_{t-1} was able to be carried out to compare FE-HK with the most general of the other random effects specifications CRE-R. The result of this test (unreported) rejects the null hypothesis that CRE-R is valid at the 5% level. By contrast with the previous paragraph, this is evidence against the null hypothesis that π , the coefficient in a linear relationship between η_i and x_{it} , is constant over time.

The result of a Hausman test of the coefficient on y_{t-1} comparing FE-HK with RE-RF was similar, rejecting the null hypothesis that RE-RF is valid at the 5% level ($p = 0.014$, see Column vi). These results give some cause for concern in assessing the validity of the Random Effects specifications, as it suggests there may be relationships between unobserved and observed characteristics beyond those postulated in the Random Effects specifications (for example, there may be nonlinear relationships between η_i and some/all of the x_{it} characteristics). If such relationships exist and are not controlled for, then the estimation techniques which use Random Effects may give rise to inconsistent parameter estimation.

To further investigate the nature of any relationships between η_i and the x_{it} characteristics, we consider the reduced form initial conditions model RE-RF, which imposes an orthogonality restriction between η_i and all of the x_{it} characteristics. Since RE-RF has an analogous linear specification which can be estimated (see Section 4.5.2 below), we focus attention upon this specification. To examine the appropriateness of the restrictions imposed by RE-RF, 8 further Hausman tests on the coefficients of each observed time-varying characteristic³¹ (including time trend coefficients) were calculated, comparing the coefficient estimated by RE-RF to that estimated by FE-HK which allows for general relationships between unobserved individual effects and observed characteristics. For example, the coefficient on marital status is negative and insignificant when allowing for correlation between unobserved individual effects and observed characteristics (allowed for by FE-HK, CRE-UN or CRE-R), but becomes positive and significant at the 5% level when the restriction of no correlation is imposed (by RE-RF, RE-EQ and RE-EX). This may simply reflect greater efficiency of estimation given no such correlation exists (i.e. that it is valid to restrict the unobserved individual effects to be orthogonal to observed characteristics). Alternatively, it may reflect that the orthogonality restriction between the individual effect and marital status (i.e. the null hypothesis) is invalid, suggesting that individuals with “more employable” unobserved characteristics are also more likely to be cohabiting. The same ambiguity is present in the interpretation of the coefficient on years of education, which is also found to be significant only when the orthogonality restriction between the individual effect and years of education is imposed; hence, the need for further tests.

The results from these further Hausman tests (unreported) showed that the more restrictive specification being tested (i.e. RE-RF) is rejected at the 5% level of significance, based on the estimated coefficients for marital status and region, but not for any other time-varying characteristics. This provides some evidence of correlation between the unobserved individual effects and one or more of the included explanatory variables, which is evidence that using the RE-RF model may lead to inconsistency of parameter estimation.

In summary, we find from the Hausman tests that there is evidence some observed individual characteristics (e.g. marital status and region) are not orthogonal to unobserved individual characteristics. This finding may imply that parameter estimation by the specifications which use Random Effects may be inconsistent. Whether the size of

³¹Recall from Section 4.3.2 and the discussion of results in Section 4.5.1, both above, that FE-HK is not capable of identifying coefficients on time-invariant characteristics. Therefore, tests can only be carried out for time-varying characteristics.

any inconsistency in material is an issue we address in Section 4.5.1, below.

Magnitude of Results

As discussed in Section 4.5.1 above, even after controlling for unobserved heterogeneity there is strong evidence of structural state dependence in employment outcomes. However, it is not obvious from inspection of these results how to characterise the extent of state dependence, which we now turn to.

For the Logit model, since the predicted employment probability is

$$\begin{aligned} \Pr(y_{it} = 1 \mid y_{i,t-1}, X_{it}) &= E(y_{it} \mid y_{i,t-1}, X_{it}) \\ &= \frac{\exp[\delta y_{i,t-1} + \beta' X_{it}]}{1 + \exp[\delta y_{i,t-1} + \beta' X_{it}]} \end{aligned}$$

then, as put forward by Chamberlain (1984), for any level of observed characteristics $\tilde{X}_{it}$ by substitution it follows that

$$\begin{aligned} &\Pr(y_{it} = 1 \mid y_{i,t-1} = 1, X_{it} = \tilde{X}_{it}) - \Pr(y_{it} = 1 \mid y_{i,t-1} = 0, X_{it} = \tilde{X}_{it}) \quad (4.20) \\ &= \frac{\exp[\delta + \beta' \tilde{X}_{it}]}{1 + \exp[\delta + \beta' \tilde{X}_{it}]} - \frac{\exp[\beta' \tilde{X}_{it}]}{1 + \exp[\beta' \tilde{X}_{it}]} \end{aligned}$$

We describe this expression, measured in probability units (i.e. percentage points scaled by 0.01), as the **marginal prior employment effect**, or for brevity just the **marginal effect**. Equation (4.7) allows us to use the coefficients identified by nonlinear estimation, to construct transition probabilities for any of the specifications that use Random Effects. (As noted in Section 4.3.2 above, the HK Fixed Effects model is incapable of generating predicted employment probabilities).

The results of applying Equation (4.20) to the RE-RF and CRE-R specifications are shown in Table 4.4, where we evaluate all transition probabilities at sample mean characteristics $\tilde{X}_{it} = \bar{X}$, $\forall i$.³² To focus attention upon the estimated δ coefficients on the lagged dependent variable reported in Table 4.3, these are repeated in Table 4.4 for the RE-RF and CRE-R specifications. Imposing the orthogonality condition (RE-RF, Column I) allows

³²Sample means are calculated across both individuals and time i.e. $\bar{X} \equiv \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N X_{it}$

Table 4.4: Marginal Effects: Nonlinear Models (Main sample: N=1285, T=1991-2000)

	(RE-RF) (i)	(CRE-R) (ii)
COEFFICIENTS		
y_{t-1}	2.361	2.322
INDEX AND PROBABILITIES		
Employment index value	2.66	2.38
Pr(StayEmployed)	0.99	0.99
Pr(NewlyEmployed)	0.93	0.91
Marginal Effect	0.06	0.08

Notes:

Employment index value is the value of the employment index $\sum_{j=1}^k \beta_j \bar{X}_j$, using estimated coefficients β_j (excluding coefficient on y_{t-1}) and evaluated at population average values.

Pr(StayEmployed) is $\Pr(\text{Employed at time } t \mid \text{Employed at time } t-1)$

Pr(NewlyEmployed) is $\Pr(\text{Employed at time } t \mid \text{Unemployed at time } t-1)$

Marginal Effect (the marginal effect of prior employment upon the probability of current employment) is $\Pr(\text{StayEmployed}) - \Pr(\text{NewlyEmployed})$

us to estimate an overall marginal prior employment effect of 6 percentage points. This is slightly lower than if unobserved individual differences are allowed to be correlated with the means of observed individual characteristics (CRE-R, Column II), in which case the estimated marginal effect is 8 percentage points. Since there is very little difference between these two results, this suggests that imposing the orthogonality restriction in Column I is relatively innocuous for δ , our main parameter of interest.

In Section 4.5.1 above, it was pointed out that there is reason to question the results from specifications using Random Effects; however, bearing in mind that the differences in estimation of the δ coefficient on the lagged dependent variable are of the order of 0.3, this is likely to have only a small effect on the estimated extent of state dependence as presented in Table 4.4. For example, if we use the FE-HK coefficient estimate of 2.079 (from Table 4.3) instead of the CRE-R coefficient estimate of 2.322, in Column ii of Table 4.4, and hold all other estimated parameters constant using parameters estimated by CRE-R, the extent of estimated state dependence is unchanged to 2 decimal places at 6 percentage points. Therefore, since the restrictions imposed by Random Effects seem relatively innocuous for our main parameter of interest, we focus further diagnostic analysis (see Section 4.5.1 below) on Random Effects specifications CRE-R and RE-RF.

Effect of Controls upon Parameter Estimation

When estimating marginal effects, we can control for three sources of variation in employment probability:

1. Observed heterogeneity;
2. Unobserved heterogeneity; and,
3. Initial conditions.

Using either of the two specifications reported in Table 4.4 above, we control for all three. To examine whether parameter estimation results are sensitive to which controls are used, we now consider results from diagnostic analysis. Table 4.5 reports estimated marginal effects over 1991-2000 using six specifications:

- Logit model with only a constant and lagged dependent variable [Column i]
- Logit model with full set of explanatory variables [Column ii]

Table 4.5: Effect of Controls upon Estimated State Dependence
(Main sample: N=1285, T=1991-2000)

	POOLED LOGIT (i)	POOLED LOGIT (ii)	RE-EX (iii)	RE-EX (iv)	RE-RF (v)	RE-RF (vi)
COEFFICIENTS						
y_{t-1}	4.018	3.607	3.903	2.781	2.725	2.361
INDEX AND PROBABILITIES						
Employment index value	-0.00	0.55	0.15	2.09	1.21	2.66
Pr(StayEmployed)	0.98	0.98	0.98	0.99	0.98	0.99
Pr(NewlyEmployed)	0.50	0.63	0.54	0.89	0.77	0.93
Marginal Effect	0.48	0.35	0.44	0.10	0.21	0.06

Notes:

Employment index value is the value of the employment index $\sum_{j=1}^k \beta_j \bar{X}_j$, using estimated coefficients β_j (excluding coefficient on y_{t-1}) and evaluated at population average values.

Pr(StayEmployed) is $\Pr(\text{Employed at time } t \mid \text{Employed at time } t-1)$

Pr(NewlyEmployed) is $\Pr(\text{Employed at time } t \mid \text{Unemployed at time } t-1)$

Marginal Effect (the marginal effect of prior employment upon the probability of current employment) is $\Pr(\text{StayEmployed}) - \Pr(\text{NewlyEmployed})$

- Random Effects Logit model with only a constant and lagged dependent variable [Column iii]
- Random Effects Logit model with full set of explanatory variables [Column iv]
- Random Effects Logit model with reduced form initial conditions and only a constant and lagged dependent variable [Column v]
- Random Effects Logit model with reduced form initial conditions and full set of explanatory variables [Column vi]

To evaluate the effect of controlling for observed and/or unobserved heterogeneity, we examine results from the four leftmost columns (assuming exogenous initial conditions throughout). The pooled Logit specification in Column i does not control for observed or unobserved heterogeneity, and estimates a relatively high marginal effect of 48 percentage points. Controlling for observed heterogeneity but not for unobserved heterogeneity, using the pooled Logit specification in Column ii, results in a lower estimated marginal effect of 35 percentage points. Similarly, controlling for unobserved heterogeneity but not for observed heterogeneity, using the Random Effects Logit specification in Column iii, results in an estimated marginal effect of 44 percentage points, which is lower than the 48 percentage point marginal effect reported in Column i. Controlling for both observed and unobserved heterogeneity, using the Random Effects Logit specification RE-EX in Column iv, results in a substantially lower estimated marginal effect of 10 percentage points.

To evaluate the effect of controlling for observed heterogeneity and/or initial conditions, we examine results from the four rightmost columns (controlling for unobserved heterogeneity throughout). The Random Effects specification in Column iii does not control for observed heterogeneity or initial conditions, and estimates a marginal effect of 44 percentage points. Controlling for observed heterogeneity but not for initial conditions, using the Random Effects specification in Column iv, results in a substantially lower estimated marginal effect of 10 percentage points. Similarly, controlling for initial conditions but not for observed heterogeneity, using the Random Effects Logit specification in Column v, results in the estimated marginal effect falling from 44 percentage points to 21 percentage points. Controlling for both observed heterogeneity and initial conditions, using the Random Effects Logit specification RE-RF in Column vi, results in a still lower estimated marginal effect of 6 percentage points.

These results indicate that controlling for all three sources of variation in employment probabilities is important, and that the incremental effectiveness of each control depends upon whether other controls are applied.

Summary of Results from Nonlinear Specifications

Among the specifications for which marginal effects can be calculated, the two specifications which seem most plausible are CRE-R and RE-RF, which estimate the marginal effect of prior employment to be, respectively, 8 percentage points and 6 percentage points. There is little difference between these results, especially when viewed in light of the results from diagnostic analysis which show that much larger marginal effects are estimated if controls for individual characteristics (i.e. observed heterogeneity), unobserved heterogeneity and initial conditions are not applied. In terms of the overall conclusions, it seems that there is little difference to results whether or not the orthogonality restriction between unobserved heterogeneity and individual characteristics is imposed. **Our most reliable estimation methods indicate that, for an individual with sample average characteristics, being employed one year earlier raises the likelihood of being currently employed, from around 91% to 93% (if previously unemployed) to around 99% (if previously employed).**

4.5.2 Results from Linear Models

Tests of Imposed Restrictions

All of the results from Section 4.5.1 above are based upon the logit specification described by (4.1). As discussed in Section 4.3.1 above, the main advantage of using the logit specification to identify parameters of interest is that it naturally provides a way of dealing with the binary nature of employment/unemployment data. However, three restrictions are used extensively to estimate parameters in the nonlinear framework. These are:

1. Strict exogeneity of observed characteristics ("Exogeneity"). $E[x_{it}\varepsilon_{is}] = 0 \quad \forall s, t$
2. Orthogonality between unobserved individual effects and observed characteristics ("Orthogonality"). $E[x_{it}\eta_i] = 0 \quad \forall i, t$
3. Homoskedasticity of time-varying random disturbances ("Homoskedasticity"). $E[\varepsilon_{it}^2] = \sigma^2 \quad \forall i, t$

It is desirable to assess whether these restrictions are valid.

Linear modelling can, although not correcting for the biases imposed by treating a fundamentally nonlinear problem as though it were linear, allow us to test the validity of these restrictions in a linear framework. If the restrictions appear to be valid in a linear framework, then this is indirect evidence that they may also be valid in a nonlinear framework. Note however that, by construction, the LPM cannot be homoskedastic: see Section 4.3.1.

Conversely, if the restrictions are invalid and the nonlinearities of the environment modest, then it may be more accurate to use the linear framework described by (4.14) to estimate results; that is, the biases imposed by linear estimation may be of second-order importance compared with potentially larger biases caused by imposing invalid restrictions. At the very least, discovering that any/all of the three restrictions above are invalid will allow us to interpret any results with necessary caution.

Going from left to right across Table 4.3, the first three specifications RE-EX, RE-EQ and RE-RF impose both Exogeneity and Orthogonality. The next two specifications CRE-R and CRE-UN impose Exogeneity but partially relax Orthogonality by allowing a linear relationship between unobserved effects and observed characteristics. The final specification FE-HK imposes Exogeneity, but completely relaxes Orthogonality by allowing any

relationship between unobserved effects and observed characteristics. All six specifications impose Homoskedasticity.

As discussed in Section 4.5.1 above, we can test whether Orthogonality is a valid restriction within the nonlinear framework by comparing the results from the fixed effects specification FE-HK, which does not impose Orthogonality, with Random Effects specifications which partially or fully impose Orthogonality. Based on these results, we concluded that, although Orthogonality does not appear to be valid for all observed characteristics, imposing the restriction does not materially alter the key estimation results. However, there is no scope within the nonlinear framework to test the validity of the Exogeneity or Homoskedasticity restrictions. This motivates study of further results from linear modelling. Therefore, we now consider to what extent these restrictions find support from the data, in the context of a linear framework.

Strict Exogeneity Since all six nonlinear models require that the variables x_{it} are strictly exogenous with respect to ε_{it} the transient error terms, it is desirable to explore further whether this type of restriction is valid. Therefore, we examine the analogous linear models, where it is possible to relax such restrictions flexibly.

