Job Mobility and Measurement Error

by

Adele Bergin

A thesis submitted in fulfilment of the requirements for the degree of
Ph.D.

In the Department of Economics, Finance and Accounting,
National University of Ireland, Maynooth (NUIM), Maynooth, Co. Kildare,
Ireland.

August 2011

Head of Department: Professor Rowena Pecchenino

Thesis Supervisors: Professor Donal O'Neill (Lead Supervisor), Dr. Aedin Doris and
Dr. Olive Sweetman

NUI MAYNOOTH

Ollscoil na hÉireann Má Nuad

Table of Contents

List of Tables

List of Figures

Summary

This thesis consists of essays investigating job mobility and measurement error. Job mobility, captured here as a change of employer, is a striking feature of the labour market. In empirical work on job mobility, researchers often depend on self-reported tenure data to identify job changes. There may be measurement error in these responses and consequently observations may be misclassified as job changes when truly no change has taken place and vice versa. These observations serve as a starting point for this thesis.

Chapter 3 explores the level and determinants of job mobility in Ireland, using the Living in Ireland Survey, the Irish component of the European Community Household Panel. One of the findings is that the rate of voluntary (i.e. employee initiated) job mobility in Ireland trebled over the period 1995 to 2000. A decomposition technique indicates that composition changes only explain around one-fifth of the increase, while the remainder reflects changes in operation of the labour market.

Chapter 4 uses Monte Carlo simulation techniques to investigate the performance of a modified probit estimator developed by Hausman, Abrevaya and Scott-Morton (1998) that controls for misclassification in a dependent variable. An analysis of the data indicates that there is the possibility of substantial measurement error which may make it difficult to capture job changes. The Hausman *et al.* (1998) estimator is used to formally control for measurement error in models of job change for Ireland (Chapter 5) and other European countries (Chapter 6). The main findings are that the true rates of job change are being severely undercounted in several countries and also that similar factors are important in determining job changes across countries.

Finally, Chapter 7 contributes to the existing literature that examines the impact of job mobility on wage growth. It finds that by controlling for measurement error in job changes, the effect of job mobility on wage growth is larger than prevailing estimates suggest.

List of Presentations and Publications

A version of chapter 3 has been published as Bergin, A. (2009), "Job Mobility in Ireland", *The Economic and Social Review*, No. 1, Spring, pp. 15-47.

Presentations:

"Job Mobility in Ireland", Irish Economic Association, 22nd Annual Conference, Mayo, Ireland, April 2008.

"Measurement Error in Survey Data: Job Mobility in Ireland", NUI Maynooth, Department of Economics, Finance and Accounting, December 2008.

"Measurement Error in Survey Data: Job Mobility in Ireland", 12th IZA European Summer School in Labor Economics, Ammersee, Germany, 27th April to 3rd May, 2009.

"Measurement Error in Survey Data: Job Mobility in Europe", NUI Maynooth, Department of Economics, Finance and Accounting, May 2010.

"Wage Changes and Job Changes: Estimation with Measurement Error in a Binary Regressor", NUI Maynooth, Department of Economics, Finance and Accounting, April 2011.

Acknowledgements

I would like to thank many people for their help in completing this thesis.

First and foremost, I would like to express my gratitude and appreciation to my supervisors, Professor Donal O'Neill, Dr. Aedin Doris and Dr. Olive Sweetman, for their guidance, insight and commitment over the course of my doctoral studies. I have learned a lot from them.

I am indebted to the late Professor Denis Conniffe for his advice and valuable comments at different stages of this work. I would like to thank Dr. Alan Barrett for his continuous encouragement. I am also grateful to the staff and my fellow students in the Department and my colleagues in the ESRI for their help, advice and constructive comments and suggestions.

Finally, I would like to express my heartfelt thanks to my friends and family, especially my parents, for their tireless support, patience and encouragement.

1. Introduction

This thesis consists of six chapters (in addition to this introduction) investigating job-to-job mobility and measurement error in capturing job changes. Job mobility is an important phenomenon to understand because the movement of workers from one job to another allows for flexibility in the labour market by providing workers and firms with a mechanism to adapt to changing economic and personal circumstances. This churning can be seen as the efficient working of the labour market; workers can seek out new employment matches that they are more productive in and for which they will be better rewarded.

Survey data are very often used in applied work on job mobility. Typically, surveys do not contain a direct question asking respondents if they have changed jobs; instead job changes are inferred from responses to a question about tenure. It is documented in the literature that there is substantial measurement error in the responses to questions that are used to determine job changes (Brown and Light (1992)). However, most research ignores this measurement error and so there is a risk of misclassifying cases as job changes when truly no change has taken place and vice versa.[1] This is one of the central concerns of this thesis. In empirical research, job changes are captured in a binary variable (an observation is coded as a job change or a job stay). Measurement error in this case is non-classical in nature, as it is negatively correlated with the true variable.

The decision to change jobs can be analysed in a binary choice model, such as a probit model. Measurement error in the dependent variable in a nonlinear model leads to estimates that are biased and inconsistent (Hausman (2001)). Hausman, Abrevaya and Scott-Morton (1998) develop an estimator that controls for misclassification in the dependent variable in discrete choice models. The estimator explicitly incorporates the misclassification probabilities as additional parameters to be estimated. Hausman *et al.* (1998) show that even when only a small amount of data is misclassified that ordinary probit estimates are severely biased.

[1] See Bound, Brown and Mathiowetz (2001) for a survey of measurement error in survey data.

There are also serious inference problems when a binary job change variable is used as an explanatory variable in a regression model, e.g. in estimating the wage effects associated with job mobility. Aigner (1973) demonstrates how, in the presence of a misclassified binary regressor, OLS coefficient estimates are biased towards zero.

Therefore, in cases where misclassification is likely, it is essential to control for it, both when the potentially misclassified variable is the dependent variable in an analysis and when it is an independent variable.

The determinants of job mobility and also the wage impacts of changing jobs have long been studied in labour economics. One of the contributions of this thesis is to examine how the results from these types of analyses are altered by measurement error in calculating job changes. In particular, it examines the extent to which job changes may be under- or over-counted in Ireland, and also in a range of other European countries. It also seeks to assess the impact that ignoring misclassification has on covariate effects in models of job change. In addition, it investigates the relationship between job mobility and wage growth for Ireland and the degree to which the estimated impact of job mobility on wage growth is affected by measurement error.

In what follows, I summarise each chapter and the main conclusions.

Chapter 2 contains a literature review on job mobility that is common to several of the subsequent chapters. It covers the theoretical approaches to job mobility and the associated wage effects, the patterns we should observe in the data and empirical evidence.

Chapter 3 investigates the factors that determine job-to-job mobility in Ireland over the period 1995 to 2001, using the Living in Ireland Survey, the Irish component of the European Community Household Panel. Changing jobs appears to be an important part of worker's experience in the labour market yet little is known, in an Irish context, about how prevalent job changing is, as well as what types of worker are most likely to switch jobs and this chapter seeks to bridge that gap. It finds that each year approximately 10 per cent of workers change jobs.

The chapter distinguishes between voluntary (employee initiated) and involuntary (employer initiated) mobility. It finds that labour market experience, working in the public sector, whether a person is overskilled, the sector they work in and their occupation are important determinants of voluntary job change. The dataset used covers most of the Celtic Tiger period, a time where growth in the Irish economy was exceptional. The chapter finds the rate of voluntary job mobility in Ireland trebled over the period 1995 to 2000. The sample is divided into two time periods and a decomposition technique is applied to ascertain how much of the increase in mobility is attributable to compositional changes and how much is due to other factors. Compositional changes explain around one-fifth of the increase, while the changes in the labour market conditions facing workers are also an important factor driving the increase.