For each variable x_{it} other than Time and Time Squared, we compare two sets of results for the specification in Equation (4.6), which are obtained by Difference GMM: the first set calculated using an instrument set including $(x_{i,t-1}, x_{i,t-2})$ and the second set including the larger instrument set $(x_{i,t+2}, x_{i,t+1}, x_{it}, x_{i,t-1}, x_{i,t-2})$. As discussed in Section 4.3.2 and again in Section 4.5.1 above, the Hausman test is available to examine reported differences in individual coefficients. Additionally, the Sargan test and Sargan Difference test are available to assess validity of restrictions, as discussed in Section 4.3.3 above.

Table 4.6 shows these results. For example, in Column 1, coefficients are reported for a specification which uses the instrument set $\{y_{i,t-2}, y_{i,t-3}, y_{i,t-4}, x_{i,t-1}, x_{i,t-2}\}$ where the x instruments include lags of the 6 explanatory variables Marital, Children, Education, Health, Housing and Region. In Column 2, the instrument set is expanded to $\{y_{i,t-2}, y_{i,t-3}, y_{i,t-4}\}$, $\{Marital_{i,t+2}, Marital_{i,t+1}, Marital_{i,t}, Marital_{i,t-1}, Marital_{i,t-2}\}$ and $\{x_{i,t-1}, x_{i,t-2}\}$ where the x instruments include lags of the 5 explanatory variables other than Marital. Columns 3 to 7 are analogous to Column 2 and Column 8 uses the instrument set $\{y_{i,t-2}, y_{i,t-3}, y_{i,t-4}, x_{i,t}, x_{i,t-1}, x_{i,t-2}\}$, where the x instruments include lags of all 6 explanatory variables.

Sargan tests and Sargan Difference tests for Table 4.6 indicate that, at a 5%

Table 4.6: Results from Linear Models: Difference GMM
(Main sample: N=1285, T=1991-2000)

	BASE-D (1)	MAR-D (2)	CH-D (3)	ED-D (4)	H-D (5)	HOU-D (6)	RE-D (7)	ALL-D (8)
y_{t-1}	0.283 (.000)	0.286 (.000)	0.277 (.000)	0.282 (.000)	0.297 (.000)	0.277 (.000)	0.293 (.000)	0.310 (.000)
Marital	-0.005 (.678)	-0.010 (.114)	-0.014 (.199)	-0.002 (.773)	0.006 (.593)	-0.006 (.514)	-0.007 (.529)	-0.009 (.063)
Children	0.026 (.606)	0.010 (.865)	0.054 (.030)	0.007 (.855)	0.008 (.897)	0.033 (.434)	0.062 (.348)	0.027 (.034)
Education	-1.076 (.803)	0.441 (.912)	-2.516 (.560)	0.703 (.069)	0.536 (.856)	0.243 (.935)	7.134 (.013)	0.393 (.375)
Health	0.007 (.678)	0.013 (.458)	0.005 (.748)	0.014 (.432)	0.017 (.112)	0.012 (.488)	0.015 (.392)	0.021 (.107)
Housing	0.032 (.035)	0.031 (.037)	0.032 (.025)	0.036 (.016)	0.034 (.018)	0.010 (.095)	0.049 (.005)	0.013 (.022)
Region	0.108 (.256)	0.139 (.119)	0.116 (.177)	0.120 (.253)	0.121 (.216)	0.136 (.097)	0.032 (.224)	0.035 (.378)
Time	0.085 (.003)	0.066 (.023)	0.098 (.000)	0.051 (.001)	0.072 (.002)	0.066 (.001)	0.054 (.034)	0.049 (.000)
Time ²	-0.050 (.003)	-0.038 (.030)	-0.057 (.000)	-0.028 (.004)	-0.043 (.004)	-0.038 (.001)	-0.040 (.021)	-0.027 (.001)
A-Bond m2	0.99	0.99	0.96	0.95	1.08	0.90	0.96	1.19
Sargan-p	0.65	0.56	0.79	0.73	0.72	0.51	0.72	0.41
Sargan df	108	129	129	129	129	129	129	129
SarganD-p	—	0.29	0.88	0.72	0.69	0.20	0.69	—
Hausman-p β_j	—	0.64	0.52	0.67	0.55	0.11	0.40	—

Notes:

Estimation is via 2 step Difference GMM with heteroskedasticity-robust standard errors.

p-values based on the Normal distribution are in parentheses.

All columns' respective instrument sets include $y_{i,t-2}, y_{i,t-3}, y_{i,t-4}$.

The Baseline specification in Column 1 includes $x_{i,t-1}, x_{i,t-2}$ for all regressors, with the exceptions of Time and Time Squared.

Columns 2 to 7 include $x_{i,t-1}, x_{i,t-2}$ for all regressors other than that being tested, with the exceptions of Time and Time Squared.

Columns 2 to 7 include $x_{i,t+2}, x_{i,t+1}, x_{i,t}, x_{i,t-1}, x_{i,t-2}$ for the regressor being tested. For example, these past and future values of Marital Status are included in the instrument set in column 2.

Column 8 includes $x_{i,t}, x_{i,t-1}, x_{i,t-2}$ for all regressors except Time and Time Squared.

A-Bond m2 is Arellano and Bond's (1991) m2 statistic (see Section 4.3.3).

Sargan-p is the Sargan Test statistic and Sargan df is the associated degrees of freedom

SarganD-p is the Sargan Difference test statistic (see Section 4.3.3). This is tested according to a χ^2 distribution with $(129 - 108 =) 21$ degrees of freedom, comparing each of the specifications in Columns 2 to 7 against the Baseline specification in Column 1.

Hausman statistic on the coefficient of the respective regressor of interest for column j β_j ($j = 2, \dots, 7$) is tested according to a χ^2 distribution with 1 degree of freedom. For example, the Hausman test reported in column 2 compares the coefficient on Marital in column 2 to the coefficient on Marital in column 1.

significance level, there is insufficient evidence to reject the null hypothesis that any of the six explanatory variables, when considered singly, is strictly exogenous with respect to transient disturbances. In broad agreement with these findings, Hausman tests on the respective coefficients corresponding to changes in instrument sets do not reject the respective null hypotheses that each coefficient is unchanged, at the 5% level.

Other than the instrument validity tests, there is evidence that the quadratic time trend and the housing variable have significant explanatory power for employment outcomes over the 1990s.

Importantly, Table 4.6 does not provide evidence that imposing Exogeneity in the linear context is invalid, and suggests that imposing the restriction of strict exogeneity in the context of the nonlinear models may be innocuous. This is also demonstrated by the apparent stability of the estimated coefficient g on the lagged dependent variable y_{t-1} across all eight specifications. Since this is the main parameter of interest, this finding is also an indication of the robustness of estimation results within the linear framework

Orthogonality of Individual Effects As for the nonlinear models, we examine whether imposing Orthogonality is valid in a linear context. If any strictly exogenous variable x satisfies the orthogonality restriction $E[x_{it}\eta_i] = 0$, then its current level x_{it} is a valid instrument for the current levels equation.

We calculate a new baseline specification which uses all the instruments for the difference equations that were used in Column 8 of Table 4.6. For comparison we calculate, for each variable x other than Time and Time Squared, a specification equivalent to the baseline specification with - additionally - the current levels of each explanatory variable x_{it} used as instruments for the level equations. We also calculate a specification which includes all 6 explanatory variables X_{it} as instruments for the level equations: these additional instruments are only valid to the extent that the Orthogonality restriction holds true, and we would expect inclusion of these instruments to lead to inconsistent parameter estimation if the restriction is not true. The Hausman, Sargan and Sargan Difference tests can again be used to test the validity of the extra restrictions.

Table 4.7 shows these results. None of the Sargan Difference statistics are significantly different from zero, so based on those we do not reject the null hypothesis that these observed characteristics are orthogonal to unobserved heterogeneity. However, the Hausman test based on the coefficient of marital status does reject orthogonality between this variable and the individual-specific effects at the 5% level ($p=0.01$). This is the only

Table 4.7: Results from Linear Models: System GMM
(Main sample: N=1285, T=1991-2000)

	BASE-S (1)	MAR-S (2)	CH-S (3)	ED-S (4)	H-S (5)	HOU-S (6)	RE-S (7)	PREF-S (8)
y_{t-1}	0.310 (.000)	0.314 (.000)	0.310 (.000)	0.313 (.000)	0.306 (.000)	0.313 (.000)	0.302 (.000)	0.319 (.000)
Constant	— —	0.582 (.000)	0.568 (.000)	0.612 (.000)	0.565 (.000)	0.513 (.000)	0.574 (.000)	0.608 (.000)
Marital	-0.009 (.063)	-0.002 (.626)	-0.009 (.048)	-0.008 (.059)	-0.009 (.066)	-0.007 (.084)	-0.007 (.077)	-0.006 (.102)
Children	0.027 (.034)	0.028 (.087)	0.014 (.399)	0.026 (.095)	0.025 (.053)	0.024 (.156)	0.019 (.187)	0.016 (.257)
Education	0.393 (.375)	0.184 (.543)	0.402 (.304)	0.075 (.339)	0.482 (.305)	0.759 (.178)	0.458 (.528)	0.009 (.795)
Health	0.021 (.107)	0.024 (.091)	0.020 (.110)	0.019 (.136)	0.018 (.121)	0.025 (.053)	0.025 (.054)	0.018 (.091)
Housing	0.013 (.022)	0.012 (.030)	0.014 (.018)	0.012 (.029)	0.014 (.014)	0.020 (.002)	0.015 (.018)	0.024 (.000)
Region	0.035 (.378)	0.039 (.342)	0.039 (.341)	0.031 (.407)	0.035 (.384)	0.035 (.367)	-0.001 (.913)	0.028 (.341)
Time	0.049 (.000)	0.046 (.000)	0.052 (.000)	0.040 (.002)	0.048 (.000)	0.047 (.000)	0.066 (.000)	0.040 (.001)
Time ²	-0.027 (.001)	-0.026 (.003)	-0.029 (.001)	-0.021 (.019)	-0.026 (.001)	-0.026 (.002)	-0.042 (.000)	-0.021 (.004)
A-Bond m2	1.19	1.22	1.19	1.22	1.16	1.22	1.16	1.30
Sargan-p	0.41	0.31	0.41	0.31	0.46	0.33	0.58	0.49
Sargan df	156	166	166	166	166	166	166	196
SarganD-p	—	0.13	0.43	0.14	0.66	0.17	0.99	0.65
Hausman-p β_j	—	0.01	N/A	0.47	0.60	N/A	0.36	—

Notes:

Estimation is via 2 step System GMM with heteroskedasticity-robust standard errors.

p-values based on the Normal distribution are in parentheses.

All instrument sets for the difference equations include $y_{i,t-2}, y_{i,t-3}, y_{i,t-4}$, and $x_{i,t}, x_{i,t-1}, x_{i,t-2}$ for all regressors other than Time and Time Squared.

Column 1 does not include any instruments for the level equations, and so is estimated entirely from the difference equations.

Columns 2 to 7 include $x_{i,t}$ for the regressor being tested as an instrument for the level equations.

Column 8 includes $x_{i,t}$, for all regressors other than Marital, Region, Time and Time Squared, as instruments for the level equations.

A-Bond m2 is Arellano and Bond's (1991) m2 statistic (see Section 4.3.3).

Sargan-p is the Sargan Test statistic and Sargan df is the associated degrees of freedom

SarganD-p is the Sargan Difference statistic (see Section 4.3.3). This is tested according to a χ^2 distribution with $(166 - 156 =) 10$ degrees of freedom, comparing each of the specifications in Columns 2 to 7 against the Baseline specification in Column 1.

Sargan Difference statistic in column 8 is tested according to a χ^2 distribution with $(196 - 156 =) 40$ degrees of freedom.

Hausman statistic on the coefficient of the respective regressor of interest in column j β_j ($j = 2, \dots, 7$) is tested according to a χ^2 distribution with 1 degree of freedom. For example, the Hausman test reported in column 2 compares the coefficient on Marital Status in column 2 to the coefficient on Marital Status in column 1.

such test where the null hypothesis of Orthogonality is rejected³³. We therefore exclude the current level of marital status from the instrument set in Column 8. Although there is little evidence that region is correlated with the individual-specific effects from diagnostic tests (Hausman test $p=0.35$), nonetheless including it in the instrument set causes computational problems due to its lack of time series variation; therefore, we exclude it from the instrument set in Column 8 as well.

There is little evidence against the validity of the PREF-S specification in Column 8, which we call the preferred linear specification. This specification (which is also the most restrictive specification in Table 4.7) thus utilises orthogonality restrictions upon 4 of the explanatory variables: number of children, years of education, health and housing status. Its full instrument set includes the following instruments for the equations in first differences:

- Lagged employment status variables $y_{i,t-2}, y_{i,t-3}, y_{i,t-4}$
- Current and lagged explanatory variables $X_{i,t}, X_{i,t-1}, X_{i,t-1}$ for marital status, number of children, years of education, health, housing and region.

Additionally it includes the following instruments for the equations in levels:

- Current explanatory variables $X_{i,t}$ for number of children, years of education, health and housing.

However, these extra restrictions do not make a great difference to the estimated coefficients, or to their estimated significance. As was the case for the most restrictive specification reported in Table 4.6 (repeated as Column 1 in Table 4.7), housing status and the quadratic time trend have strongly significant effects upon employment outcomes. Good health also has a significant positive effect upon being employed, while cohabitation is found to be negatively correlated with being employed. However, the most restrictive model in Column 8 does not identify a significant effect of children upon employment status.

Again, it is noteworthy that these results are robust, in the sense that the estimated coefficient g on the lagged dependent variable y_{t-1} is apparently stable across these eight specifications.

³³Two of the Hausman tests were unavailable due to the standard errors on respective coefficients in the more restrictive models in Columns 3 and 6 being greater than those in the least restrictive model of Column 1, and are marked "N/A".

Homoskedasticity The assumption that the time-varying error terms are homoskedastic $E[\varepsilon_{it}^2] = \sigma^2 \quad \forall i, t$ is one which we can readily test using the linear framework. However, we are wary of drawing conclusions concerning the nonlinear models from any results using the linear framework. This is because MLE is used to estimate parameters for the logit models in Section 4.5.1 above, and MLE conceivably may be inconsistent for the parameters of interest if disturbances are heteroskedastic. By contrast, since GMM is used to estimate the linear models then, if parameters are estimated using a non-heteroskedasticity robust procedure, estimation of the standard errors may be inconsistent but estimation of the parameters of interest themselves is not impaired.

From the discussion in Section 4.3.1, we know that Homoskedasticity is not a valid restriction within the LPM since the binary choice model estimated within the LPM framework makes the errors heteroskedastic by construction. To see the results from this, we separately estimate results using a heteroskedasticity-robust specification and a specification which assumes disturbances are homoskedastic. All of the specifications estimated in Table 4.7 are heteroskedasticity-robust in the White (1980) sense (see Appendix C.2 for further details). We therefore compare the baseline linear specification BASE-S and the preferred linear specification PREF-S, from Columns 1 and 8 of Table 4.7, with corresponding specifications which estimate standard errors that are not heteroskedasticity-robust.

Table 4.8 shows these results. It can be seen that several explanatory variables, especially Region, have significant effects upon employment status when estimated using standard errors that are not heteroskedasticity-robust but are insignificant or marginally significant using heteroskedasticity-robust standard errors. However, the main parameter of interest, the coefficient on lagged employment y_{t-1} is clearly significant regardless of whether heteroskedasticity-robust standard errors are used. The estimated standard error on y_{t-1} is 0.032 for BASE-S when heteroskedasticity-robust standard errors are used, compared to 0.009 when standard errors that are not heteroskedasticity-robust are used. Similarly, the estimated standard error on y_{t-1} is 0.030 for PREF-S when heteroskedasticity-robust standard errors are used, compared to 0.008 when standard errors that are not heteroskedasticity-robust are used. Using heteroskedasticity-robust standard errors more than trebles the estimated standard errors.

We therefore conclude that the main result of a significantly positive coefficient g on the lagged dependent variable is robust to whether estimation is proceeds using the restriction of Homoskedasticity (which is clearly invalid within the linear model). Although this result cannot be directly applied back to the nonlinear context, it nonetheless does

not provide substantial evidence against the suitability of imposing Homoskedasticity in a nonlinear environment.