Chapter 4 analyses in detail the estimator developed by Hausman *et al.* (1998) to control for misclassification in binary choice models. The chapter uses Monte Carlo simulations to compare the estimates from the Hausman *et al.* model with those from an ordinary probit model for different model specifications with varying levels of misclassification. There are many sources of misclassification in models that employ binary dependent variables, such as coding error, self-reporting, recall error and where a dummy variable is used to serve as a proxy for some true underlying variable. The aim of the chapter is to provide insights into when it is reasonable to use the Hausman *et al.* estimator to control for misclassification. In general, the Hausman *et al.* estimator outperforms an ordinary probit model but in some cases, especially when the range of $x_i^{'}\beta$ is limited or when both the sample size is small and the level of misclassification is low, a probit model is superior.

Chapter 5 explores the extent of measurement error evident in responses to the question used to capture job changes for Ireland using the Living in Ireland Survey. It finds that the extent of measurement error in the data is considerable but it is similar to what Brown and Light (1992) find for the US using the Panel Study of Income Dynamics. The chapter uses the Hausman *et al.* (1998) estimator to formally control for misclassification in job changes. It finds that, by ignoring misclassification, the true number of job changes is underestimated by around 60 per cent. To put this figure in context, the average mobility rate in the dataset is calculated at around 9.7 per cent and this estimate implies that the true mobility rate is around 15.6 per cent. In addition, the chapter finds that ignoring misclassification leads to diminished covariate effects. The chapter also explores the possibility of covariate dependent misclassification but no support is found for this hypothesis.

Chapter 6 extends the analysis in Chapter 5 to other European countries in the European Community Household Panel dataset. The chapter finds that the true rates of job mobility are undercounted in several countries, typically in the peripheral countries of the EU. In addition, it finds that similar factors are important in determining job mobility across countries. Apart from age, personal and household characteristics are generally not important in explaining job changes, while firm and job characteristics have an important role in explaining job changes. For the countries, where the model incorporating misclassification is accepted, the impact of the covariates is much stronger indicating that ordinary probit estimates are biased downwards.

Chapter 7 focuses on the wage effects associated with changing jobs. Previous research has examined the role that job mobility makes to life cycle wage growth (e.g. (Topel and Ward (1992)) or how gender differences in the returns to job mobility may help explain the gender pay gap (e.g. Loprest (1992)). Estimates of the returns to job mobility are crucial to such studies. The chapter finds OLS estimates (that ignore misclassification) of the effect of job mobility on wage growth of around 8 per cent. This effect persists even after controlling for unobserved individual heterogeneity. This estimate is very similar to what has been found in other studies (e.g. OECD (2010)). However these estimates ignore measurement error in job changes. The chapter adopts a two-step approach to controlling for misclassification in a binary explanatory variable. It finds that controlling for misclassification has a substantial effect on the estimated impact changing jobs has on wage growth. The effect of job mobility on wage growth is estimated to be closer to 14 per cent when measurement error is controlled for.

2. Literature Review on Job Mobility

2.1 Introduction

This chapter discusses the theoretical and empirical background for some of the subsequent chapters.[2] Section 2.2 reviews theories of job mobility and the associated implications for wages. Section 2.3 explores the types of patterns we should observe in the data and the related empirical evidence. Finally, Section 2.4 focuses on the empirical research on the wage effects of job mobility.

2.2 Why do Workers Change Jobs and how does Job Mobility affect Wages? Theoretical Considerations

It is widely accepted that human capital accumulation, both general and specific to the job, is a major driving force of wage growth over the life-cycle. In equilibrium, in a competitive labour market with perfect information, workers are paid their marginal product which depends on human capital investment. However, if the labour market is characterised by heterogeneity in workers and firms and by incomplete information, it is only after a series of job matches, separations and new job matches that this equilibrium is achieved. For example, imperfect information may mean that firms are uncertain about the productivity of workers at the beginning of employment relationships. As a result, workers may not be initially employed in the jobs in which they are the most productive. Job mobility provides a mechanism for the labour market to move towards a more efficient allocation of resources whereby workers sort themselves into jobs that maximise their productivity. Therefore job mobility may make an important contribution to life-cycle wage growth.

In the empirical literature, it is quite difficult, and not always possible, to distinguish between different models of job mobility and often they are seen as complementary in the sense that each contributes to our understanding of a worker's decision to change

[2] Sections 2.2 and 2.3 essentially provide the theoretical background and empirical evidence on job mobility relevant for Chapters 3, 5 and 6. In addition to Sections 2.2 and 2.3, Section 2.4 is important for Chapter 7.

jobs and the effect this can have on wages (Veum (1997)).[3] A few papers attempt to disentangle the wage effects from different models of job mobility (e.g. Light and McGarry (1998), Campbell (2001) and Munasinghe and Sigman (2004)). Underlying the various models are different causal mechanisms through which job mobility affects wages. Also, some of the theoretical models make different predictions about the impact of changing jobs on wages. As a result, it is useful to understand and to identify these different mechanisms and the associated predicted wage effects.[4] Four main theoretical approaches can be distinguished in the literature, namely the mover-stayer model, job search models, job matching models and human capital models.

Mover-Stayer Model

The mover-stayer model of Blumen *et al.* (1955) comes from the sociology literature and is one of the earliest models of job mobility. In this model some individuals are instinctively more likely to change jobs than others. This inherent "itch" or "hobo syndrome" arises from some underlying unobservable personal characteristic(s). This characteristic results in people being consistently high or low mobility individuals over time. It is assumed that this instability makes movers less productive than stayers. The implication for wages is that movers earn lower wages because they are less productive workers. In this model, mobility is negatively correlated with wages because it is correlated with the unobservable characteristic that determines productivity. Therefore, in empirical work controlling for unobserved individual heterogeneity should mean that the wages of movers and stayers do not differ. Empirical evidence tends to refute the predictions of this model (e.g. Light and McGarry (1998) and Munasinghe and Sigman (2004)).

Job Search Models

In job search models (e.g. Burdett (1978)) there is heterogeneity in worker productivity across jobs and so a worker's productivity depends on the firm they are employed in. The quality (i.e. productivity) of an employment match is known in advance. Workers search for better matches and so they move from lower to higher paying jobs as the opportunity arises.

[3] In fact, many models of job mobility combine some of the approaches outlined here. For example, the job separation model in Jovanovic (1979a) merges the specific capital approach with job search theory.
[4] In the models examined in this section, job mobility refers to voluntary (employee initiated) job changes.

In Burdett (1978) when the worker enters the labour market, if they accept the first job offer they obtain, their wage can be seen as a random draw from a distribution of wage offers reflecting their differing productivities in each of the jobs available. Once employed, the worker can engage in search activity. Any job offer they receive from a firm will reflect their productivity in that firm. The more intensely a worker searches, the faster is the arrival rate of alternative offers. Workers will have an incentive to switch jobs if the present value of the wage in the alternative job exceeds that of the existing job, net of any mobility costs. Therefore, the model predicts that job mobility has a positive effect on wages.

In this type of search model, the wage gain arising from mobility comprises a discrete jump in the wage level at the point of job change. Wages are not affected by mobility as such, but rather by an improvement in match quality which is constant (within the job). Therefore, mobility should have no effect on wages once time constant job-specific effects are controlled for. The wage effects predicted by this model have received mixed empirical support. Light and McGarry (1998) find that job mobility has an effect on wages, even after controlling for observed and unobserved personal and job characteristics. However, Campbell (2001) estimates how much of the wage effect associated with mobility is attributable to a change in the wage level when the job change occurs and how much is due to a change in wage growth. He finds that around 40 per cent of the gain associated with job mobility is attributable to an increase in the wage level at the time of changing jobs and the remainder is due to higher on-the-job wage growth in the new job so this type of search model only provides a partial explanation of the wage effect associated with changing jobs.