Power of Tests The diagnostic tests used in Tables 4.6, 4.7 and 4.8 do not reject many of the restrictions imposed by DGMM or SGMM estimation, at the 5% level. Although this is consistent with these restrictions being valid, it is nonetheless also consistent with the possibility that these tests have low power at the sample size of this study. To investigate this issue further, further specifications which use demonstrably invalid, or otherwise implausible, restrictions were also considered. The purpose of this is to ascertain whether the same diagnostic tests were sufficiently powerful to reject demonstrably invalid or otherwise implausible restrictions. This is a similar check to the diagnostic analysis carried out for the nonlinear specifications and reported in Table 4.5, where we examined how robust parameter estimation is to different specifications.

Table 4.9 shows estimation results for the linear specifications. In the context of Difference GMM estimation, Columns 1 and 2 respectively show results from the most restrictive Difference GMM estimator (Column 1, see also in Column 8 of Table 4.6) side by side with those from the same estimator with the level of y_{t-1} added to the instrument set for equations in *first differences* (Column 2). Due to the introduction of the error term from the previous period into the first-differenced error term, the dependent variable from the previous period is not a valid instrument. All diagnostic tests overwhelmingly reject the model in Column 2 ($p=0.00$).

In the context of SGMM estimation, results from the most restrictive SGMM estimator (Column 3, see also in Column 8 of Table 4.7) are juxtaposed with those from an estimator which adds the level of y_{t-1} added to the instrument set for equations in *levels* (Column 4). If unobserved individual-specific effects are non-zero, these must be correlated with the level of the lagged dependent variable y_{t-1} . Although this model is not rejected by the Sargan test ($p=0.22$), it is nonetheless rejected by the Hausman and Sargan Difference tests ($p=0.00$).

Additionally, results from estimating a linear regression using Ordinary Least Squares are included (Column 5). These results show evidence of second order residual correlation (A-Bond $m2 = 3.42$), and also allow us to compare results with the preferred linear estimation method in Column 3. The Hausman test's null hypothesis that the coefficient of y_{t-1} is the same in Columns 3 and 5 is overwhelmingly rejected ($p=0.00$).

These exercises provide evidence that the failure of our tests to reject the restric-

Table 4.8: Diagnostic Analysis on Standard Errors from Linear Models
(Main sample: N=1285, T=1991-2000)

	(BASE-S) (Robust) (1)	(BASE-S) (Non-Robust) (2)	(PREF-S) (Robust) (3)	(PREF-S) (Non-Robust) (4)
COEFFICIENTS				
y_{t-1}	0.310 (.000)	0.310 (.000)	0.319 (.000)	0.319 (.000)
Constant	— —	— —	0.608 (.000)	0.608 (.000)
Marital	-0.009 (.063)	-0.009 (.000)	-0.006 (.102)	-0.006 (.001)
Children	0.027 (.034)	0.027 (.007)	0.016 (.257)	0.016 (.077)
Education	0.393 (.375)	0.393 (.157)	0.009 (.795)	0.009 (.765)
Health	0.021 (.107)	0.021 (.010)	0.018 (.091)	0.018 (.004)
Housing	0.013 (.022)	0.013 (.000)	0.024 (.000)	0.024 (.000)
Region	0.035 (.378)	0.035 (.000)	0.028 (.341)	0.028 (.000)
Time	0.049 (.000)	0.049 (.000)	0.040 (.001)	0.040 (.000)
Time ²	-0.027 (.001)	-0.027 (.000)	-0.021 (.004)	-0.021 (.000)

Notes:

Estimation is via 2 step System GMM. Columns (1) and (3) are estimated using heteroskedasticity-robust standard errors. Columns (2) and (4) are estimated using standard errors which are not robust to heteroskedasticity.

p-values based on the Normal distribution are in parentheses.

Columns (1) and (2) both use the same instrument set, as for Column (1) of Table 4.7.

Columns (3) and (4) both use the same instrument set, as for Column (8) of Table 4.7.

Table 4.9: Results from Linear Models using Invalid or Implausible Restrictions
(Main sample: N=1285, T=1991-2000)

	ALL-D (1)	DGMM1A (2)	ALL-S (3)	SGMM1A (4)	OLS (5)
y_{t-1}	0.310 (.000)	-0.111 (.000)	0.319 (.000)	0.416 (.000)	0.461 (.000)
Constant	— —	— —	0.608 (.000)	0.506 (.000)	0.392 (.000)
Marital	-0.009 (.063)	-0.004 (.201)	-0.006 (.102)	-0.007 (.056)	0.011 (.038)
Children	0.027 (.034)	0.030 (.036)	0.016 (.257)	0.017 (.242)	-0.040 (.017)
Education	0.393 (.375)	1.423 (.052)	0.009 (.795)	0.009 (.705)	0.169 (.020)
Health	0.021 (.107)	0.000 (.913)	0.018 (.091)	0.021 (.055)	0.040 (.002)
Housing	0.013 (.022)	0.003 (.533)	0.024 (.000)	0.022 (.000)	0.044 (.000)
Region	0.035 (.378)	0.064 (.125)	0.028 (.341)	0.044 (.279)	0.005 (.110)
Time	0.049 (.000)	0.010 (.183)	0.040 (.001)	0.036 (.004)	0.057 (.059)
Time ²	-0.027 (.001)	-0.004 (.373)	-0.021 (.004)	-0.019 (.027)	-0.027 (.244)
A-Bond m2	1.19	-4.68	1.30	1.87	3.42
Sargan-p	0.41	0.00	0.49	0.23	—
Sargan df	156	116	196	205	—
SarganD-p	—	0.00	—	0.00	—
Hausman-p	—	0.00	—	0.00	0.00

Notes:

Estimation is via 2 step Difference GMM (Columns 1-2), 2 step System GMM (Columns 3-4), Ordinary Least Squares (Column 5) with heteroskedasticity-robust standard errors.

p-values based on the Normal distribution are in parentheses.

All four instrument sets for the difference equations include $y_{i,t-2}, y_{i,t-3}, y_{i,t-4}$, and $x_{i,t}, x_{i,t-1}, x_{i,t-2}$ for all other regressors than Time and Time Squared.

Instrument set for the difference equations in Column 2 includes $y_{i,t-1}$.

Instrument set for the level equations in Column 3 includes $x_{i,t}$ for all regressors other than Marital Status, Region, Time and Time Squared.

Instrument set for the level equations in Column 4 is as for Column 3, and also $y_{i,t-1}$.

A-Bond m2 is Arrelano and Bond's (1991) m2 statistic (see Section 4.3.3).

Sargan-p is the Sargan Test statistic and Sargan df is the associated degrees of freedom

SarganD-p is the Sargan Difference test statistic (see Section 4.3.3). Sargan Difference statistic in Column 2 is tested according to a χ^2 distribution with $(116 - 108 =) 8$ degrees of freedom.

Sargan Difference statistic in Column 4 is tested according to a χ^2 distribution with $(205 - 196 =) 9$ degrees of freedom.

Hausman statistics on the coefficient of $y_{i,t-1}$ are tested according to a χ^2 distribution with 1 degree of freedom. The Hausman test reported in column 2 compares the coefficient on $y_{i,t-1}$ in column 2 to the coefficient on $y_{i,t-1}$ in column 1. The Hausman tests reported in columns 4 and 5 compare the coefficient on $y_{i,t-1}$ in column 3 to the respective coefficients on $y_{i,t-1}$ in columns 4 and 5.

tions in Tables 4.6, 4.7 and 4.8 is not purely because of a lack of power of the tests in all circumstances. It should be noted that, in contrast with tables 4.6, 4.7 and 4.8, the estimated coefficient g on the lagged dependent variable y_{t-1} is highly sensitive to these specifications, in one case (Column 2) even estimated to be negative. This reassures us that the restrictions being placed on the earlier specifications in Tables 4.6, 4.7 and 4.8 are indeed relatively innocuous and that our main estimation results for the linear models are reliable.

Magnitude of Results

For the linear results, where (as here) outcomes are binary the coefficient on a lagged dependent variable is equivalent to a transition probability. This is helpful as it allows us to interpret the estimated coefficient $\hat{g}_{GMM}$ as the extent to which a respondent employed one year previously is more likely to be employed currently than a respondent unemployed one year previously, after controlling for the two respondents' observed characteristics. That is, the estimated marginal effect is independent of the values of explanatory variables $\tilde{X}_{it}$:

$$\Pr(y_{it} = 1 \mid y_{i,t-1} = 1, X_{it} = \tilde{X}_{it}) - \Pr(y_{it} = 1 \mid y_{i,t-1} = 0, X_{it} = \tilde{X}_{it}) = \hat{g}_{GMM} \quad \forall i, t \quad (4.21)$$

Therefore, in contrast to Equation (4.20) used for the nonlinear framework, we do not need to evaluate transition probabilities at particular values of the explanatory variables if all we wish to do is estimate marginal effects: Equation (4.21) is sufficient for this within the linear framework. However, since we are also interested in the transition probabilities themselves, we need to specify at which values of $\tilde{X}_{it}$ the transition probabilities are to be evaluated. As we did for the nonlinear specifications, we evaluate transition probabilities at sample mean characteristics $\tilde{X}_{it} = \bar{X}$, $\forall i$.

We focus upon the preferred linear specification PREF-S (see Table 4.7 and discussion in Section 4.5.2 above), which uses the following instrument set for the equations in first differences:

- Lagged employment status variables $y_{i,t-2}, y_{i,t-3}, y_{i,t-4}$
- Current and lagged explanatory variables $X_{i,t}, X_{i,t-1}, X_{i,t-1}$ for marital status, number of children, years of education, health, housing and region

Table 4.10: Marginal Effects: Nonlinear and Linear Models
(Main sample: N=1285, T=1991-2000)

	(RE-RF) (1)	(CRE-R) (2)	(PREF-S) (3)
COEFFICIENTS			
y_{t-1}	2.361	2.322	0.319
INDEX AND PROBABILITIES			
Employment index value	2.66	2.38	—
Pr(StayEmployed)	0.99	0.99	0.99
Pr(NewlyEmployed)	0.93	0.91	0.67
Marginal Effect	0.06	0.08	0.32

Notes:

Employment index value is the value of the employment index $\sum_{j=1}^k \beta_j \bar{X}_j$, using estimated coefficients β_j (excluding coefficient on y_{t-1}) and evaluated at population average values.

Pr(StayEmployed) is $\Pr(\text{Employed at time } t \mid \text{Employed at time } t-1)$

Pr(NewlyEmployed) is $\Pr(\text{Employed at time } t \mid \text{Unemployed at time } t-1)$

Marginal Effect (the marginal effect of prior employment upon the probability of current employment) is $\Pr(\text{StayEmployed}) - \Pr(\text{NewlyEmployed})$

and includes the following instrument set for the equations in levels:

- Current explanatory variables $X_{i,t}$ for number of children, years of education, health and housing

It can therefore be seen from Table 4.10 that the marginal effect estimated using the most reliable linear specification is much greater than the marginal effect estimated by either of the nonlinear specifications.

Summary of Results from Linear Specifications

Results from the linear model in (4.14) estimate a marginal effect of 32 percentage points, compared with 6 percentage points to 8 percentage points for the nonlinear model. This marginal effect is made up of two components, which we summarise.

Our results suggest an individual with sample average characteristics who was employed last period has a likelihood of being employed of around 99%. These results

are robust across most of the specifications estimated in this sub-section, and are almost identical to the results from nonlinear estimation summarised in Section 4.5.1 above.

However, there is a noticeable difference in employment likelihoods if unemployment is the prior state. Using our linear model, we estimate an individual with sample average characteristics who was unemployed last period has a likelihood of being employed of around 67%. Although these results are fairly robust within the linear specifications examined, they suggest a much lower probability of being employed this period for such an individual, compared with results from nonlinear estimation which estimate the employment likelihood around 91% to 93%. In this sense, the estimated marginal effect (i.e. the extent of unemployment persistence) is much greater when a linear model is used, which may be as a result of misspecification caused by the linear model.

4.6 Robustness Analysis

Restricting attention to a 10 year balanced panel may potentially result in misleading estimates of state dependence for two reasons:

1. Cross-sectional parameter variation (“Selection Bias”).
2. Time-series parameter variation. (“Regime Shifts”)

Selection Bias

If different individuals experience state dependence to differing extents, and this is related to our sample selection rule (as described in Section 4.4.2 above), then our estimators of state dependence will cease to be representative of the population as a whole. For example, if individuals who are persistently employed are more likely to be contactable by survey organisers (e.g. because they are more likely to maintain stable addresses) then they will be over-represented in our sample.

If so, our selection rule systematically excludes a part of the population with distinct characteristics relevant to the extent of state dependence, and parameter estimation will be inconsistent due to selection bias [see, for example, Verbeek and Nijman(1996)]. The population can be divided into two cross-sections – depending on whether included and excluded in our eventual sample - and the parameters of interest are different between cross-sections. Of course, if the two parameters of interest are the same across the two

cross-sections, then we are not concerned, and our sample selection rule results in consistent parameter estimation.

Regime Shifts

By only estimating one set of parameters across the sample of 10 years, we assume that the parameters are stable across time. In the event of a regime shift, where the environment fundamentally changes during the sample period such that the functional relationships remain the same but the parameters of interest change, then our methods would fail to capture this. Therefore, our parameter estimation would be inconsistent for both the subperiods before and after a regime shift.

It is therefore important to examine how robust our results are to these two possible misspecifications.

4.6.1 Sample Selection

To examine the effect of using a balanced panel, we perform estimation on two auxiliary samples. These are:

- Sample E1.

Time period 1991-1995. Selected according to all individuals reporting employment or unemployment during all 10 years 1991-2000. Total observations = 1285 individuals x 5 waves = 6425.

- Sample F1.

Time period 1991-1995. Selected according to all individuals reporting employment or unemployment during all 5 years 1991-1995, but missing values for at least one year in 1996-2000. Total observations = 769 individuals x 5 waves = 3845.

Sample E1 is exactly the same cross-section of surveyed men as our main sample, but we only look at the shorter time period 1991-1995. Sample F1 is the sample of men that had no missing observations during 1991-1995 but were excluded from our main sample on the grounds that they were missing at least one observation over 1996-2000. If the selection rule for the main sample had been to include all men with 5 observations over 1991-1995, then both E1 and F1 would have been included in the main sample over 1991-2000. Therefore, E1 and F1 can be viewed as complementary samples: respectively included

Table 4.11: Marginal Effects over different Sample Selection Rules

	MAIN SAMPLE RE-RF (i)	MAIN SAMPLE CRE-R (ii)	E1 RE-RF (iii)	E1 CRE-R (iv)	F1 RE-RF (v)	F1 CRE-R (vi)
COEFFICIENTS						
y_{t-1}	2.4 (.000)	2.3 (.000)	2.2 (.000)	2.2 (.000)	1.6 (.000)	1.6 (.000)
INDEX AND PROBABILITIES						
Employment index value	2.66	2.38	2.03	2.40	2.96	1.51
Pr(StayEmployed)	0.99	0.99	0.98	0.99	0.99	0.96
Pr(NewlyEmployed)	0.93	0.91	0.88	0.92	0.95	0.82
Marginal Effect	0.06	0.08	0.10	0.07	0.04	0.14

Notes:

Main Sample: N=1285, T=1991-2000

Sample E1: N=1285, T=1991-1995

Sample F1: N=769, T=1991-1995

p-values based on the Normal distribution are in parentheses.

Employment index value is the value of the employment index $\sum_{j=1}^k \beta_j \bar{X}_j$, using estimated coefficients β_j (excluding coefficient on y_{t-1}) and evaluated at population average values.Pr(StayEmployed) is $\Pr(\text{Employed at time } t \mid \text{Employed at time } t-1)$ Pr(NewlyEmployed) is $\Pr(\text{Employed at time } t \mid \text{Unemployed at time } t-1)$ Marginal Effect (the marginal effect of prior employment upon the probability of current employment) is $\Pr(\text{StayEmployed}) - \Pr(\text{NewlyEmployed})$

and excluded in the main sample 1991-2000 according to whether there were any missing observations over 1996-2000.

We report summary statistics and commentary for auxiliary samples E1 and F1 in Appendix C.3. We also report results from these samples using nonlinear specifications RE-RF and CRE-R: see discussion in Section 4.5.1 above for details of these models in relation to our main sample. These results for samples E1 and F1 are shown in Table 4.11.

The most important result from estimation on these auxiliary sample is that the estimated coefficient on lagged employment status remains positive and significantly different from zero, for both estimators and across both auxiliary samples.

In terms of magnitude of results, the more restrictive RE-RF results from either Sample E1 or F1 estimate a marginal effect which is only slightly higher than for the main

sample. Hausman tests (unreported) of the orthogonality restrictions for the specifications considered in Table 4.11, similar to those described in Table 4.3 for the main sample, fail to be rejected for either Sample E1 or F1. However, the same caveats as for the main sample discussed in Section 4.5.1 above (e.g. the ambiguity in interpreting the coefficient on marital status) apply to both of these samples, so it is not clear that the orthogonality restrictions imposed by RE-RF are valid.

The estimated marginal effect is higher in Sample E1 than in Sample F1, regardless of whether the orthogonality restrictions are imposed. Focussing on RE-RF, the coefficient on lagged employment status is 2.2 for Sample E1 and 1.6 for Sample F1. A Hausman Test is not available to formally test the difference between these two coefficients. However, as the standard error of estimate is approximately 0.18 for each of the two coefficients, it seems likely that the observed difference is significantly different from zero. In terms of marginal effects, results are fairly stable across estimation methods using Sample E1, with marginal effects ranging between 7 percentage points and 10 percentage points. For sample F1, this range is wider: using RE-RF the estimated marginal effect is 4 percentage points, while using CRE-R it is 14 percentage points.

Therefore, we conclude that the sample selection rule has some impact on parameter estimation. In particular, it seems likely that those individuals present in 1991-1995 who are more likely to be missing from the sample during 1996-2000 also experience less persistence in employment status relative to those more likely to be present in all 10 waves 1991-2000. This may be attributable to individuals more likely to change employment status also being more likely to move residence and inadvertently drop out of the survey.

4.6.2 Time Period

To examine whether parameters have remained stable over the period 1991-2000, we perform estimation on two additional samples. These are:

- Sample E1.

As per Section 4.6.1 above. Time period 1991-1995. Selected according to all individuals reporting employment or unemployment during all 10 years 1991-2000. Total observations = 1285 individuals x 5 waves = 6425.

- Sample E2.

Table 4.12: Estimated State Dependence over different Time Periods

	MAIN SAMPLE RE-RF (i)	MAIN SAMPLE CRE-R (ii)	E1 RE-RF (iii)	E1 CRE-R (iv)	E2 RE-RF (v)	E2 CRE-R (vi)
COEFFICIENTS						
y_{t-1}	2.4 (.000)	2.3 (.000)	2.2 (.000)	2.2 (.000)	3.4 (.000)	3.2 (.000)
INDEX AND PROBABILITIES						
Employment index value	2.66	2.38	2.03	2.40	1.66	0.96
Pr(StayEmployed)	0.99	0.99	0.98	0.99	0.99	0.98
Pr(NewlyEmployed)	0.93	0.91	0.88	0.92	0.84	0.72
Marginal Effect	0.06	0.08	0.10	0.07	0.15	0.26

Notes:

Main Sample: N=1285, T=1991-2000

Sample E1: N=1285, T=1991-1995

Sample E2: N=1285, T=1996-2000

p-values based on the Normal distribution are in parentheses.

Employment index value is the value of the employment index $\sum_{j=1}^k \beta_j \bar{X}_j$, using estimated coefficients β_j (excluding coefficient on y_{t-1}) and evaluated at population average values.Pr(StayEmployed) is $\Pr(\text{Employed at time } t \mid \text{Employed at time } t-1)$ Pr(NewlyEmployed) is $\Pr(\text{Employed at time } t \mid \text{Unemployed at time } t-1)$ Marginal Effect (the marginal effect of prior employment upon the probability of current employment) is $\Pr(\text{StayEmployed}) - \Pr(\text{NewlyEmployed})$

Time period 1996-2000. Selected according to all individuals reporting employment or unemployment during all 10 years 1991-2000. Total observations = 1285 individuals x 5 waves = 6425.

Results in Table 4.12 show that estimated state dependence is considerably higher during the last 5 waves 1996-2000 than during the first 5 waves 1991-1995. This could be partly due to all individuals in Sample E2 being five years older than in Sample E1 (see also the discussion of age effects in Section 4.6.3 below).

Nonetheless, our main conclusion from Sections 4.6.1 and 4.6.2 is that the main result of nonzero persistence (i.e. positive marginal effects) established in Section 4.5 above is unchallenged by these findings. Therefore, our main result is robust both to different selection rules and to different time periods

during the 1990s.

4.6.3 Comparison of Results with those of other Studies

As described in Section 4.5.1 above, we use Equation (4.20) to transform the estimated coefficient on lagged employment status into two probabilities: first, the probability of currently being employed, conditional upon being employed one year previously; second, the probability of currently being employed, conditional upon being unemployed one year previously. Both probabilities are evaluated at sample average observed characteristics, and the difference between these two probabilities is interpreted as the marginal effect on current employment status due to the previous period's employment status.

Three other studies have estimated state dependence using the BHPS, and it is natural to compare the results here with results from those studies. It is also desirable, for purposes of clarity, to apply Equation (4.20) to the estimated parameters of interest in those studies to be able to make a more direct comparison. In other words, following Chamberlain (1984), marginal effect on current employment status due to the previous period's employment status is our metric of comparison.

Table 4.13 shows the results of these comparisons.

We explain our results first, which are based on BHPS data over 1991-2000. Using the RE-RF model with initial conditions estimated via reduced form, but with no correlation allowed between unobserved heterogeneity and observed characteristics, an individual with average observed characteristics who was employed one year ago has a 99% probability of currently being employed. The probability of currently being employed for such an individual falls to 93% if he was unemployed one year ago. Thus, the estimated marginal effect of previous employment upon current employment is the difference between the two probabilities, which is 6 percentage points.

Stewart (2007) uses the random effects probit model, with initial conditions estimated via reduced form, on BHPS data over 1991-1996. He estimates a main marginal effect of 15 percentage points using both men and women in an unbalanced panel.³⁴

Arulampalam (2004) uses the random effects probit model, with initial conditions estimated via reduced form, on BHPS data over 1991-1997. Her results are divided into four

³⁴Wage stratification is performed, so Stewart (2007) actually estimates two marginal effects - one of 15 percentage points (for high wage employment) and another of only 2 percentage points (for low wage employment). However, he focusses more upon the former. He also distinguishes between a single spell of unemployment and multiple spells of unemployment, which is not an issue we have addressed in our study.

Table 4.13: Estimated Marginal Effects by Studies

STUDY	PROB STAYEMP (I)	PROB NEWEMP (II)	MARGINAL EFFECT (III)	REF. (IV)	MODEL (V)
K2011	0.99	0.93	0.06	Table 4 Column 1	RE Logit
S2007	0.97	0.82	0.15	Table III Column 2	RE Probit
A2004 Minimum	—	—	0.40	Table 3 Column 1	RE Probit
A2004 Maximum	—	—	0.70	Table 3 Column 1	RE Probit
A2000 Minimum	0.98	0.90	0.08	Table 4 Row 8	RE Probit
A2000 Maximum	0.96	0.73	0.23	Table 4 Row 8	RE Probit
H2004	—	—	0.28	Table IV Column 4	RE Probit

Notes:

Column I shows $\Pr(\text{Employed at time } t \mid \text{Employed at time } t - 1)$

Column II shows $\Pr(\text{Employed at time } t \mid \text{Unemployed at time } t - 1)$

Column III (the extent of State Dependence) is calculated as $\Pr(\text{StayEmp}) - \Pr(\text{NewEmp})$

K2011 = this study. Results from RE-RF model estimated on main sample.

S2007 = Stewart (2007)

A2004 = Arulampalam (2004)

A2000 = Arulampalam et al (2000)

H2004 = Henley (2004)

The words 'Minimum' or 'Maximum' refer to the estimated marginal effect. For example, the second row shows the minimum marginal effect estimated by Arulampalam (2004) was 0.40.

subsamples depending upon whether, at the time of each survey, an individual is younger than 25 years and whether the individual has an educational qualification of O-level (i.e. obtained at or around the minimum school-leaving age) or higher qualification. Therefore, she reports four estimated marginal effects, ranging from 40 percentage points (25 and Over, O-level) to 70 percentage points (25 and Over, no O-level). These effects are much larger than the ones we estimate.

Arulampalam *et al* (2000) use the random effects probit model, with initial conditions estimated via reduced form, on BHPS data over 1991-1995. Their results are divided into eight sub-samples depending upon whether, at the time of each survey, an individual is younger than 25 years and which year between 1992-1995 the individual's current employment status is measured in (there is no previous year's observation for 1991 so it is not considered). Therefore, they report eight estimated marginal effects, ranging from 8 percentage points (Under 25, 1995) to 23 percentage points (25 and Over, 1992). At the lower end, these are similar to the marginal effects we estimate.

Henley (2004) uses the the random effects probit model, with initial conditions estimated via reduced form, on BHPS data over 1991-1999. His study specifically examines self-employment rather than employment *per se*. His results are divided into two subsamples depending upon whether information relating to housing tenure, real house value and monthly investment income are used to condition employment probabilities. The results across the two subsamples are similar – taking the subsample which uses the housing information, he estimates a marginal effect of 28 percentage points, which is considerably above the marginal effects we estimate.

It is useful to investigate what the sources of these discrepancies are. There are some differences between these other studies and ours, namely:

- Specification of baseline age and education;
- Definition of employment used; and,
- Estimation period.

Arulampalam (2004) divides the cohort of men into four groups, stratified by baseline age (i.e. at least 25 years of age in 1991 cf. less than 25 years), and by educational attainment (i.e. with at least one O-level qualification cf. with no O-levels). By comparison, we allow for the effect of age using a quadratic specification and education using a linear trend. This avoids the creation of even smaller sub-samples, which is important

Table 4.14: Effects of Age and Educational Partitioning upon Estimated Marginal Effects (Sample: T=1991-1995, except for Column VI where T=1991-2000)

	OVERALL	LOW ED	HIGH ED	YOUNGER	OLDER	MAIN SAMPLE
	(I)	(II)	(III)	(IV)	(V)	(VI)
COEFFICIENTS						
y_{t-1}	2.074 (.000)	2.003 (.000)	1.979 (.000)	1.027 (.009)	2.430 (.000)	2.361 (.000)
INDEX AND PROBABILITIES						
Sample Size	2054	1184	870	296	1758	1285
Employment index value	2.31	2.32	2.51	1.96	0.21	2.66
Pr(StayEmployed)	0.99	0.99	0.99	0.95	0.93	0.99
Pr(NewlyEmployed)	0.91	0.91	0.92	0.88	0.55	0.93
Marginal Effect	0.08	0.08	0.07	0.07	0.38	0.06

Notes:

Results from Random Effects Model with Reduced form Initial Conditions (RE-RF) reported throughout.

p-values based on the Normal distribution are in parentheses.

Employment index value is the value of the employment index $\sum_{j=1}^k \beta_j \bar{X}_j$, using estimated coefficients β_j (excluding coefficient on y_{t-1}) and evaluated at population average values.

Pr(StayEmployed) is $\Pr(\text{Employed at time } t \mid \text{Employed at time } t-1)$

Pr(NewlyEmployed) is $\Pr(\text{Employed at time } t \mid \text{Unemployed at time } t-1)$

Marginal Effect (the marginal effect of prior employment upon the probability of current employment) is $\Pr(\text{StayEmployed}) - \Pr(\text{NewlyEmployed})$

since parameter estimation only effectively takes into account the employment histories of a small fraction of the entire sample (see discussion in Section 4.4.3 above). To examine whether results are sensitive to these differences in specification, we present results from partitioning an auxiliary sample of 2054 men who were in the labour force for each wave during 1991-1995 by each of age and educational attainment, using the same rules as Arulampalam (2004). This sample of 2054 men is the union of auxiliary samples E1 (1285 men) and F1 (769 men) in Section 4.6.1 above. Results are reported in Table 4.14.

The estimated marginal effects of partitioning men according to educational attainment show that for those men with 12 years of education or less (1184 men in total), the estimated marginal effect using the RE-RF model is 8 percentage points. For those men with more than 12 years of education (870 men in total), the estimated marginal effect

is 7 percentage points. For this sample over 1991-1995 as a whole, the estimated marginal effect is 8 percentage points, which is similar to the estimated marginal effect of 6 percentage points for the main sample as a whole over 1991-2000. The results from partitioning according to educational attainment over 1991-1995 are almost identical to those reported for the main sample over 1991-2000.

However, the estimated marginal effects of partitioning men according to age are quite different. There are 296 "younger" men who were aged 24 years or younger in 1991, for whom we estimate a marginal effect of 7 percentage points. This compares with 1758 "older" men who were aged 25 years or older in 1991, for whom we estimate a much larger marginal effect of 38 percentage points, so we additionally report descriptive statistics for these two subsamples in Tables 4.15 and 4.16.

Table 4.15: Sample Characteristics, BHPS 1991-1995, among men younger than 25 years in 1991

	Full Sample (I)	Employed 10 years (II)	Employed 0 years (III)	Single Transition from Work (IV)	Single Transition to Work (V)	Multiple Transitions (VI)
Means and Standard Deviations						
Employment (1=Employed)	0.85 (0.36)	1.00 (0.00)	0.00 (0.00)	0.32 (0.48)	0.75 (0.43)	0.64 (0.48)
Age/100 (1991)	0.21 (0.02)	0.21 (0.02)	0.20 (0.02)	0.20 (0.03)	0.20 (0.03)	0.20 (0.02)
Marital (1=Cohabiting)	0.36 (0.50)	0.39 (0.52)	0.50 (0.51)	0.36 (0.49)	0.18 (0.39)	0.31 (0.46)
Children/10	0.05 (0.09)	0.04 (0.07)	0.15 (0.13)	0.08 (0.07)	0.05 (0.07)	0.08 (0.11)
Years of Education/100	0.12 (0.02)	0.12 (0.02)	0.10 (0.01)	0.11 (0.01)	0.12 (0.02)	0.12 (0.01)
Health (1 = Good Health)	0.97 (0.17)	0.97 (0.17)	0.98 (0.14)	1.00 (0.00)	0.98 (0.12)	0.96 (0.20)
Housing (1 = Own House)	0.70 (0.46)	0.75 (0.43)	0.14 (0.35)	0.76 (0.44)	0.69 (0.46)	0.65 (0.48)
Region (1 = South England)	0.33 (0.47)	0.36 (0.48)	0.30 (0.46)	0.20 (0.41)	0.32 (0.47)	0.25 (0.43)
Sample Size	296	188	10	5	26	67

Notes:

Standard deviations are in parentheses.

Variable definitions are contained in Appendix C.1.