More recent models extend the approach in Burdett (1978). For example, Burdett and Mortensen (1998) incorporate the reactions of firms. Naticchiono and Panigo (2004) classify this type of search model as “static” in the sense that is does not allow for any within-job wage dynamics. Other search models allow for on-the-job wage growth. For example, in Burdett and Coles (2003) and Stevens (2004), employers post wage-tenure contracts (instead of just wages) that allow wages to increase with tenure, to reduce the quit probability of their employees. In Postel-Vinay and Robin (2002) wages increase on-the-job because of outside offers. In these more recent search models, the gain from changing jobs will not necessarily comprise a discrete jump in the wage level at the time of job change but rather depend on the wage growth in the new job.

Job Matching Models

The key feature of the job matching approach is that match quality is not known *ex ante*. In matching models (e.g. Johnson (1978), Jovanovic (1979b) and Viscusi (1979)), jobs are considered as 'experience goods' and it is only over time that information about the quality of the match is revealed. Similar to the job search approach, matching models tend to assume there is heterogeneity in worker productivity across jobs.

In Jovanovic (1979b), workers face a distribution of wages reflecting firm's estimates of their productivity, rather than their actual productivity, in different jobs. As tenure on the job increases firms accumulate additional information about a worker's true productivity. This can lead to an upward or downwards adjustment to wages.[5] Workers who experience wage cuts or have low on-the-job wage growth are likely to separate from their employer. Specifically, job changes occur if the expected value of an offer at another firm exceeds the expected value of the wage in the existing job. This model predicts a positive relationship between job mobility and wage growth, although it is not a direct relationship but rather wage growth is affected by superior perceptions of match quality. One of the implications of the model is that a worker may be willing to accept a wage cut at the time of changing jobs (i.e. in the short-run) if they receive higher on-the-job wage growth in the new job. Naticchiono and Panigo (2004) classify this type of model as "dynamic" as it incorporates within-job wage dynamics. However, on-the-job wage growth is not attributable to changes in productivity but rather from changes in the firm's assessment of the worker's productivity.

In matching models, mobility is driven by changes in observations of match quality. Although true match quality is constant within the job, views about match quality can change over time. Therefore, there will be a relationship between job mobility and

[5] Other models (e.g. Greenwald (1986) and Gibbons and Katz (1991)) explore the possibility of informational asymmetries where employers have private information about the ability of their employees and how this affects employees' wages and mobility.

wages even after individual and job specific observable and unobservable effects are controlled for. Although we expect the relationship between job mobility and wages to be positive, it can be negative in the short-term. Light and McGarry (1998) find that mobility has an effect on wages after controlling for individual and job specific unobservable effects and argue this finding is consistent with the job matching model.

Specific Human Capital Approach

The specific human capital approach (e.g. Becker (1962), Jovanovic (1979a), Mortensen (1978) and Parsons (1972)) highlights a negative relationship between investment in specific human capital and the likelihood of job change. Workers acquire specific human capital as tenure in a job increases. As these skills are not transferable to a new job, workers and firms share the costs and benefits of this investment. The acquisition of this specific human capital raises productivity and consequently wages in the current job so the likelihood of a job separation decreases with tenure. In the short-term, we expect changing jobs to result in wage losses because of the non transferability of specific capital to another job.

However, as the rate of specific human capital accumulation declines with job tenure, wage growth will also decline as tenure on the job increases. Although existing specific capital is lost when changing jobs which can lead to a lower starting wage in an alternative job, there may be more opportunities for investment in specific human capital in an alternative job which means that the worker could have faster on-the-job wage growth than in the current job (Mortensen (1988)).

Other Approaches

Many modern theoretical models build on the models described above. However, there are alternative approaches. For example, in Lazear's (1986) raiding model, firms compete for high quality workers and they use workers' previous wages as an indicator of their quality. Consequently, high productivity workers experience more mobility than low productivity workers because the high paying firms will poach workers from other firms. This is contrary to the prediction of the mover-stayer model; in this raiding model mobility acts as a positive signal of productivity and leads to wage gains.

2.3 Patterns we should observe in the Data

Labour Market Conditions

Turnover rates should vary over the course of the business cycle. Shimer (2003) highlights the fact that vacancies are procyclical. Therefore, during an upturn there is an increase in vacancies and so there are more potential employment opportunities available to workers. In job search models, this would lead to an increase in the job offer arrival rate. In job matching models there is an increase in the number of alternative jobs a worker can switch to. In general, we would expect workers to have a higher probability of quitting when they have a good chance of obtaining a better job quickly. Therefore, when labour market conditions are tight we would expect to see more quits than when they are loose (e.g. Cornelissen, Hubler and Schneck (2007)). Conversely, layoff rates tend to be anti-cyclical; when demand falls employers will layoff workers.

Burgess, Lane and Stevens (2000) say that the relationship between job turnover and the business cycle is more complex. They argue and find empirical support for their hypothesis that in an upturn, where there is a surge in hiring by firms, there is also an increase in the number of workers whose productivity is unknown and this may lead to higher subsequent job separations. Similarly, downturns in activity provide employers with an opportunity to shed their least productive workers, thus reducing the need for subsequent adjustments in their workforce.

Other studies focus on the effect of labour market conditions at the time of labour market entry for workers; the impact this has on wages and how it affects job mobility (e.g. Oreopoulos, von Wachter and Heisz and (2008) and Bachmann, Bauer and David (2010)).

Cross Country Differences

Borghans and Golsteyn (2011) document differences in mobility patterns across countries for college graduates. They find that graduates in the US are more mobile than European graduates and that there are large differences in job mobility within Europe; graduates in Norway and the Netherlands are more mobile than those in France, Sweden and Germany. They also find that mobility rates in Japan are close to the European average mobility rate. We would expect to see higher mobility rates in countries with more flexible labour markets. Several studies highlight differences in labour market institutions and regulations across countries which may help explain differences in mobility rates and wage changes (e.g. Davia (2005), Dustmann and Pereira (2008), Ibsen, Trevisan and Westergaard-Nielsen (2008) and Pavlopoulos, Fouarge, Muffels and Vermunt (2007)).

Tenure, Age and Experience

Age is an important factor determining job mobility and turnover declines with age. In Stigler (1962) younger workers are more likely to try a variety of jobs in order to acquire knowledge of the labour market and their own preferences and ability for different jobs (a process known as "job shopping"), so we expect to see higher mobility rates for younger workers. In job search models, as workers gain labour market experience, they have more opportunities to search for, assess and accept superior job offers. Consequently, as experience increases so does the worker's reservation wage for changing jobs so the probability of job mobility declines with experience. This is supported empirically by numerous studies. For example, Topel and Ward (1992) find that for young men, two thirds of their total lifetime job mobility occurs within the first ten years of their career. They see job mobility for young workers as a crucial phase in workers' movement to more long-term stable employment relationships. Booth, Francesconi and Garcia-Serrano (1999) use retrospective work-history data from the British Household Panel Survey to study mobility over the period 1915 to 1990 and they find that on average workers hold five jobs over the course of their working lives and that half of all lifetime job changes occur within the first ten years of labour market entry.

The probability of job mobility also declines with tenure. In job matching models, when the quality of the match is revealed, workers in a successful match may be rewarded with higher wages or match specific rents. If tenure indicates the existence of a successful match then these rents may reduce job mobility for workers with longer tenures.[6] In human capital models, the relative value of an existing job, in terms of productivity and wages, increases with tenure because of specific human

[6] This negative relationship between job mobility and job tenure is usually evident using annual data. However, using more frequent data can show there is an increase in the hazard of a job ending in the first few months of an employment relationship (e.g. Farber (1994)).

capital accumulation. Therefore, as workers acquire specific human capital the probability of turnover is reduced.