Table 4.16: Sample Characteristics, BHPS 1991-1995, among men 25 years or older in 1991

	Full Sample (I)	Employed 10 years (II)	Employed 0 years (III)	Single Transition from Work (IV)	Single Transition to Work (V)	Multiple Transitions (VI)
Means and Standard Deviations						
Employment (1=Employed)	0.94 (0.23)	1.00 (0.00)	0.00 (0.00)	0.29 (0.46)	0.75 (0.43)	0.67 (0.47)
Age/100 (1991)	0.40 (0.10)	0.40 (0.10)	0.39 (0.09)	0.41 (0.07)	0.38 (0.10)	0.39 (0.10)
Marital (1=Cohabiting)	0.84 (0.37)	0.85 (0.36)	0.65 (0.48)	0.64 (0.48)	0.82 (0.39)	0.81 (0.39)
Children/10	0.09 (0.11)	0.09 (0.11)	0.15 (0.15)	0.11 (0.10)	0.10 (0.10)	0.10 (0.12)
Years of Education/100	0.12 (0.02)	0.12 (0.02)	0.11 (0.02)	0.12 (0.03)	0.11 (0.02)	0.11 (0.02)
Health (1 = Good Health)	0.97 (0.17)	0.97 (0.17)	0.96 (0.19)	0.98 (0.14)	0.99 (0.08)	0.96 (0.19)
Housing (1 = Own House)	0.83 (0.38)	0.86 (0.35)	0.16 (0.37)	0.60 (0.49)	0.67 (0.47)	0.74 (0.44)
Region (1 = South England)	0.41 (0.49)	0.40 (0.49)	0.41 (0.49)	0.36 (0.48)	0.47 (0.50)	0.46 (0.50)
Sample Size	1758	1511	22	19	59	147

Notes:

Standard deviations are in parentheses.

Variable definitions are contained in Appendix C.1.

The most noteworthy difference across these subsamples is the considerably higher employment rate among older men (94% cf. 85%). However, also note that the percentage of men making transitions (i.e. those men in columns IV, V and VI) accounts for a much larger percentage of the subsample for the younger men, compared to the sample of older men (33% cf. 13%).³⁵

These auxiliary results explain some of the differences between the results we report here and those from other studies, although even so the marginal effects estimated by Arulampalam (2004) are between 40 to 70 percentage points, which are higher than any we estimate in this section. Although there are considerably more older men than younger men in the sample, since transitions are more common among younger men than older men and the size of estimated marginal effects depends crucially upon information from those making transitions³⁶, this may explain why the overall marginal effect of 8 percentage points is very close to the estimated marginal effect for younger men of 7 percentage points since results are mainly driven by observations on these men. If persistence is much stronger among older men, this may also partly explain why we estimate higher marginal effects over the later period in Section 4.6.2 above, since men are 5 years older in 1996-2000 than they were in 1991-1995.

The higher estimated marginal effects for older men are also much closer to the results from linear modelling. From Table 4.10, we can see that the most plausible linear specification estimates a marginal effect of 32 percentage points, which is not dissimilar to the 38 percentage point marginal effect for older men considered here.

Definition of employment used

Stewart (2007), Arulampalam (2004) and Arulampalam *et al* (2000) all define employment as wage or salary earning, by contrast with this study which includes self-employed as well as wage and salary earners and with Henley (2004) who focusses exclusively upon the self-employed. This may have an impact upon results if the behaviour of the self-employed is very different from the behaviour of wage and salary earners. However, both Henley (2004) and Arulampalam (2004) estimate very high marginal effects (i.e. above 25 percentage points) despite the differences between their definitions of employment. Also

³⁵Unsurprisingly, there is also a considerably higher percentage of older men who are cohabiting compared with younger men, and older men tend to have more children, higher rates of house ownership and a higher likelihood of living in the south of England.

³⁶As discussed in Section 4.4.3, above.

Arulampalam (2004) estimates much higher marginal effects than Stewart (2007), despite both excluding the self-employed from the employment category. Our decision to include the self-employed is based upon increasing the sample size while at the same time controlling for the possibility of endogenous. Based upon the evidence discussed in this paragraph from different studies, it appears that other factors (e.g. age, gender) may have a larger impact upon results than the definition of employment status.

Estimation period

To examine whether results are sensitive to estimation period, we separately estimate results from the subsample in Section 4.6.3 above, which is estimated on BHPS data over 1991-1995. Specifically, the selection rule used is to include all individuals who have recorded observations for all relevant variables in each of the first five waves of the BHPS.

Because Arulampalam *et al* (2000) use an unbalanced panel, they have 2524 individuals in their sample, while by comparison this balanced panel has 2054 individuals. However they note (p.27) that they also estimated the models on a balanced panel with unaltered conclusions, so this is a valid comparison.

On this sample, we estimate the state dependence marginal effect to be 7 percentage points, which is close to the 6 percentage point effect estimated on the main sample over 10 years. Therefore, we conclude that estimation period does not explain the discrepancies between their study and ours. However, our results are the same order of magnitude as those of Stewart (2007), who uses a similar period of 1991-1996.

4.7 Conclusion

This Chapter adds to the body of literature which finds that state dependence (i.e. persistence) in male unemployment is a real phenomenon. We use data for UK men from the first 10 waves of the British Household Panel Survey over the period 1991-2000. The results suggest that men continually in the labour force who are currently employed have around a 99% likelihood of also being employed a year into the future, while for men currently unemployed future employment likelihood is around 91% to 93%. This represents a significant difference, although not so substantial a difference as some other studies have reported. These differences may be partly due to state dependence in unemployment be-

ing stronger among older men, and to this extent other studies which look separately at persistence among different age groups may report different results to ours.

Methodologically, we focussed upon ensuring consistent estimation of the marginal effect of past employment status upon current employment likelihood. All of the estimation methods considered show that this marginal effect is positive; that is, that prior employment increases the likelihood of currently being employed. The evidence that state dependence exists is robust to everything we control for, and in particular robust to allowing for the influence of unobserved heterogeneity between individuals. The conclusions we reach are that unemployment among men is explained partly by heterogeneity - both observed and unobserved heterogeneity - and partly by true state dependence.

However, the extent of state dependence is less clear. All specifications which model labour market disturbances using a nonlinear process indicate that marginal effects are no higher than 10 percentage points. However, all linear specifications considered estimate much higher marginal effects, and most linear methods estimate marginal effects of around 30 percentage points.

Evidence both from this Chapter and other studies suggests that persistence varies over workers' employment lives. The employment prospects of younger workers tend to be less affected by their recent employment status, compared with older workers who may be more likely to be subject to persistence in employment status.

Among other findings, we note that allowing for unobserved heterogeneity often involves specifying to what extent unobserved differences between individuals are related to observed differences. For our main purpose of estimating marginal effects due to prior employment status, relationships between observed and unobserved heterogeneity are not sufficiently strong to conclude that one particular specification stands out strongly as the most credible. However, were the focus of the study the precise estimation of employment effects from observed characteristics, such as the ones we control for in this Chapter, there is evidence that results would be more sensitive to considering the nature of relationships between observed characteristics and unobserved effects.

Chapter 5

Conclusion

As stated in the Introduction, the primary aim of the thesis is to shed light upon the extent and causes of unemployment persistence.

At an empirical level, it is an already-established regularity that unemployment persistence exists. The empirical results reported here do not challenge this consensus within the literature, for the key conclusion emerging from each of Chapter 2 and Chapter 4 is that individual employment histories have predictive value for the likelihood of currently being employed, even after controlling for observed and unobserved differences between individuals.

However, our empirical results contain some new findings concerning the extent of measured persistence. The results from estimation using standard Heckman-type techniques indicate relatively low extents of persistence, where the employment prospects of an individual with sample mean characteristics who has been recently employed are between 2 and 10 percentage points higher than for a similar individual who has been recently unemployed. By comparison with most other studies which have investigated the issue, these are relatively small state dependence effects. The Australian data set confirms that these results are not just restricted to men.

The likely explanation for these findings relates to age of respondents, since our robustness analysis in Chapter 4 reveals that persistence is stronger among younger men. Since the Australian data set used in Chapter 2 only contains respondents aged up to 29 years, it is then less surprising that we measure a relatively small amount of persistence

from it.

The key empirical contribution made by the thesis is therefore not the finding that unemployment persistence exists, but that its measured extent depends heavily upon how it is modelled. We show that when using a linear model of employment status we are able to estimate results for persistence of magnitudes similar to that estimated by several other studies. We also show that our results from linear modelling are robust to several plausible specifications. It is not yet clear which specification is most likely to be appropriate, or indeed whether appropriate specifications depend upon country and/or time period. However, the results reported here point to the need to pursue this issue more widely.

So at an empirical level we have shown that unemployment persistence is a complicated phenomenon. This is also true at a theoretical level, but our response there is different. For our empirical analysis, we have emphasised the need to examine sensitivity of results to estimation method used. By contrast, the theoretical models we put forward in Chapter 3 have been deliberately simplified. This is most obvious not only through the fixed-wage assumption which we have maintained throughout, but also by exploiting the restrictive assumption of Leontief technology which removes the firm's ability to substitute between labour and capital.

The key theoretic contribution made by the thesis is that we have shown it is possible to construct a tractable explanation for unemployment persistence based upon incomplete information about productivity, and involving rational hiring behaviour by the firm. This fills a gap in the literature, and provides a foundation for further generalisation.

Our results raise the question of how far our theoretical framework can be extended in its current form. If we acknowledge that, ultimately, an explanation for persistence which relies on a fixed-wage firm with fixed-coefficients technology is unlikely to be very convincing, then developing a more general approach will be of great importance. However, the basic theoretical model used is already intractable to the extent that results from both our extended models rely heavily on tabulations. The question remains: can this be improved?

We suggest there are two major directions in which the theoretical work may be usefully extended. First, it would seem imperative to investigate refinements which enable the model to make less sweeping predictions about the extent of persistence. For example, in our model even with exogenous separations the new workers have the same average characteristics as the incumbents in the labour force. However, Pryor and Schaffer

(1999) among others suggest that in fact younger workers (entrants) typically have more skills than those they replace in the workforce. Therefore, in our context, this may be an effect which actually causes the composition of the unemployed pool to *improve*, which would offset against the endogenous worsening of the unemployed pool's composition in the models presented here.

Second, since the theory as set out in Chapter 3 does not predict for certain whether persistence will result, then it would be helpful to devise a test capable of falsifying the theory. Falsification could only arise if it were shown - *under the maintained assumptions of the theory* - that employment outcomes were merely transitory. This sort of evidence is unlikely to arise; so, the hypothesis of information-driven persistence may be difficult either to prove or disprove. However, if we seek to match the theory with our empirical work closely, it is noteworthy that the theoretical models developed here often (i.e. for many configurations of parameter values) either predict complete persistence or no persistence at all. Since none of our empirical estimates even remotely resemble complete persistence (i.e. $\delta = 1$), then it would seem that the theory as it stands is still not capable of standing up against rigorous empirical testing.

Ultimately, the reason for pursuing a very restrictive specification in the theoretical analysis lies in its tractability. Although this is a limitation, the tractability of the model in Chapter 3 also indicates its key theoretic contribution to the literature.

The third and final area in which the thesis makes a contribution is methodological. We set out in Chapter 4 a framework for modelling different stochastic processes which drive binary employment outcomes. As part of this, we put forward as a reference work a wide-ranging survey of estimation techniques from the empirical literature. We then show, using the example of modelling unemployment persistence using data from the BHPS, how an empirical investigation into the determinants of binary outcomes may be conducted using panel data. A crucial part of our methodological contribution is formed by robustness analysis and discussion of model selection, as these demonstrate how practical issues may be addressed in the context of applied panel data econometrics.

However, the methodological focus is wider than this, and is intertwined through all chapters of the thesis. In Chapter 2, we motivate the probabilistic approach which underpins our empirical model by pointing out that the natural alternative of attempting to model labour demand and labour supply simultaneously is infeasible due to problems with under-identification of independent demand and supply relationships. In Chapter 3, we discuss at length the differences between our modelling approach and others used in the

literature, and point out that our model is equivalent to that of Montgomery (1999) for particular parameter configurations.

It is convenient to note that Chapters 2 and 4 report empirical results and Chapter 3 develops theoretical results, but there is an overarching theme to this work. The thesis starts from and finishes with the position of bringing together, and endeavouring to reconcile in detail, several methodological issues. All of these issues arise in the course of explaining, and measuring, the empirical regularity of serial correlation in employment outcomes.

Bibliography

- [1] Akerlof, G. (1970). "The Market for Lemons: Qualitative Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, 84, pp.488-500.
- [2] Alonso-Borrego, C. and Arellano, M. (1996). "Symmetrically Normalised Instrumental-Variable Estimation Using Panel Data," *Journal of Business and Economic Statistics*, 14, pp.36-49.
- [3] Amemiya, T. (1981). "Qualitative Response Models: A Survey," *Journal of Economic Literature*, 19, pp.481-536.
- [4] Anderson, T. and Hsiao, C. (1981). "Estimation of Dynamic Models with Error Components," *Journal of the American Statistical Association*, 76, pp.598-606.
- [5] Arellano, M. and Bond, S. (1991). "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, pp.277-297.
- [6] Arulampalam, W. (1998). "State Dependence in Unemployment Incidence: Evidence for British Men 1991-1995," *mimeo*, University of Warwick.
- [7] Arulampalam, W. (2000). "Unemployment-Non-employment Dynamics: Evidence for British men," *mimeo*, University of Warwick.
- [8] Arulampalam, W. (2004). "Unemployment-Non-employment Dynamics: Evidence for British men revisited," *mimeo*, University of Warwick.
- [9] Arulampalam, W., Booth, A. and Taylor, M. (2000). "Unemployment Persistence," *Oxford Economic Papers*, 52, pp.24-50.

- [10] Arulampalam, W., Gregg, P. and Gregory, M. (2001). "Unemployment Scarring," *Economic Journal*, 111, pp.F577-584.
- [11] Arulampalam, W. and Stewart, M. (2009). "Simplified Implementation of the Heckman Estimator of the Dynamic Probit Model and a Comparison with Alternative Estimators," *Oxford Bulletin of Economics and Statistics*, 71, pp. 659-681.
- [12] Aungles, P. and Stewart, B. (1986). "The Duration of Unemployment of CES Registered Job-seekers: An Econometric Analysis", *Bureau of Labour Market Research Working Paper No. 66*.
- [13] Blundell, R. and Bond, S. (1998). "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87, pp.115-143.
- [14] Bond, S. (2002). "Dynamic Panel Data Models: A Guide to Micro Data Methods and Practice", *Portugese Economic Journal*, 1, pp. 141-162.
- [15] Booth, A., Jenkins, S. and Garcia Serrano, C. (1999). "New Men and New Women? A Comparison of Paid Work Propensities from a Panel Data Perspective," *Oxford Bulletin of Economics and Statistics*, 61(2), pp.167-198.
- [16] Bradley, S., Crouchley, R. and Oskrochi, R. (2003). "Social Exclusion and Labour Market Transitions: a Multi-State Multi-Spell Analysis using the BHPS," *Labour Economics*, 10, pp.659-679.
- [17] Brooks, C. and Volker, P. (1985). "The Influence of Unemployment Duration and Heterogeneity on the Transition from Unemployment for Australian Youth," *Bureau of Labour Market Research Working Paper No. 55*.
- [18] Buddelmeyer, H., Lee, W. and Wooden, M. (2010). "Low-Paid Employment and Unemployment Dynamics in Australia," *Economic Record*, 86, pp. 28-48.
- [19] Bureau of Labour Market Research (1986). "The First Wave of the ALS: Facts and Figures about Young CES Registrants," *Monograph Series No. 15, Bureau of Labour Market Research*, Canberra: AGPS.
- [20] Burgess, S. and Turon, H. (2005). "Unemployment Dynamics in Britain," *Economic Journal*, 115, pp.423-448.

- [21] Butler, J. and Moffitt, R. (1982). "A Computationally Efficient Quadrature Procedure for the One Factor Multinomial Probit Model," *Econometrica*, 50, pp.761-764.
- [22] Cappellari, L., Dorsett, R. and Haile, G. (2010). "State Dependence and Unobserved Heterogeneity in the Employment Transitions of the Over-50s," *Empirical Economics*, 38, pp.523-554.
- [23] Chamberlain, G. (1984). "Panel Data," in (Z. Griliches and M. Intrilligator, eds.), *Handbook of Econometrics, Volume II*, Amsterdam: North-Holland.
- [24] Chamberlain, G. (1985). "Heterogeneity, Omitted Variables Bias, and Duration Dependence," in (J. Heckman and B. Singer, eds.), *Longitudinal Analysis of Labor Market Data*, Cambridge University Press.
- [25] Chay, K. and Hyslop, D. (2000). "Identification and Estimation of Dynamic Binary Response Panel Data Models: Empirical Evidence Using Alternative Approaches," *mimeo*, University of California, Berkeley.
- [26] Clasen, J., Davidson, J., Ganßmann, H. and Mauer, A. (2004). "Non-Employment and the Welfare state: UK and Germany Compared," *mimeo*, Anglo-German Foundation for the Study of Industrial Society.
- [27] Cripps, T. and Tarling, R. (1974). "An Analysis of the Duration of Male Unemployment in Great Britain, 1932-1973," *Economic Journal*, 84, pp.289-316.
- [28] Department of Employment, Education and Training (1990). *Australian Longitudinal Survey, 1985-1988: A Research Bibliography*, Canberra: AGPS.
- [29] Doornik, J., Arellano, M. and Bond, S. (2002). "Panel Data Estimation using DPD for Ox," *mimeo*, University of Oxford.
- [30] Efron, B. (1979). "Bootstrap Methods: Another Look at the Jackknife," *Annals of Statistics*, 7, pp.1-26.
- [31] Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*, Chapman and Hall.
- [32] Ellwood, D. (1982). "Teenage Unemployment: Permanent Scars of Temporary Blemishes?" in *The Youth Labor Market Problem: Its Nature, Causes and Consequences*, (eds. R.B. Freeman and D. Wise), pp. 349-385, Chicago, NBER.

- [33] 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-Neilsen eds.), *Panel Data and Labour Market Dynamics*, Elsevier Science Publishers.
- [34] Greene, W. (2000). *Econometric Analysis*, 4th edition, New Jersey: Prentice-Hall.
- [35] Guasch, J. and Weiss, A. (1980). "Wages as Sorting Mechanisms in Competitive Markets with Asymmetric Information: A Theory of Testing," *Review of Economic Studies*, 47, pp.653-664.
- [36] Hansen, L. (1982). "Large Sample Properties of Generalised Method of Moment Estimators," *Econometrica*, 50, pp.1029-1054.
- [37] Harris, M.N. (1993). "The Australian Longitudinal Survey: A Reference Document," *memo*, Department of Econometrics, Monash University.
- [38] Harris, M.N. (1996). "Modelling the Probability of Youth Unemployment in Australia," *Economic Record*, 72, pp. 118-129.
- [39] Hausman, J. (1978). "Specification Tests in Econometrics," *Econometrica*, 46, pp.1251-1371.
- [40] Heath, A and Swann, T. (1999). "Reservation Wages and the Duration of Unemployment," *Reserve Bank of Australia Research Discussion Paper 1999-02*.
- [41] Heckman, J. (1981a). "Statistical Models for Discrete Panel Data," in (C. Manski and D. McFadden, eds.), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press.
- [42] Heckman, J. (1981b). "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process," in (C. Manski and D. McFadden, eds.), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press.
- [43] Heckman, J. (1981c), "Heterogeneity and State Dependence," in *Studies in Labor Markets*, (ed. S. Rosen), pp.91-139, Chicago: University of Chicago Press.
- [44] Heckman, J. (1991). "Identifying the Hand of Past: Distinguishing State Dependence from Heterogeneity," *American Economic Review*, 81, pp.75-79.

- [45] Heckman, J. (1995), "Lessons from the Bell Curve", *Journal of Political Economy*, 103(5), pp.1091-1120.
- [46] Heckman, J. and Borjas, G. (1980), "Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence," *Economica*, 47, pp. 247-283.
- [47] Henley, A. (2004). "Self-Employment Status: The Role of State Dependence and Initial Circumstances," *Small Business Economics*, 22(1), pp.67-82.
- [48] Honore, B. and Kyriazidou, E. (2000). "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica*, 68, pp.839-874.
- [49] Hosios, A.(1990). "On the Efficiency of Matching and Related Models of Search and Unemployment," *Review of Economic Studies*, 57, pp.279-298.
- [50] Hotz, V.J., Kydland, F. and Sedlacek, G. (1988). "Intertemporal Preferences and Labor Supply," *Econometrica*, 56, pp.335-360.
- [51] Hsiao, C. (2003). *Analysis of Panel Data*, 2nd edition, Cambridge University Press.
- [52] Hui, W. (1986). "State Dependence and Youth Unemployment in Australia: A Survival Analysis," *ANU Working Papers in Economics and Econometrics No. 138*.
- [53] Hyslop, D. (1999). "Serial Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women," *Econometrica*, 67, pp.1255-1294.
- [54] Inkermann, J. (2000). "Misspecified Heteroskedasticity in the Panel Probit Model: A Small Sample Comparison of GMM and SML Estimators," *Journal of Econometrics*, 97(2), August, pp.227-59.
- [55] Jackman, R. and Layard, R. (1990). "Does Long-term Unemployment Reduce a Person's Chance of a Job? A Time-series Test," *Economica*, 58, pp.93-106.
- [56] Johnston, J. (1984). *Econometric Methods*, 3rd edition, New York: McGraw-Hill.
- [57] Keller, G., Roberts, K. and Stevens, M. (2007). "Unemployment, Participation and Market Size," *University of Oxford Department of Economics Discussion Paper Series*, 362, October.

- [58] Knights, S., Harris, M. and Loundes, J. (2002). "Dynamic Relationships in the Australian Labour Market: Heterogeneity and State Dependence," *Economic Record*, 78(3), pp.284-298.
- [59] Lancaster, T. (1990). *The Econometric Analysis of Transition Data*, Cambridge, UK: Cambridge University Press.
- [60] Lawless, J. (1987). "Negative Binomial and Mixed Poisson Regression," *Canadian Journal of Economics*, 15, pp.209-225.
- [61] Layard, R. and Nickell, S. (1985). "The Causes of British Unemployment," *National Institute Economic Review*, 0, pp. 62-85.
- [62] Layard, R., Nickell, S. and Jackman, R. (2005). *Unemployment: Macroeconomic Performance and the Labour Market*, 2nd edition, Oxford University Press.
- [63] Lazear, E. (1999). "Output-based Pay: Incentives or Sorting?" *NBER Working Paper 7419*.
- [64] Le, A. and Miller, P. (2001). "Is a Risk Index Approach to Unemployment Possible?" *Economic Record*, 77(1), pp.51-70.
- [65] Lindbeck, A. and Snower, D. (1986). "Wage Setting, Unemployment, and Insider-Outsider Relations," *American Economic Review*, 76, pp.235-239.
- [66] Lindbeck, A. and Snower, D. (1988). "Cooperation, Harassment, and Involuntary Unemployment: An Insider-Outsider Approach," *American Economic Review*, 78, pp.167-188.
- [67] MacLeod, W. and Malcomson, J. (1988). "Reputation and Hierarchy in Dynamic Models of Employment," *Journal of Political Economy*, 96, pp.832-854.
- [68] Manski, C. (1987). "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, 55, pp.357-362.
- [69] McRae, I. (1984). "The Australian Longitudinal Survey," *BLMR Conference Paper No. 44*, Canberra: AGPS.
- [70] Miller, P. (1989). "The Structure of Aboriginal and Non-Aboriginal Youth Unemployment", *Australian Economic Papers*, 28(52).

- [71] Montgomery, J. (1991). "Social Networks and Labor-Market Outcomes: Toward an Economic Analysis," *American Economic Review*, 81, pp.1408-1418.
- [72] Montgomery, J. (1999). "Adverse Selection and Employment Cycles," *Journal of Labor Economics*, 17, pp.281-297.
- [73] Mortensen, D. (1989). "The Persistence and Indeterminacy of Unemployment in Search Equilibrium," *Scandinavian Journal of Economics*, 91, pp.347-370.
- [74] Muhleisen, M. and Zimmermann, K. (1994). "A Panel Analysis of Job Changes and Unemployment," *European Economic Review*, 38, pp.793-801.
- [75] Mundlak, Y. (1978). "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 46, pp.69-85.
- [76] Narendranathan, W., Nickell, S. and Stern, J. (1985). "Unemployment Benefits Revisited," *Economic Journal*, 95, pp.307-329.
- [77] 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, pp.161-185.
- [78] Nickell, S. (1981). "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 47, pp.1249-1266.
- [79] Olsen, C. (1996). "Unemployment Duration in the United States: An Application of a Competing Risks Model," *Master of Commerce Research Report, University of Melbourne*.
- [80] Ong, P. and Mar, D. (1992). "Post-Layoff Earnings among Semiconductor Workers," *Industrial and Labor Relations Review*, 45, pp.366-379.
- [81] Orme, C. (1997). "The Initial Conditions Problem and Two-Step Estimation in Discrete Panel data Models," *mimeo*, University of Manchester.
- [82] Paull, G. (1996). "Dynamic Labour Market Behaviour in the British Household Panel Survey: the Effects of Recall Bias and Panel Attrition," *mimeo*, Institute for Fiscal Studies.
- [83] Paull, G. (2002). "Biases in the Reporting of Labour Market Dynamics," *IFS Working Paper Series*, WP02/10, London: Institute for Fiscal Studies.

- [84] Phelps, E. (1972). *Inflation Policy and Unemployment Theory: The Cost Benefit Approach to Monetary Planning*, Macmillan.
- [85] Pissarides, C. (1991). "Real Wages and Unemployment in Australia," *Economica*, 58(1), pp. 35-55.
- [86] Pryor, F. and Schaffer, D. (1999). *Who's not Working and Why*, Cambridge University Press.
- [87] Raaum, O. and Røed, K. (2006). "Do Business Cycle Conditions at the Time of Labor Market Entry Affect Future Employment Prospects?" *The Review of Economics and Statistics*, 88, pp.193-210.
- [88] Rodriguez-Planas, N. (2009). "A Signaling Model of Temporary Layoffs," *Oxford Economic Papers*, 61, pp.566-585.
- [89] Rudin, W. (1976). *Principles of Mathematical Analysis*, 3rd edition, New York: McGraw-Hill.
- [90] Schoeni, R. (1995). "Marital Status and Earnings in Developed Countries," *Journal Of Population Economics*, 8(4), pp.351-359.
- [91] Shapiro, C. and Stiglitz, J. (1984). "Equilibrium Unemployment as a Worker Discipline Device," *American Economic Review*, 74(3), pp. 433-444.
- [92] Silcock, H. (1954). "The Phenomenon of Labor Turnover," *Journal of the Royal Statistical Society*, 117, pp.429-440.
- [93] Spence, M. (1973). "Job Market Signaling," *Quarterly Journal of Economics*, 77, pp.355-374.
- [94] Staiger, D. and Stock, J. (1997). "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, pp.557-586.
- [95] Stewart, M. (2007). "The Interrelated Dynamics of Unemployment and Low-wage Employment," *Journal of Applied Econometrics*, 22, pp. 511-531.
- [96] Stromback, T., Dockery, M., and Ying, W. (1998). "Transitions in the Labour Market: Evidence from the Survey of Employment and Unemployment Patterns," *Melbourne Institute Working Paper Series No. 1/98*.

- [97] Taylor, M. (1994). *British Household Panel Survey user manual*, Economic and Social Research Council, UK.
- [98] Trivedi, P. and Alexander, J. (1988). "Incorporating International Competitiveness into the demand for Labour Function: Some Issues of Specification and Interpretation," *Economic Record*, 64(186), pp.196-208.
- [99] Trognon, A. (1978). "Miscellaneous Asymptotic Properties of Ordinary Least Squares and Maximum Likelihood Estimators in Dynamic Error Components Models," *Annales de L'INSEE*, 30-1, pp.631-657.
- [100] Van Den Berg, G. and Van Ours, J. (1994). "Unemployment Dynamics and Duration Dependence in France, the Netherlands and the United Kingdom," *Economic Journal*, 104, pp.432-443.
- [101] Verbeek, M. and Nijman, T. (1996). "Incomplete Panels and Selection Bias," in (L. Matyas and P. Sevestre, eds.), *The Econometrics of Panel Data : A Handbook of the Theory With Applications*, 2nd edition, Kluwer Academic Publishers.
- [102] White, H. (1980). "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, pp.817-838.
- [103] Windmeijer, F. (2005). "A Finite Sample Correction for the Variance of Linear Efficient Two-step GMM Estimators," *Journal of Econometrics*, 126, pp.25-51.
- [104] Zaretsky, A. and Coughlin, C. (1995). "An Introduction to the Theory and Estimation of a Job-Search Model," *Federal Reserve Bank of St-Louis Review*, 77(1), pp. 53-65.

Appendix A

Appendix for Chapter 2

A.1 Data Appendix

Descriptions of variables used

All variables are, either directly or indirectly, from the Australian Longitudinal Survey (ALS) over the period 1985-1988.

Dependent Variable: Employment Status

A binary variable was created.

Employment Status = 0 denotes respondent was unemployed as at time of interview.

Employment Status = 1 denotes respondent was employed as at time of interview.

Those absent for the full duration of the sample 1985-1988 were treated as missing values. Those present for part of the sample had values recorded for those periods and missing values recorded for periods spent outside the labour force. Thus, all individuals who are counted in a given period were either in employment or actively seeking work during that period.

Explanatory Variables: Overall Model

Constant

Estimated constant term.

Employment Status in previous period

As for Employment Status.

Experience

Defined as period elapsed between time of first entry into labour market and time of current survey. For females, this calculation was subsequently weighted according to number of children [using weights inferred from Harris (1996)].

Marital Status & Separated

2 binary variables were created to denote marital status.

The default category (i.e. Marital Status = Separated = 0) denotes single.

Marital Status = 1 denotes married.

Separated = 1 denotes separated, divorced, or widowed.

City Dwelling & Rural Dwelling

2 binary variables were created to denote place of residence.

The default category (i.e. City Dwelling = Rural Dwelling = 0) denotes dwelling in country town.

City Dwelling = 1 denotes dwelling in city.

Rural Dwelling = 1 denotes dwelling in rural area which is not a country town.

Buying House & Renting/Boarding House & Rent Free House

3 binary variables were created to denote financial commitments related to dwelling.

The default category (i.e. Buying House = Renting/Boarding House = Rent Free House = 0) denotes ownership of house.

Buying House = 1 denotes buying house.

Renting House = 1 denotes renting house or boarding.

Rent Free House = 1 denotes living in rent free accommodation or already owning house.

Year 10 & Year 11

2 binary variables were created to denote highest level of education for low education respondents (i.e. those respondents with a level of education no higher than completion of Year 11)

The default category (i.e. Year 10 = Year 11 = 0) denotes lower than completion of Year 10.

Year 10 = 1 denotes completion of Year 10.

Year 11 = 1 denotes completion of Year 11.

Year 12 & Trade Qualification & Diploma & Degree

4 binary variables were created to denote highest level of education for high education respondents (i.e. those respondents with a level of education no lower than completion

of Year 11).

The default category (i.e. Year 12 = Trade Qualification = Diploma = Degree = 0) denotes completion of other qualification.

Year 12 = 1 denotes completion of Year 12.

Trade Qualification = 1 denotes completion of trade qualification.

Diploma = 1 denotes completion of diploma or certificate from TAFE or business college.

Degree = 1 denotes completion of degree.

Partner's Employment Status

As for Employment Status. The binary variable was only recorded for respondents with Marital Status = 1, with missing values recorded for other respondents.

Health Status

A binary variable (Health Status = 0 denotes no disability) was created.

Health Status = 1 denotes respondent disabled insofar as to limit the amount or type of work they could perform.