In addition, Groot and Verberne (1997) argue that mobility is likely to be higher for younger people or for those with less labour market experience or less tenure due to the presence of mobility costs. There are both financial and psychological costs to changing jobs. Older people are more likely to have made investments in housing and be more settled or attached to their environment. Therefore, the costs of changing jobs are likely to be higher for older people, especially if changing jobs involves moving house. Workers with longer tenure are also likely to have higher psychological costs in changing jobs. To the extent that longer tenure reflects high quality matches, these workers may feel a stronger attachment to their organisation and colleagues. In addition, even if the costs associated with changing jobs are the same for younger and older people, younger people have more time before retirement to make up these costs. Workers change jobs if the expected utility from doing so exceeds the costs. If the gains involved in changing jobs put a worker on a higher wage path, younger workers will benefit for longer from these gains. Finally, older workers may have higher time preferences and therefore apply a higher discount rate on future earnings so job mobility declines with age.[7]

Gender

Central to why we might expect differential mobility rates by gender is that women have a lower attachment to the labour force. Barron, Black and Loewenstein (1993) develop a job-matching model where workers differ in their attachment to the labour force. The model predicts that those with a weaker attachment to the labour force are sorted into jobs that offer less training and that use less capital and as a result have less to lose by changing jobs in terms of specific capital. On the other hand, women's mobility decisions may be more constrained by nonmarket variables such as their partner's location or the rearing of children. In addition, women spend more time on household tasks than men (Hersch and Stratton (1997)). Therefore the opportunity cost of job search may be higher for women. Empirically, several studies have found that by controlling for characteristics, such as labour market experience, gender

[7] While there are theoretical arguments supporting the importance of variables such as age, experience and tenure, empirically, they tend to be correlated with each other, which may make the identification of separate effects difficult.

differences in turnover rates diminish or disappear (e.g. Blau and Kahn (1981) for worker-initiated separations in the US and Booth and Francesconi (2000) for worker-initiated separations in the UK).

However, there are instances where we expect to see differences in job change rates. For example, within voluntary mobility, women may be more likely to make changes for non-job related reasons, for example they may be more motivated to change jobs to help them combine their professional and family life (e.g. Keith and McWilliams (1999)).

Education

There are several reasons to expect a relationship between education and job mobility but there is no consensus in the literature as to whether it is positive or negative. Barron *et al.* (1993) argue that education may qualify workers for high training jobs or capital-intensive jobs and so incentives are offered to decrease the expected number of quits for better-educated workers. Connolly and Gottschalk (2006) observe that less educated workers may invest less in human capital and consequently have less to lose by changing jobs. They will therefore have a lower reservation value when approached with an alternative job offer. Weiss (1984) suggests that there is an unobservable characteristic, which he calls "stick-to-itiveness", that affects both the value of education and the value of staying in an existing job.

Neal (1999) proposes a model of job search that involves both employment matches and career matches. He argues that less educated workers are likely to experience more job turnover because they experience mobility that involves career change and then they search for a good employment match. Therefore, it is possible that the process of finding a good career match may add considerably to the wage growth of younger workers, especially the less educated. To the extent that better educated workers (especially those with college degrees) use time spent in education as a form of pre-market search, they are less likely to experience mobility that involves career changes.

However, it is also possible that there could be a positive relationship between education and mobility. Weiss (1984) argues that education increases workers' alternative opportunities and so may increase job mobility. Johnson (1979) argues that higher wage variance may increase the option value of job mobility, so highly educated workers may experience more job turnover as they face more variable but potentially more rewarding alternative job offers. In addition, Greenwood (1975) contends that highly educated individuals may be more efficient job searchers and so have lower transactions costs and therefore may change jobs more easily. It is possible that better educated workers are more likely to have 'faster' careers and will change jobs more frequently as a means of advancing up the career ladder (Borsch-Supan (1987)). Finally, Bartel and Lichtenberg (1987) put forward the idea that highly educated workers have a comparative advantage in learning and implementing new technologies and so firms may provide incentives to reduce job quits.

2.4 Empirical Findings on Wage Impacts

Many empirical studies of the wage changes associated with job mobility find a positive effect on wage growth of around 10 per cent. This result is reasonably robust across countries. Campbell (2001) finds that the wage gain associated with changing job over a three-year period is around 10 per cent in the UK. Abbott and Beach (1994) find that the average wage gain for Canadian women who change jobs is around 8-9 per cent. Topel and Ward (1992) report a 10 per cent return to mobility for young men in the US. OECD (2010) find an average of a 3 to 4 percentage point wage premium associated with changing jobs for a range of European countries. Their estimate for the Irish wage premium is higher at around 9 per cent.

Several studies (e.g. Light and McGarry (1998), Topel and Ward (1992)) do not distinguish between voluntary and involuntary mobility. There are several reasons why it may not be entirely useful to differentiate between them. In some instances, the distinction between the two may be somewhat arbitrary. For example, an employee in a firm that is in difficulty may be concerned about being laid off and so may search for and obtain alternative employment. This would count as a voluntary quit even though it is motivated by the risk of being laid off. In addition, if we count all employee initiated separations as voluntary, then voluntary quits include those caused by illness and family reasons, in addition to those due to finding a better job. Finally, very often a significant proportion of respondents in surveys either do not give a reason for why they left their previous job or they do so for a reason not included in the questionnaire. These people either have to be arbitrarily assigned as voluntary or involuntary movers or excluded from the analysis (Light and McGarry (1998)).

However, other studies have found the distinction between voluntary and involuntary mobility to be important when estimating wage impacts. A common result in the literature is that voluntary mobility leads to higher wage growth than not changing jobs or for involuntary moves. (e.g. Mincer (1986)). In addition, in many instances involuntary moves are associated with wage losses, especially when there is an intervening spell of unemployment between jobs (e.g. Garcia Perez and Rebollo Sanz (2005)). Different types of separations (within voluntary and involuntary changes) have been found to have differential wage impacts. For example, Bartel and Borjas (1981) and Keith and McWilliams (1999) find higher wage growth for voluntary separations that are job-related and not due to personal reasons. Keith and McWilliams (1997) find that being laid off has a smaller wage penalty associated with it than being fired. Keith and McWilliams (1999) also find that those who engaged in employed job search prior to separating from their employer receive higher returns to mobility.

There is consistent evidence in the literature that mobility related wage gains decrease with age as well as tenure (e.g. Bartel and Borjas (1981)). In fact most of the empirical literature focuses on younger workers because job mobility is more common in the earlier stages of individuals working lives. For example, Topel and Ward (1992) find that job mobility accounts for one third of total wage growth for men in their first ten years in the labour market.

Other studies argue that the effect of changing jobs on wage growth depends on the position in the wage distribution. For example, Pavlopoulos *et al.* (2007) find that job mobility leads to a wage increase for low-paid workers but not for high-paid workers.

Other studies focus on gender differences in the returns to mobility and the extent to which this could account for the gender wage gap (e.g. Caparros Ruiz, Navarro Gomez and Rueda Narvaez (2004)). Gender differences in the impact of mobility may arise as women are more likely than men to separate from their jobs for family-related reasons (e.g. Keith and McWilliams (1997)). In addition, the opportunity cost of job search may be higher for women, especially married women, and so they may search less intensively than men for a better job match. Kahn and Griesinger (1989) argue that female job quits may be less responsive to wages than male ones because women place a higher value on non-monetary aspects of a job than men. Loprest (1992) finds that women only experience half the wage growth of men when changing jobs. However, Keith and McWilliams (1997) do not find gender differences in the returns to mobility once the reason for job separation is controlled for.

Another strand of the literature focuses on the effect of prior mobility on current wages. Keith (1993) argues that highly mobile workers generate greater turnover costs for employers so they view people with a high probability of turnover as undesirable and so there may be "reputation effects" associated with a high mobility past. Mincer and Jovanovic (1981) aggregate voluntary and involuntary changes and they find that prior mobility does not affect current wages. However Keith (1993) finds that each voluntary separation increases wages by 2 per cent while each involuntary quit reduces wages by a similar magnitude so that aggregating the two types of mobility disguises the offsetting impact of each. Light and McGarry (1998) also find that persistent mobility leads to a lower wage.

3. Job Mobility in Ireland

3.1 Introduction

The focus of this chapter is to investigate the various factors that determine job-to-job mobility in Ireland. The dataset used covers most of the Celtic Tiger period, a time where growth in the Irish economy was exceptional, and the chapter addresses the effect the changing labour market had on job mobility. Job mobility is an important phenomenon to understand because the movement of workers from one job to another allows for flexibility in the labour market by providing workers and firms with a mechanism to adapt to changing economic and personal circumstances. This churning can be seen as the efficient working of the labour market; as workers can seek out new employment matches that they are more productive in and for which they will be better rewarded.

Over the course of the 1990s the Irish economy experienced spectacular growth rates with GNP growth averaging 7.9 per cent per annum over the period 1995 to 2001. The success of the Irish economy over this period was built on factors that affected labour supply, such as the favourable demographic structure of the labour force, a dramatic rise in female participation rates and net immigration, particularly towards the end of the period, and accompanied by factors that affected the demand for labour such as foreign direct investment and competitiveness. Over this period, labour supply growth averaged 3.4 per cent per annum and employment increased by an average of 67,000 per annum (on a PES basis), implying that the number of jobs created over the period far exceeded the number of jobs that were destroyed. Existing research tells us that some of these jobs were filled by those returning to the labour market, particularly women (see Doris (2001)), and immigrants or returning nationals (see Barrett, Fitz Gerald and Nolan (2002)). Changing jobs appears to be an important part of worker's experience in the labour market yet little is known, in an Irish context, about how prevalent job changing is, as well as what types of worker are most likely to switch jobs and this chapter seeks to bridge that gap.

The chapter finds that labour market experience, working in the public sector, whether a person is overskilled, the sector they work in and their occupation are important determinants of voluntary job change. The chapter also finds the rate of voluntary job mobility in Ireland trebled over the period 1995 to 2000. It investigates the potential causes of this increase - is it simply driven by changes in the composition of workers or do other factors such as changes in the labour market conditions facing workers play a role? To do this, the sample is divided into two time periods and a decomposition technique is applied to ascertain how much of the increase in mobility is attributable to compositional changes and how much is due to other factors. Compositional changes explain around one-fifth of the increase, while the remainder seems to reflect fundamental changes in the operation of the labour market.

The chapter is organised as follows: Section 3.2 describes the dataset, the construction of key variables and provides some descriptive statistics. Section 3.3 presents the results of some multinomial probit models of job change and also discusses how we take account of changes in the labour market environment over the period. Section 3.4 outlines a decomposition technique that is used to ascertain the extent to which the increase in mobility is driven by changes in the composition of the sample. Section 3.5 concludes.

3.2 Dataset, Defining Job Mobility and Descriptive Statistics

3.2.1 Dataset and Sample Construction

The Living in Ireland Survey (LIS) is used to investigate the determinants of job change. The LIS constitutes the Irish component of the European Community Household Panel (ECHP) which began in 1994 and ended in 2001.[8] It involved an annual survey of a representative sample of private households and individuals aged 16 years and over in each EU member state, based on a standardised questionnaire. The fact that the same households were interviewed each year means that it is possible to study changes in characteristics or circumstances of individuals or households over

[8] Additional details on the structure of the LIS Survey and the sample design are available at: http://issda.ucd.ie/documentation/esri/lii-overview.pdf

time.[9] A wide range of information on variables such as labour force status, occupation, income and education level is collected. There is also data on job and firm characteristics.

Job mobility refers to the ability of workers to change jobs; in practice realised job changes are used as a proxy for job mobility. The panel dimension of the LIS is exploited to identify job changes. A revolving balanced panel of people aged 20 to 60, roughly the prime working age, is selected from the LIS. This means that individuals are included in the sample in every year that they meet this age restriction.[10] A revolving balanced panel is preferable to a pure balanced panel as a balanced panel prevents the entry of younger people into the sample and so, over time, as the fixed sample ages the proportion of younger people would decline.[11] Essentially, a revolving balanced panel allows younger people to enter into the sample and older people to leave the sample in later years. In addition, respondents must have completed the interview in each year in question. Each individual's labour force status is then categorised on a PES basis. Individuals, who are categorised as employed, are those who work or usually work at least 15 hours per week. Finally, around 120 cases are deleted from the sample each year; these cases refer to where the respondent is working but the start date with their employer (which is needed to capture job changes as described in Section 3.2.2) is missing in any year.

Table 3.1 presents the total sample size for each year and provides some basic characteristics of the sample. The average age of the sample declines over the period implying that the impact of the baby boom generation outweighs the effect of the ageing of the sample. The sample labour force participation rate appears high but it is measured as the number of people in the labour force aged 20 to 60 as a percentage of the total number of people age 20 to 60.[12] The male participation rate is significantly

---

[9] There was some attrition in the sample in the earlier years, although the representativeness of the sample was improved in 2000 with the addition of new households. These new entrants to the LIS sample have been excluded from the analysis.

[10] This approach to selecting a sample is similar to that of Baker and Solon (1999).

[11] For example, someone who is 20 in 1995 will be 26 in 2001 and if we only considered the same group of people over time (a balanced panel), there would be no one below the age of 26 in the panel by 2001.

[12] Using Central Statistics Office (CSO) data, the participation rate for those aged 15 to 64 rose from 60 to 66 per cent over the period. This is below the participation rate of the sample given in Table 3.1, but it considers more younger and older people who are less likely to be in the labour force and so it is at least consistent with the rate given in Table 3.1.

above the female participation rate; however there is a dramatic rise in the female rate over the period. The table also shows that participation rates decline with age, as we would expect, and that the participation rates for those over the age of 30 increased between 1995 and 2001.[13]

Table 3.1: Revolving Balanced Panel of Individuals aged 20 to 60

	1995	1996	1997	1998	1999	2000	2001
Sample Size	2,417	2,367	2,338	2,299	2,294	2,314	2,357
less cases where starting date with employer is missing for workers	125	120	121	120	115	119	118
Revolving Balanced Panel	2,292	2,247	2,217	2,179	2,179	2,195	2,239
Average Age	42.2	42.0	41.7	41.4	41.0	40.5	40.1
	%	%	%	%	%	%	%
Participation Rate	64	65	67	69	71	72	73
Participation Rate: Male	90	89	90	90	91	90	90
Participation Rate: Female	40	42	46	50	53	55	57
Participation Rate: 20-29	79	81	81	81	84	78	81
Participation Rate: 30-39	71	72	73	77	76	78	77
Participation Rate: 40-49	65	67	71	72	76	77	75
Participation Rate: 50-60	48	48	50	53	55	58	60

3.2.2 Calculation of Job Mobility

To capture job changes we need to be able to identify those who have separated from their employer between waves. The LIS does not contain an explicit question about changing jobs. Instead job changes are captured using the information about when a worker reports that they started working with their current employer. In this chapter, job mobility is defined in terms of employment-to-employment transitions, so to capture this, workers need to be employed in two consecutive waves. Workers are asked to report the month and year that they started working with their current

[13] There is a 6-percentage point drop in the participation rate for people aged 20 to 29 between 1999 and 2000. This is explained by an increase in the proportion of younger people staying on in education, in particular those aged 20.

employer. The specific question asked in the survey is: '*When did you begin work with your present employer (or in your present business)? Please specify the month and the year*'. If the reported date is after their interview in one year but before their interview in the following year, this indicates that the person has changed jobs between the two waves. Other types of labour market transitions such as moving from being unemployed or not participating in the labour market to being employed are excluded from the analysis. For example, someone who is employed in one year and then unemployed for two years and then employed again is not included in the analysis. Even though this person has moved to a new job over the four-year period, they have moved from being employed to being unemployed for two years to being employed again. These types of transitions are excluded because the decision to change jobs is different to the decision to move from, say nonparticipation or unemployment to employment. This definition of job mobility only allows people to be unemployed or to not participate in the labour market for a relatively short amount of time between jobs, essentially less than a year (or more precisely less than the amount of time between interviews).