Number of Children

Defined as respondent's number of children.

Average Experience

Defined as the arithmetic mean of the variable Experience for the respondent's (gender and education) subgroup.

Replacement Ratio

Defined as the ratio $\frac{Unemployment\ benefits}{Average\ weekly\ wage_i}$, $i = 1, \dots, 4$

$i = 1$: High education females

$i = 2$: high education males

$i = 3$: low education females

$i = 4$: low education males

Explanatory Variables: Initial Conditions Model

Constant

As for overall model.

Experience

As for overall model.

Marital Status & Separated

As for overall model.

City Dwelling & Rural Dwelling

As for overall model.

Buying House & Renting or Boarding House & Rent Free House

As for overall model.

Year 10 & Year 11

As for overall model.

Year 12 & Trade Qualification & Diploma & Degree

As for overall model.

Partner's Employment Status

As for overall model.

Health Status

As for overall model.

Number of Children

As for overall model.

Western origin

A binary variable was created.

Western = 1 denotes of western racial origin.

Western = 0 denotes of aboriginal, Asian, or other non-western racial origin.

State school

A binary variable was created.

State School = 0 denotes attendance at state or catholic school.

State School = 1 denotes attendance at other type of school.

Manufacturing industry

A binary variable was created.

Manufacturing Industry = 1 denotes first job was in manufacturing industry.

Both parents

A binary variable was created.

Both parents = 1 denotes both parents were present at age 14.

Less than six months & Six months

2 binary variables were created to denote length of search time before finding first job.

The default category (i.e. Less than six months = Six months = 0) denotes length of search time before finding first job was less than two weeks.

Less than six months = 1 denotes length of search time before finding first job was between 2 weeks and six months.

Six months = 1 denotes length of search time before finding first job was more than six months.

Employment History & Unemployment History

2 binary variables were created to denote the respondent's employment status at survey times.

The default category (i.e. Employment history = Unemployment history = 0) denotes respondent was employed in as many surveys as unemployed.

Employment history = 1 denotes respondent was employed in more surveys than unemployed.

Unemployment history = 1 denotes respondent was unemployed in more surveys than employed.

Appendix B

Appendix for Chapter 3

Due to the analytical intractability of certain issues dealt with in Chapter 3, some key Propositions within the chapter are proved by exhaustive tabulation of admissible parameter configurations.

Two tabulations have been prepared: one for Section 3.3.2 and one for Section 3.4.1. For each Section, this is done by using a table of values for different combinations of values of the three parameters γ^* , h and q . Each table uses a step size of 0.05 for each parameter: γ^* goes from 0.05 to 0.95 inclusive (19 distinct values), h goes from 0.05 to 1.00 inclusive (20 distinct values) and q_H goes from 0.50 to 1.00 (11 distinct values). Therefore, each table has a total number of $19 \times 20 \times 11 = 4,180$ unique configurations. Each table is prepared for 10 periods (i.e. for $t = 1, 2, \dots, 10$).

Each table comprises the following measures.

Column C contains the value of γ^* (an input).

Column D contains the value of h (an input).

Column E contains the value of q_H for the Section 3.3.2 table and q for the Section 3.4.1 table (an input in each case).

Columns F to O evaluate Condition 1 in (3.13) for each period: TRUE if the firm is not turnover-constrained (small firm) and FALSE if constrained (large firm).

Columns P to Y contain the value of α_t for each period, assuming that a repeated hire-fire policy is used.

Columns AA to AJ contain the value of $\alpha_t - \alpha_{t-1}$ for each period. For example,

Column AA contains $\alpha_2 - \alpha_1$ assuming that a repeated hire-fire policy is used.

Columns AL to AU contain the firm's desired turnover for each period, assuming that a repeated hire-fire policy is used. Units are as a proportion of the entire labour force.

Columns AV to BE contain the firm's desired turnover for each period, assuming that a repeated hire-fire policy is used. Units are as a proportion of the firm's labour force.

Columns BF to BO contain the firm's actual turnover for each period, assuming that a repeated hire-fire policy is used. If not constrained, this number is identical to its desired turnover and otherwise the constrained turnover is given. Units are as a proportion of the entire labour force.

Columns BP to BY contain the firm's actual turnover for each period, assuming that a repeated hire-fire policy is used. If not constrained, this number is identical to its desired turnover and otherwise the constrained turnover is given. Units are as a proportion of the entire labour force.

Columns CA to CD prove that, for all 4,180 parameter configurations used, Propositions 3, 5, 6 and 8 are true respectively. Method of proof is to assume the converse of each Proposition and show that this converse is false for every one of the 4,180 configurations.

Each of the examples from the main body of Section 3.3.2 (i.e. Examples 3.1 to 3.4) are marked in **bold** within the tabulations.

Due to their large size, these tabulations are provided as additional information accompanying this work on a portable medium or otherwise available upon request from the author.

Appendix C

Appendices for Chapter 4

C.1 Data Appendix

Descriptions of variables used

All variables are, either directly or indirectly, from the British Household Panel Survey (BHPS) over the period 1991-2000.

Dependent Variable: Employment

Variables *wJBSTAT* are used. The directions to each respondent were “Please look at this card and tell me which best describes your current situation”. The card specifies the following options:

1. Self employed.
2. In paid employment.
3. Unemployed.
4. Retired.
5. Family care.
6. Full time student.
7. Long term sick/disabled.

8. On maternity leave.
9. Government training scheme.
10. Something else.

Recodes:

"Self-employed" and "Employed" = employed (1)

"Unemployed" and "Government training scheme" = unemployed (0)

All else = missing, and missing values are then excluded to obtain a balanced panel.

Variable: Age

Age is defined as age in 1991, and divided by 100. Variable *wAGE* for 1991 is used.

Variable: Age Squared

Age is defined as age in 1991, and squared then divided by 1000. Variable *AAGE* for 1991 is used.

Variable: Marital Status

Variables *wMASTAT* are used.

Recodes:

"Married" = 1

All else ("single", "widowed", divorced" or "separated") = 0

Variable: Number of Children

Variables *wNKIDS* are used, and divided by 10.

Variable: Years of Education

Variables *wQFACHI* (Highest Qualification Obtained) and *wSCEND* (age left school) are used.

Recodes:

If *wQFACHI* = 1 (higher degree) then recoded var = 18

If *wQFACHI* = 2 (degree) then recoded var = 16

If *wQFACHI* = 3 (teaching qualification or diploma) then recoded var = 13

If *wQFACHI* = 4 (A level) then recoded var = 13

If *wQFACHI* = 5 (O level) then recoded var = 11

If *wQFACHI* = 6 (CSE qualification) then recoded var = 11

If $wQFACHI = 7$ (none of the above) then recoded var = $ASCEND-6$ (ie. 1991 Age left school less 6).

If $wQFACHI = -7$ then recoded var is imputed accoring to previous year's value. (there are no proxies in 1991)

This recoded variable is then divided by 100.

Variable: Health

Variables $wHLSTAT$ are used. Respondents were asked: "Please think back over the last 12 months about how your health has been. Compared to people of your own age, would you say that your health has on the whole been..."

Recodes:

If $wHLSTAT = 1$ (Excellent), 2 (Good) or 3 (Fair) then recoded variable =1

If $wHLSTAT = 4$ (Poor) or 5 (Very Poor) then recoded variable = 0

Variable: Housing

Variables $wHSOWND$ are used. Respondents were asked: "Does your household own or rent this accommodation or does it come rent-free?"

Recodes:

If $wHSOWND = 1$ (Owned or on Mortgage) or 2 (Shared Ownership) then recoded variable =1

If $wHSOWND = 3$ (Rented), 4 (Rent Free) or 5 (Other) then recoded variable =0

Variable: Domicile Region

Variables $wREGION$ are used. Responses are derived from the BHPS questionnaire.

Recodes:

If $wREGION = 1$ (Inner London), 2 (Outer London), 3 (Rest of South East) or 4 (South West) then recoded variable =1

If $wREGION = 5$ (East Anglia), 6 (East Midlands), 7 (West Midlands Conurb), 8 (Rest of West Midlands), 9 (Greater Manchester), 10 (Merseyside), 11 (Rest of North West), 12 (South Yorkshire), 13 (West Yorkshire), 14 (Rest of Yorks and Humber), 15 (Tyne and Wear), 16 Rest of North, 17 (Wales) or 18 (Scotland) then recoded variable = 0

Variable: Time A time trend is used, where 1991=0.1,1992=0.2,...,2000=1.0.

Variable: Time Squared

1991=0.01,1992=0.04,...,2000=1.0.

Explanatory Variables: Initial Conditions Models

Constant

As for overall model.

Age

As for overall model.

Age Squared

As for overall model.

Marital Status

As for overall model.

Number of Children

As for overall model.

Health

As for overall model.

Years of Education

As for overall model.

Housing

As for overall model.

Domicile Region

As for overall model.

C.2 Further Details of Estimation Procedures used

C.2.1 Nonlinear Models

FE-HK = Honore/Kyriazidou (2000) fixed effects estimator.

RE-EX = Heckman (1981b) random effects exogenous initial conditions estimator.

RE-EQ = Heckman (1981b) random effects reduced form initial conditions estimator.

CRE-R = Heckman (1981b) random effects restricted correlation and reduced form initial conditions estimator.

CRE-UN = Heckman (1981b) random effects unrestricted correlation and reduced form initial conditions estimator.

Further notes:

- p values for all estimators are in parentheses.
- Standard errors for the Honore-Kyriazidou fixed effects model were obtained using a bootstrap procedure.
- Starting values set for all main parameters are identical across the five random effects estimators, and obtained by calculating a simple (i.e. not allowing for initial conditions) dynamic random effects logit. [STATA command: xtlogit empstat const empstat_1 age agesq manuf91 marital nchild educ health wave, re].
- Starting values set for parameters directly related to initial conditions (used for the final three estimators which model initial conditions in reduced form) are identical across these three estimators, and obtained by calculating a simple static logit. [STATA command: logit empstat age agesq manuf91 marital nchild educ health wave].
- Starting values for $var(\eta_i)$ were set to 1 across all five random effects estimators.

C.2.2 Linear Models

Difference GMM: One-Step Estimation To obtain standard errors, an estimator of $var\left(\hat{\boldsymbol{\mu}}_1^{DGMM}\right)$ is $\hat{V}_1\left[\hat{\boldsymbol{\mu}}_1^{DGMM}\right]$, where:

$$\begin{aligned}\hat{\varepsilon} &\equiv \mathbf{y}_{DGMM}^* - \mathbf{W}_{DGMM}^* \hat{\boldsymbol{\mu}}_1^{DGMM} \\ \hat{\sigma}_{DGMM}^2 &\equiv \frac{\hat{\varepsilon}'\hat{\varepsilon}}{N(T-2) - K} \\ \hat{V}_1 &= \hat{\sigma}_{DGMM}^2 \mathbf{M}_1^{-1}\end{aligned}$$

K is the number of parameters to be estimated, and $\mathbf{y}_{DGMM}^*$ and $\mathbf{W}_{DGMM}^*$ are, respectively, $\mathbf{y}_i^*$ and $\mathbf{W}_i^*$ stacked over individuals.

Because $\hat{V}_1\left[\hat{\boldsymbol{\mu}}_1^{DGMM}\right]$ is only a consistent estimator of $var\left(\hat{\boldsymbol{\mu}}_1^{DGMM}\right)$ under the assumption that the ε_{it} terms are homoskedastic, it may be desirable to use White's (1980) procedure for standard errors which are robust to heteroskedasticity. In this case, define $\hat{V}_{1r}\left[\hat{\boldsymbol{\mu}}_1^{DGMM}\right]$ as:

$$\hat{V}_{1r}\left[\hat{\boldsymbol{\mu}}_1^{DGMM}\right] = \mathbf{M}_1^{-1} \left(\sum_i \mathbf{W}_{iDGMM}^* \mathbf{Z}_{iDGMM} \right) \mathbf{A}_{1N} \mathbf{A}_{2N}^{-1} \mathbf{A}_{1N} \left(\sum_i \mathbf{Z}_{iDGMM}' \mathbf{W}_{iDGMM}^* \right) \mathbf{M}_1^{-1}$$

where $\mathbf{A}_{2N} \equiv \left(\sum_i \mathbf{Z}_{iDGMM}' \left(\hat{\varepsilon}_i \hat{\varepsilon}_i' \right) \mathbf{Z}_{iDGMM} \right)^{-1}$ and $\hat{\varepsilon}_i \equiv \mathbf{y}_{iDGMM}^* - \mathbf{W}_{iDGMM}^* \hat{\boldsymbol{\mu}}_1^{DGMM}$. [$\mathbf{A}_{2N}$ is discussed below in the context of the two-step estimator].

Difference GMM: Two-Step Estimation To obtain standard errors, an estimator of $var\left(\hat{\boldsymbol{\mu}}_2^{DGMM}\right)$ is $\hat{V}_2\left[\hat{\boldsymbol{\mu}}_2^{DGMM}\right] \equiv \mathbf{M}_2^{-1}$, and the two step estimated residuals are $\hat{\varepsilon}_i \equiv \mathbf{y}_{iDGMM}^* - \mathbf{W}_{iDGMM}^* \hat{\boldsymbol{\mu}}_2^{DGMM}$. Although $\hat{V}_2\left[\hat{\boldsymbol{\mu}}_2^{DGMM}\right]$ is a consistent estimator of $var\left(\hat{\boldsymbol{\mu}}_2^{DGMM}\right)$ as $N \rightarrow \infty$, simulations carried out by Arellano and Bond (1991) suggest that in small samples it can exhibit severe downward bias. As a result, inference on $\boldsymbol{\mu}$ can be more reliably conducted using the one-step estimator discussed above, and we adopt this practice throughout the paper. Calculating the two-step estimator is nonetheless valuable, since to do so is necessary for obtaining heteroskedasticity-robust standard errors to go with the one-step estimator, as discussed above. The two-step estimator also has some valuable properties for evaluating post-estimation diagnostics.

Note that $\mathbf{H}_{2i}^{DGMM}$, unlike $\mathbf{H}_{1i}^{DGMM}$, does vary across individuals. More impor-

tantly, it depends on parameters estimated at the first step, and used to construct $\widehat{\epsilon}_i$. This is why the usual asymptotic standard errors are not reliable in small samples. Windmeijer (2005) also shows that a useful small-sample correction can be implemented, which is installed as part of the software package DPD for OX.

Difference GMM: Diagnostic Tests To test the assumption of no second-order serial correlation in terms of the *transformed* shocks $\varepsilon_{it} - \varepsilon_{i,t-1}$, Arellano and Bond (1991) develop test statistics (“ m_1 ” and “ m_2 ”) for first and second order serial correlation, respectively, in the transformed residuals. Define, for individual i

$$w_{it} \equiv \begin{cases} \widehat{\epsilon}_{i,t-m} & \text{for } t = m, \dots, T \\ 0 & \text{for } t < m \end{cases}$$

$$\mathbf{w}_i \equiv (w_{i1}, w_{i2}, \dots, w_{iT})'$$

where m is the order of serial correlation one wishes to test for. Further, construct - from the full sample - the following statistics:

$$\begin{aligned} d_0 &\equiv \sum_i \mathbf{w}_i' \widehat{\epsilon}_i \\ d_1 &\equiv \sum_i \mathbf{w}_i' \mathbf{H}_{si}^{DGMM} \mathbf{w}_i \\ d_2 &\equiv -2 \left(\sum_i \mathbf{w}_i' \mathbf{W}_{iDGMM}^* \right) \mathbf{M}_s^{-1} \left(\sum_i \mathbf{W}_{iDGMM}^{*'} \mathbf{Z}_{iDGMM} \right) A_{sN} \left(\sum_i \mathbf{Z}_{iDGMM}' \mathbf{H}_{si}^{DGMM} \mathbf{w}_i \right) \\ d_3 &\equiv \left(\sum_i \mathbf{w}_i' \mathbf{W}_{iDGMM}^* \right) \widehat{V}_s \left(\sum_i \mathbf{W}_{iDGMM}^{*'} \mathbf{w}_i \right) \end{aligned}$$

where $s = 1, 2$ depending upon whether the 1 or 2 step estimator is being used. Then the m_1 and m_2 statistics are defined as:

$$m_1 \equiv \left[\frac{d_0}{(d_1 + d_2 + d_3)^{1/2}} \mid (m = 1) \right]$$

$$m_2 \equiv \left[\frac{d_0}{(d_1 + d_2 + d_3)^{1/2}} \mid (m = 2) \right]$$

Each statistic converges as $N \rightarrow \infty$ to the standard normal distribution under the null hypothesis of no serial correlation. Since, as discussed in the text, the tests are specified in

terms of the transformed errors $\varepsilon_{it} - \varepsilon_{i,t-1}$, only m_2 is of direct relevance to the question of the model's validity, so a valid model will typically reject the null for m_1 and fail to reject the null for m_2 .