This measure of job mobility refers to a change of employers and so may be considered as a measure of external job mobility, there are other types of mobility that can take place within a firm, such as promotion etc but we cannot capture this type of internal mobility in the data.[14] In addition, this measure of job mobility may underestimate total mobility if more than one job change takes place between interviews. It is well known that that there is a high hazard of jobs ending within the first year of an employment relationship (e.g. Farber (1999)).

Individuals who are employed in successive two-year periods are selected from the revolving balanced panel. The resulting sample is one with workers who have a high attachment to the labour force. There are 1,817 people in the analysis and 8,976 person-year observations. Table 3.2 shows the number of workers employed in consecutive two-year periods and the rate of job change. Each year approximately 10 per cent of workers change jobs. However, this figure masks an important trend

[14] Several studies (e.g. Booth and Francesconi (2000)) have found this distinction to be important.

evident in the data. In 1995, 6.5 per cent of workers changed jobs and this rate increased over the period so that by 2000 the mobility rate was 13.4 per cent.

Table 3.2: Job Mobility Rate

	1995	1996	1997	1998	1999	2000	2001
Number of workers	1,163	1,175	1,211	1,276	1,341	1,376	1,434
No. Job Changes	76	85	102	139	146	184	156
Job Mobility Rate	6.5%	7.2%	8.4%	10.9%	10.9%	13.4%	10.9%

A total of 888 job changes are identified, however, some people changed jobs more than once so Table 3.3 shows the number of jobs held by the 1,817 workers between the beginning and end of the 7-year period.

Table 3.3: Number of Job Changes per Worker

0	1,254
1	359
2	121
3	54
4	20
5	9

To put Ireland in an international context, Table 3.4 shows average rates of job mobility for young workers over the period 1995 to 2001 across a range of European countries from Davia (2005). From the table we can see that young workers in Ireland have a relatively high rate of job mobility. The mobility rates reported in Table 3.4 refer to workers who were under the age of 30 in 1994 as the sample considered by Davia (2005) is restricted to younger people. This chapter focuses on workers aged 20 to 60. The mobility rate reported by Davia (2005) for Ireland is consistent with the mobility rate this chapter finds for younger people.

Table 3.4: Job Mobility Rates, Average between 1995-2001 for Workers under 30 in 1994

Germany	6%
Netherlands	8%
Austria	8%
Portugal	9%
Belgium	10%
France	10%
Italy	10%
Greece	13%
Ireland	16%
UK	19%
Finland	22%
Spain	23%

Source: Davia (2005), estimates derived from the European Community Household Panel Survey.

The LIS also asks the main reason for the previous employment relationship ending. This allows us to identify worker initiated or voluntary quits such as obtaining a better job, family-related quits etc and employer related or involuntary quits such as redundancy, dismissal, business closure etc. It may be important to distinguish between voluntary and involuntary job turnover as the reason for job separation is likely to have different impacts on subsequent wage growth.[15] Table 3.5 gives the main reason why job changers stopped working in their previous jobs. In each year the bulk of job changes were voluntary, with 49 per cent of job changes being voluntary in 1995 rising to 65 per cent in 2001.[16] In 1995, 33 per cent of mobility was involuntary and this tended to fall over the period so that by 2001 around 21 per cent of all job changes were involuntary. Unfortunately, around 15 per cent of people who changed jobs each year did so for another reason that wasn't included in the questionnaire or they did not answer the question. These workers are excluded from the analysis that follows and Table 3.6 shows the number of workers employed in consecutive two-year periods and the rate of job change for the resulting sample. The

[15] Keith and McWilliams (1999) amongst others find differential rates of return to job mobility in the US depending on whether the reason for separation is voluntary or involuntary.
[16] Included in the 'Other Reasons Given' category in Table 3.5 are explanations such as childbirth or looking after children, looking after an old, sick or disabled person, that their partner's job required them to move to another place, study, or that the person became ill or disabled.

table shows that the voluntary job mobility rate trebled over the period 1995 to 2000, while the involuntary job mobility rate remained roughly constant over the time period.

Table 3.5: Reason for Stopping Previous Job

	1995	1996	1997	1998	1999	2000	2001
Voluntary Turnover:							
Got Better Job	46%	44%	43%	41%	47%	53%	56%
Other Reasons Given	3%	7%	11%	14%	16%	14%	9%
Involuntary Turnover:							
Obliged to Stop	11%	12%	12%	17%	8%	10%	10%
End of Contract	22%	24%	15%	12%	16%	12%	12%
Rest	18%	14%	20%	16%	14%	11%	13%

Table 3.6: Job Mobility Rate

	1995	1996	1997	1998	1999	2000	2001
Number of workers	1,149	1,163	1,191	1,254	1,321	1,355	1,413
No. Job Changes	62	73	82	117	126	163	135
Overall Job Mobility Rate	5.4%	6.3%	6.9%	9.3%	9.5%	12.0%	9.6%
Voluntary Mobility Rate	3.2%	3.7%	4.6%	6.1%	7.0%	9.1%	7.2%
Involuntary Mobility Rate	2.2%	2.6%	2.3%	3.2%	2.6%	3.0%	2.3%

3.2.3 Descriptive Statistics

This section examines some individual characteristics of workers and of those who change jobs. The aims are to identify differences in characteristics between those who change jobs and those who stay in their jobs and also to identify any compositional changes in the total number of workers that might help explain the rise in the rate of voluntary job change.

Age

The age distribution of all workers in the sample (from Table 3.6) is given in Table 3.7. The proportion of workers in the 20 to 29 age group increases over time, and the increase is more marked in 2000 and 2001, reflecting the fact that these younger people only have to be working for a relatively short period of time to be included in the sample. The proportion of workers in the 30-39 age group declines over the period, consistent with the ageing of the sample over time. The proportion of the workers in the 40 to 49 age group increases up to 2000, again indicating the ageing of the sample.[17] The proportion of workers between 50 and 60 declines slightly over the period because the impact of people dropping out of the sample at 60 slightly dominates the effect of ageing. There is a slight decrease in the average age of workers over the period due to the impact of the 'baby boom' generation.

Table 3.7: Age Distribution of Workers & Job Change Rate by Age Group

	1995	1996	1997	1998	1999	2000	2001
	Age Distribution of Workers						
20-29	17%	17%	19%	20%	22%	24%	24%
30-39	31%	30%	28%	27%	25%	23%	22%
40-49	26%	28%	29%	29%	30%	30%	29%
50-60	26%	25%	24%	23%	23%	23%	24%
Average Age	41.0	40.8	40.3	40.1	40.0	39.8	39.8
	Voluntary Job Change Rate by Age Group						
20-29	12%	11%	15%	15%	17%	21%	13%
30-39	2%	4%	3%	6%	7%	8%	9%
40-49	2%	2%	3%	3%	4%	4%	3%
50-60	0%	1%	1%	2%	2%	4%	5%
Average Age	28.8	32.1	30.0	32.7	30.9	31.8	34.1
	Involuntary Job Change Rate by Age Group						
20-29	3%	5%	4%	5%	3%	4%	3%
30-39	2%	2%	2%	4%	3%	3%	2%
40-49	2%	3%	2%	2%	3%	1%	2%
50-60	2%	1%	3%	3%	2%	3%	1%
Average Age	39.4	35.6	38.9	38.2	38.8	37.5	36.9

[17] The proportion declines slightly from 2000 to 2001. The numbers leaving to enter the older age group roughly cancels out the number of people entering this age group and because the number of people in the younger age group is increasing quite dramatically the share of the total accounted for by the 40 to 49 age group declines somewhat.

The table also shows the percentage of each age group who experience voluntary and involuntary mobility over time. From the table, we can see that the propensity to voluntarily change jobs declines with age and this finding is consistent with the empirical literature. The increasing proportion of young people aged 20 to 29 is, at least in part, driving the increase in the overall mobility rate. Interestingly, the mobility rates for workers over the age of 30, although somewhat volatile over the period, show quite large increases. For example, the rate of job change for those aged between 30 and 39 quadruples over the period, albeit from a much lower base than the comparable rate for workers aged between 20 and 29. Workers who change jobs are on average 8/9 years younger than the sample average. The table also shows that the rates of involuntary change are more evenly distributed across age groups, although those in the 20 to 29 age category experience the highest rate of involuntary job separations.

Gender

Table 3.8 shows the gender distribution of workers over time and the proportion of men and women who experience voluntary and involuntary mobility. Female workers account for a rising proportion of workers over time, capturing female workers who returned to the labour market over the period. In addition, female workers experience a higher rate of voluntary mobility than male workers. Both the male and female rates of voluntary job mobility increase over the period 1995 to 2000 with the female rate increasing at a faster pace. The female voluntary job separation rate is around 1 percentage point above the male rate so the changing gender distribution of workers may be contributing somewhat to the rise in the voluntary job mobility rate over the period.

Table 3.8: Gender Distribution of Workers & Job Change Rate by Gender

	1995	1996	1997	1998	1999	2000	2001
	Gender Distribution of Workers						
Male	69%	67%	66%	63%	63%	63%	61%
Female	31%	33%	34%	37%	37%	37%	39%
	Voluntary Job Change Rate by Gender						
Male	4%	3%	4%	5%	6%	8%	7%
Female	2%	4%	5%	9%	8%	10%	7%
	Involuntary Job Change Rate by Gender						
Male	2%	2%	3%	3%	3%	2%	2%
Female	3%	3%	1%	3%	1%	4%	2%

Education

The education distribution of all workers is shown in the top panel of Table 3.9. In the table low-skilled workers are those who have, at most, Junior Certificate education, medium-skilled are those who have, at most, a diploma and high-skilled are those with degrees.[18] From the table, an improvement in the educational attainment of workers is apparent with low-skilled workers accounting for a declining proportion of the total over time.

The table also shows the percentage of workers within various education groups who change jobs by type of change. Medium-skilled workers have a higher propensity to experience voluntary separations than low-skilled workers or high-skilled workers. The rise in the proportion of medium-skilled workers may be contributing to the rise in the voluntary mobility rate. Low-skilled workers have the highest rate of involuntary separations. On average 3 per cent of low-skilled workers experience involuntary mobility, while the comparable rates for medium-skilled and high-skilled workers are 2 per cent and 1 per cent respectively.

[18] There are between 1 and 7 cases each year where the answer to the educational attainment question is missing. For these people, their educational attainment is assigned on the basis of the age at which they left full time education.

Table 3.9: Education Distribution of Workers & Job Change Rate by Education Level

	1995	1996	1997	1998	1999	2000	2001
	Education Distribution of Workers						
Low-Skilled	49%	48%	48%	45%	44%	43%	41%
Medium-Skilled	37%	38%	39%	41%	42%	43%	46%
High-Skilled	14%	13%	13%	14%	15%	14%	14%
	Voluntary Job Change Rate by Education						
Low-Skilled	2%	3%	5%	5%	5%	6%	7%
Medium-Skilled	4%	6%	6%	8%	9%	11%	8%
High-Skilled	5%	3%	1%	5%	7%	10%	4%
	Involuntary Job Change Rate by Education						
Low-Skilled	3%	3%	3%	4%	4%	4%	3%
Medium-Skilled	2%	3%	2%	3%	2%	2%	2%
High-Skilled	1%	0%	1%	1%	2%	3%	2%

Occupation

The occupations workers have may also provide a measure of human capital or skills. The occupational distribution of workers and the propensity for workers in different occupations to change jobs is given in Table 3.10.[19] As job changes may also involve a change in occupation the data in the table refer to the occupation held in the previous year or in the previous job. The table shows that over the period there is generally some decline in the proportion of workers who are managers, professionals and skilled workers, while the proportion of workers in elementary occupations and clerks increases over the period. There is much more variability in the rates of job mobility by occupation than by education level. Clerks and those in elementary occupations have roughly double the rate of job change of managers, professionals and skilled workers. The changing occupational structure could be contributing to the overall increase in mobility. Over half of the job changes identified involve a change in occupation. Clerks and those in elementary occupations also experience a higher rate of involuntary separations.

[19] In the LIS, occupations are classified according to the International Standard Classification of Occupations, version 1988 (COM) 1-digit codes. In the tables the 'Manager' category comprises managers, senior officials and legislators; the 'Professional' category includes those working in the armed forces, professionals, technicians and associated professionals; the 'Clerks' category includes clerks, service, shop and sale workers; the 'Skilled' category comprises skilled agricultural or fishery workers and skilled craft or trade workers and finally the 'Elementary Occupations' category includes those in elementary occupations, plant or machine operators and assemblers.

Table 3.10: Occupational Distribution of Workers & Job Change Rate by Occupation

	1995	1996	1997	1998	1999	2000	2001
	Occupational Distribution of Workers						
Manager	12%	11%	10%	9%	8%	8%	10%
Professional	26%	26%	26%	25%	26%	24%	24%
Clerk	21%	21%	22%	23%	24%	26%	26%
Skilled	23%	23%	22%	21%	21%	21%	22%
Elementary	18%	19%	20%	22%	22%	21%	19%
	Voluntary Job Change Rate by Occupation						
Manager	4%	3%	3%	2%	6%	5%	6%
Professional	3%	4%	2%	4%	6%	7%	4%
Clerk	5%	5%	6%	10%	8%	13%	11%
Skilled	3%	2%	5%	2%	5%	6%	8%
Elementary	2%	5%	6%	11%	10%	12%	7%
	Involuntary Job Change Rate by Occupation						
Manager	1%	2%	0%	1%	0%	0%	1%
Professional	1%	1%	2%	1%	1%	3%	2%
Clerk	2%	3%	2%	4%	3%	3%	2%
Skilled	2%	1%	1%	2%	3%	2%	3%
Elementary	5%	5%	6%	7%	5%	5%	4%

Sector

The share of workers in each sector is given in Table 3.11. The average shares over the period are broadly comparable to the employment shares from the Labour Force Survey and Quarterly National Household Survey, with the exception of the share employed in agriculture which exceeds the CSO data by around 5 percentage points and the share in market services which is around 5 percentage points lower than the CSO data.[20] The declining importance of agriculture in terms of its share in employment and the rising importance of market services are evident in the table. As with occupations, a job change may also involve changing sector, so the data in the table refers to the sectors workers were in the previous year or in their previous jobs.

[20] The market services sector comprises distribution, hotels and restaurants, transport, storage and communications, financial intermediation, real estate, renting and business activities and other services; the non-market services sector includes public administration and defence, education, health and social work.