The Sargan test statistic of overidentifying restrictions is defined as

$$\begin{aligned} S_1 &\equiv (\sum_i \hat{\varepsilon}_i' \mathbf{Z}_{iDGMM}) \mathbf{A}_{1N} (\sum_i \mathbf{Z}_{iDGMM}' \hat{\varepsilon}_i) \hat{\sigma}_{DGMM}^{-2} \\ S_2 &\equiv (\sum_i \hat{\varepsilon}_i' \mathbf{Z}_{iDGMM}) \mathbf{A}_{2N} (\sum_i \mathbf{Z}_{iDGMM}' \hat{\varepsilon}_i) \end{aligned}$$

(depending on whether one-step or two-step estimation is used), and follows an asymptotic chi square distribution if the null hypothesis of valid instruments is correct. The degrees of freedom for these tests are equal to the number of overidentifying restrictions. Because only S_2 is heteroskedasticity consistent, then it is advisable in most situations to use the two-step Sargan test in preference to the one-step test. This provides a further demonstration of the usefulness of the two-step estimation procedure for obtaining robust results, even though (as discussed above) this procedure is not usually appropriate for parameter inference unless the sample size N is extremely large [or unless finite sample corrected errors are used, as proposed by Windmeijer (2005)].

System GMM: One-Step Estimation Under one-step estimation of $\boldsymbol{\mu}$, the weighting matrix is:

$$\mathbf{H}_1^{SGMM} \equiv \begin{pmatrix} \mathbf{H}_1^{DGMM} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2}I \end{pmatrix}$$

where I is the identity matrix. The data are now organised as follows:

$$\begin{aligned} \mathbf{y}_{iSGMM}^* &\equiv [\Delta \mathbf{y}_{i3} \ \Delta \mathbf{y}_{i4} \ \dots \ \Delta \mathbf{y}_{iT} \ \mathbf{y}_{i3} \ \mathbf{y}_{i4} \ \dots \ \mathbf{y}_{iT}]' \\ &= [\mathbf{y}_{iDGMM}^* \ \mathbf{y}_{i3} \ \mathbf{y}_{i4} \ \dots \ \mathbf{y}_{iT}]' \end{aligned}$$

$$\begin{aligned} \mathbf{W}_{iSGMM}^* &\equiv [\Delta \mathbf{w}_{i3}' \ \Delta \mathbf{w}_{i4}' \ \dots \ \Delta \mathbf{w}_{iT}' \ \mathbf{w}_{i3}' \ \mathbf{w}_{i4}' \ \dots \ \mathbf{w}_{iT}']' \\ &= [\mathbf{W}_{iDGMM}^* \ \mathbf{w}_{i3}' \ \mathbf{w}_{i4}' \ \dots \ \mathbf{w}_{iT}']' \end{aligned}$$

where $\mathbf{w}_{it} \equiv (y_{it-1} \ x'_{it})$, $\Delta \mathbf{w}_{it} = (\Delta y_{it-1} \ \Delta x'_{it})$, $\Delta ()_t \equiv ()_t - ()_{t-1}$ and the apostrophe ' denotes matrix transposition. The form of the system one-step estimator $\hat{\boldsymbol{\mu}}_1^{SGMM}$ is then identical to the difference estimator $\hat{\boldsymbol{\mu}}_1^{DGMM}$:

$$\begin{aligned} \mathbf{A}_{1N}^S &\equiv (\sum_i \mathbf{Z}'_{iSGMM} \mathbf{H}_1^{SGMM} \mathbf{Z}_{iSGMM})^{-1} \\ \mathbf{M}_{1S} &\equiv (\sum_i \mathbf{W}_{iSGMM}^* \mathbf{Z}_{iSGMM}) \mathbf{A}_{1N}^S (\sum_i \mathbf{Z}'_{iSGMM} \mathbf{W}_{iSGMM}^*) \\ \hat{\boldsymbol{\mu}}_1^{SGMM} &= \mathbf{M}_{1S}^{-1} (\sum_i \mathbf{W}_{iSGMM}^* \mathbf{Z}_{iSGMM}) \mathbf{A}_{1N}^S (\sum_i \mathbf{Z}'_{iSGMM} \mathbf{y}_{iSGMM}^*) \end{aligned}$$

To obtain standard errors, note there are two sets of residuals - one from the transformed equations and the other from the equations in levels. Utilising the latter, we proceed analogously:

$$\begin{aligned} \hat{\varepsilon}_S &\equiv \mathbf{y}_i - \mathbf{W}_i \hat{\boldsymbol{\mu}}_1^{SGMM} \\ \hat{\sigma}_{SGMM}^2 &\equiv \frac{\hat{\varepsilon}_S' \hat{\varepsilon}_S}{N-1} \\ \hat{V}_{1S} &= \hat{\sigma}_{SGMM}^2 \mathbf{M}_{1S}^{-1} \end{aligned}$$

and heteroskedasticity-robust standard errors can again be calculated::

$$\hat{V}_{1Sr} [\hat{\boldsymbol{\mu}}_1^{SGMM}] = \mathbf{M}_{1S}^{-1} (\sum_i \mathbf{W}_{iSGMM}^* \mathbf{Z}_{iSGMM}) \mathbf{A}_{1N} \mathbf{A}_{2N}^{-1} \mathbf{A}_{1N} (\sum_i \mathbf{Z}'_{iSGMM} \mathbf{W}_{iSGMM}^*) \mathbf{M}_{1S}^{-1}$$

where $\mathbf{A}_{2N}^S \equiv (\sum_i \mathbf{Z}'_{iSGMM} (\hat{\varepsilon}_S \hat{\varepsilon}_S') \mathbf{Z}_{iSGMM})^{-1}$ again arises from two-step estimation.

System GMM: Two-Step Estimation Under two-step system GMM estimation of $\boldsymbol{\mu}$, the weighting matrix is again the vector product of the estimated residuals from the corresponding one-step estimation immediately above, so that $\mathbf{H}_2^{SGMM} \equiv \hat{\varepsilon}_S \hat{\varepsilon}_S'$. Then the two-step estimator $\hat{\boldsymbol{\mu}}_2^{DGMM}$ is as follows :

$$\begin{aligned} \mathbf{A}_{2N}^S &\equiv (\sum_i \mathbf{Z}'_{iSGMM} \mathbf{H}_2^{SGMM} \mathbf{Z}_{iSGMM})^{-1} \\ \mathbf{M}_{2S} &\equiv (\sum_i \mathbf{W}_{iSGMM}^* \mathbf{Z}_{iSGMM}) \mathbf{A}_{2N}^S (\sum_i \mathbf{Z}'_{iSGMM} \mathbf{W}_{iSGMM}^*) \\ \hat{\boldsymbol{\mu}}_2^{SGMM} &= \mathbf{M}_{2S}^{-1} (\sum_i \mathbf{W}_{iSGMM}^* \mathbf{Z}_{iSGMM}) \mathbf{A}_{2N}^S (\sum_i \mathbf{Z}'_{iSGMM} \mathbf{y}_{iSGMM}^*) \end{aligned}$$

Standard errors can again be obtained by $\widehat{V}_{2S} \left[\widehat{\boldsymbol{\mu}}_2^{SGMM} \right] \equiv \mathbf{M}_{2S}^{-1}$, and the two- step estimated residuals are $\widehat{\boldsymbol{\varepsilon}}_{iS} \equiv \mathbf{y}_{iSGMM}^* - \mathbf{W}_{iSGMM}^* \widehat{\boldsymbol{\mu}}_2^{SGMM}$. Further simulations by Blundell and Bond (1998) suggest that, as is the case for Difference GMM, the two-step estimator is less reliable for parameter inference than the one-step estimator.

System GMM: Diagnostic Tests To test the assumption of no serial correlation in the error terms ε_{it} , Arellano and Bond's (1991) m_1 and m_2 test statistics can be used, although they are defined slightly differently for System GMM estimation. Use the same specification for w_{it} (i.e. based on the equations in first differences) as before:

$$w_{it} \equiv \begin{cases} \widehat{\varepsilon}_{i,t-m} & \text{for } t = m, \dots, T \\ 0 & \text{for } t < m \end{cases}$$

$$\mathbf{w}_i \equiv (w_{i1}, w_{i2}, \dots, w_{iT})'$$

where m is the order of serial correlation one wishes to test for. Now construct the following statistics from a mixture of system and difference data:

$$\begin{aligned} d_4 &\equiv \sum_i \mathbf{w}_i' \widehat{\boldsymbol{\varepsilon}}_i \\ d_5 &\equiv \sum_i \mathbf{w}_i' \widehat{\boldsymbol{\varepsilon}}_i \widehat{\boldsymbol{\varepsilon}}_i' \mathbf{w}_i \\ d_6 &\equiv -2 \left(\sum_i \mathbf{w}_i' \mathbf{W}_{iDGMM}^* \right) \mathbf{M}_s^{-1} \left(\sum_i \mathbf{W}_{iDGMM}^{*'} \mathbf{Z}_{iSGMM} \right) A_{sN} \left(\sum_i \mathbf{Z}_{iSGMM}' \widehat{\boldsymbol{\varepsilon}}_i \widehat{\boldsymbol{\varepsilon}}_i' \mathbf{w}_i \right) \\ d_7 &\equiv \left(\sum_i \mathbf{w}_i' \mathbf{W}_{iDGMM}^* \right) \widehat{V}_{sS} \left(\sum_i \mathbf{W}_{iDGMM}^{*'} \mathbf{w}_i \right) \end{aligned}$$

where $s = 1, 2$ depending upon whether the 1 or 2 step estimator is being used. Then the m_1 and m_2 statistics are analogously defined as:

$$m_1 \equiv \left[\frac{d_4}{(d_5 + d_6 + d_7)^{1/2}} \mid (m = 1) \right]$$

$$m_2 \equiv \left[\frac{d_4}{(d_5 + d_6 + d_7)^{1/2}} \mid (m = 2) \right]$$

and each statistic again converges as $N \rightarrow \infty$ to the standard normal distribution under the null hypothesis of no serial correlation.

To test the *full* set of restrictions exploited by System GMM estimation, a Sargan test statistic, of which the preferred two-step statistic is:

$$S_{2S} \equiv (\sum_i \hat{\epsilon}_{iS}' \mathbf{Z}_{iSGMM}) \mathbf{A}_{2N}^S (\sum_i \mathbf{Z}_{iSGMM}' \hat{\epsilon}_{iS})$$

can be used. Alternatively, the *additional* restrictions imposed by System GMM can be tested using a "Sargan Difference test", which compares the System GMM Sargan statistic to the Difference GMM Sargan statistic. This last statistic is:

$$\Delta S \equiv S_{2S} - S_2$$

and follows an asymptotic chi square distribution if the null hypothesis, that the additional instruments are valid, is correct. The test's degrees of freedom are equal to the number of extra overidentifying restrictions introduced by System GMM relative to Difference GMM.

C.3 Tables of Summary Statistics for Auxiliary Samples E1 and F1

Table C.1: Sample Characteristics for BHPS: Sample E1 1991-1995

	Full Sample (1)	Employed 5 years (2)	Employed 0 years (3)	Single Transition from Work (4)	Single Transition to Work (5)	Multiple Transitions (6)
Means and Standard Deviations						
Employment (1=Employed)	0.94 (0.23)	1.00 (0.00)	0.00 (0.00)	0.59 (0.49)	0.66 (0.47)	0.70 (0.46)
Age/100 (1991)	0.36 (0.10)	0.36 (0.10)	0.34 (0.12)	0.38 (0.11)	0.30 (0.09)	0.31 (0.10)
Marital Status	0.76 (0.43)	0.79 (0.42)	0.64 (0.48)	0.69 (0.46)	0.65 (0.48)	0.59 (0.49)
Children/10	0.09 (0.11)	0.09 (0.11)	0.15 (0.12)	0.07 (0.10)	0.10 (0.11)	0.09 (0.12)
Education/100	0.12 (0.02)	0.12 (0.02)	0.10 (0.01)	0.11 (0.02)	0.11 (0.02)	0.12 (0.02)
Health	0.97 (0.16)	0.98 (0.16)	0.98 (0.14)	0.97 (0.17)	1.00 (0.06)	0.95 (0.21)
Housing	0.82 (0.39)	0.85 (0.36)	0.06 (0.24)	0.67 (0.47)	0.62 (0.49)	0.74 (0.44)
Region	0.45 (0.50)	0.46 (0.50)	0.30 (0.46)	0.28 (0.45)	0.47 (0.50)	0.50 (0.50)
Sample Size	1285	1088	10	35	66	86

Notes:

Standard deviations are in parentheses.

Variable definitions are contained in Appendix C.1.

The noteworthy differences are that those in Sample F1 are – on average – 3 years older than those in Sample E1, have employment rates which are 3 percentage points lower (91% cf. 94%) and are 15 percentage points less likely to live in the South of England (30% cf. 45%). Those in F1 are also marginally more likely to be in cohabitation and marginally less likely to own their place of residence. However, among those who make multiple transitions (who predominantly are the individuals who allow us to identify state dependence effects), the age gap is smaller (2 years) and 30% live in the South of England in Sample F1 cf. 50% of those in Sample E1 who make multiple transitions.

Table C.2: Sample Characteristics for BHPS: Sample F1 1991-1995

	Full Sample (1)	Employed 5 years (2)	Employed 0 years (3)	Single Transition from Work (4)	Single Transition to Work (5)	Multiple Transitions (6)
Means and Standard Deviations						
Employment (1=Employed)	0.91 (0.29)	1.00 (0.00)	0.00 (0.00)	0.59 (0.49)	0.68 (0.47)	0.66 (0.48)
Age/100 (1991)	0.39 (0.13)	0.40 (0.12)	0.33 (0.12)	0.40 (0.12)	0.34 (0.14)	0.33 (0.12)
Marital Status	0.78 (0.42)	0.81 (0.39)	0.59 (0.49)	0.72 (0.45)	0.62 (0.49)	0.65 (0.48)
Children/10	0.08 (0.11)	0.08 (0.10)	0.15 (0.15)	0.10 (0.11)	0.09 (0.10)	0.09 (0.12)
Education/100	0.12 (0.02)	0.12 (0.02)	0.11 (0.02)	0.11 (0.02)	0.11 (0.02)	0.11 (0.02)
Health	0.96 (0.19)	0.96 (0.19)	0.96 (0.19)	0.95 (0.22)	0.98 (0.14)	0.97 (0.18)
Housing	0.80 (0.40)	0.84 (0.37)	0.20 (0.40)	0.85 (0.36)	0.63 (0.49)	0.69 (0.46)
Region	0.30 (0.46)	0.29 (0.45)	0.41 (0.49)	0.38 (0.49)	0.33 (0.47)	0.30 (0.46)
Sample Size	769	611	22	31	39	66

Notes:

Standard deviations are in parentheses.

Variable definitions are contained in Appendix C.1.