There is considerable variability in job mobility by sector. Workers in construction and market services display the highest rate of job turnover, while those in non-market services and in the agricultural sector are least likely to change jobs. A similar pattern holds for involuntary mobility.

Table 3.11: Sectoral Distribution of Workers & Job Change Rate by Sector

	1995	1996	1997	1998	1999	2000	2001
	Sectoral Distribution of Workers						
Agriculture, Mining & Utilities	16%	16%	13%	13%	12%	11%	10%
Manufacturing	18%	18%	19%	21%	19%	19%	19%
Construction	7%	8%	7%	8%	8%	9%	9%
Market Services	33%	32%	35%	33%	36%	36%	38%
Non Market Services	26%	26%	26%	26%	25%	25%	24%
	Voluntary Job Change Rate by Sector						
Agriculture, Mining & Utilities	1%	2%	2%	1%	2%	3%	1%
Manufacturing	2%	3%	6%	6%	6%	8%	8%
Construction	6%	7%	11%	6%	13%	16%	12%
Market Services	6%	6%	6%	9%	10%	12%	10%
Non Market Services	1%	2%	2%	4%	4%	5%	3%
	Involuntary Job Change Rate by Sector						
Agriculture, Mining & Utilities	1%	2%	1%	3%	2%	1%	2%
Manufacturing	1%	1%	1%	3%	2%	2%	1%
Construction	12%	10%	5%	1%	2%	5%	6%
Market Services	3%	2%	3%	4%	3%	3%	2%
Non Market Services	1%	2%	2%	3%	3%	3%	2%

From the preceding analysis age, occupation and sector appear to be important in explaining job changes. The following section explores the factors that determine job change more formally. The increase in voluntary job mobility over the period may be driven by changes in the composition of the sample, or, it may be related to the rapid output and employment growth observed over the period and we try to capture this effect in the next section.

3.3 Determinants of Job Change

Table 3.12 reports the marginal effects from a multinomial probit model examining the factors that determine voluntary quits and involuntary changes relative to the base of staying in the same job with the same employer. The coefficient estimates from a probit regression pooling both types of mobility are used as starting values for the multinomial probit model. The data for 1995 to 2001 have been pooled so there are 8,615 observations from which 506 voluntary job changes and 224 involuntary job changes have been identified.[21] The explanatory variables are defined in Appendix Table 3.1. All the explanatory variables are lagged by one year so they refer to the workers' characteristics and situation in the previous year or in their previous jobs.

[21] The number of observations is lower than reported in Table 3.6 as some observations are excluded because data is missing or not available for at least one of the explanatory variables.

Table 3.12: Multinomial Probit Model of Job Mobility*

Variable	*Marginal Impact*	*P>|Z|*	*Marginal Impact*	*P>|Z|*
	Voluntary Mobility		Involuntary Mobility	
Experience	-0.0059	0.00	-0.0020	0.00
Experience squared	0.0001	0.00	0.0000	0.01
Female	-0.0026	0.86	0.0122	0.06
Children	0.0028	0.62	0.0013	0.71
Living in a Couple	0.0095	0.22	0.0025	0.60
Female*Living in a Couple	-0.0181	0.02	-0.0145	0.01
Education (Ref: low):				
Education- medium	-0.0058	0.18	-0.0091	0.01
Education- high	-0.0060	0.33	-0.0137	0.02
Working Part-Time	-0.0057	0.99	0.0561	0.00
Female*Working Part-Time	0.0232	0.17	-0.0159	0.00
Recent Training	0.0285	0.00	0.0027	0.39
Public Sector	-0.0236	0.00	0.0008	0.86
Number of Employees > 50	-0.0141	0.00	-0.0053	0.07
Overskilled	0.0166	0.00	0.0090	0.00
Occupation of Origin:				
(Ref: Elementary Occ's)				
Manager	-0.0160	0.02	-0.0175	0.00
Professional	-0.0147	0.02	-0.0157	0.00
Clerk	-0.0114	0.04	-0.0148	0.00
Skilled	-0.0167	0.00	-0.0151	0.00
Sector of Origin:				
(Ref: Non Market Services)				
Agriculture, Mining & Utilities	-0.0273	0.00	-0.0054	0.28
Manufacturing	-0.0086	0.28	-0.0080	0.20
Building	0.0254	0.02	0.0277	0.01
Market Services	0.0101	0.15	0.0050	0.36
Year Dummies:				
(Ref: 1995)				
1996	0.0060	0.48	0.0020	0.64
1997	0.0085	0.35	-0.0024	0.73
1998	0.0289	0.00	0.0058	0.16
1999	0.0332	0.00	0.0007	0.60
2000	0.0509	0.00	0.0050	0.12
2001	0.0353	0.00	0.0004	0.62
N		8,615		
Wald chi2		611.10		
Prob > chi2		0.0000		
Log pseudolikelihood		-2543.4988		

*Notes: Standard errors are adjusted to take account of the fact that there are multiple observations on the same people. The marginal effects are computed at the sample means of the explanatory variables.

Beginning with the results for voluntary mobility, the signs and significance of the marginal effects are, in general, what would be expected. The marginal effect of experience is negative and highly significant implying that for a worker with mean characteristics, an additional year of experience is associated with a 0.6 percentage point decrease in the probability of changing jobs.[22] Experience may have a non-linear effect on the probability of changing jobs so, to capture the fact that job changes are more likely to occur early in one's career, a squared term is included in the specification. The positive effect on the experience-squared variable implies that as years of experience increase the predicted probability of changing jobs decreases at a diminishing rate.

The model contains a range of individual controls that include household structure and personal characteristics. The marginal effect on gender is small and insignificant implying that there are no gender differences in the probability of experiencing voluntary mobility. Looking at household structure, workers who are married or living in a couple are more likely to change jobs but the effect is not significant. If people are constrained by their partner's job we might expect the effect to be bigger for women. The results show that women who are married or living in a couple are 1.8 per cent less likely to experience voluntary mobility. The marginal effect on the children variable is small and insignificant implying that having children does not affect the probability of changing jobs.[23] This is somewhat surprising but may partly be explained by the fact that the sample considers people who have a high attachment to the labour force.

The education variables capture general human capital. The marginal effects of higher levels of education are small and insignificant implying that education does not affect the probability of voluntary mobility. Booth and Francesconi (2000) find a similar result for the UK. The occupation variables may also serve as a proxy for human capital and the negative effects on the occupations of origin variables imply that people in occupations that embody more human capital than the base category

[22] Age was also included in the specification but as age and experience are highly correlated the model did not support the inclusion of both. Experience is used in the final specification because the resulting model has a better fit.

[23] Alternative formulations of this variable such as including the number of children and specific age groups of children were examined. A gender and children interaction term was included but was dropped because it was not significant.

(elementary occupations) are less likely to change jobs. In addition, workers who have undergone recent training are more likely to change jobs. This may reflect the fact that, typically, training is undertaken at the beginning of a job and there is a high hazard of new jobs ending early.

A range of variables to capture job and firm characteristics are also included in the model. We would expect a positive relationship between working part-time (less than 30 hours per week) and job mobility as part-time workers are less attached to the labour force etc and we may expect there to be differences by gender. However, the results do not support either of these hypotheses. A variable to capture overskilling, meaning that workers report they have skills and qualifications necessary to do a more demanding job, is included in the analysis as overskilling may indicate a poor job match. Workers who report that they are overskilled, have a higher probability of changing jobs. In addition, a firm size effect is included to capture the fact that those working in large firms may be less likely to change jobs because they have more alternative opportunities within the firm. The results indicate that workers in firms with more than 50 employees are 1.4 per cent less likely to change jobs.

Workers in the public sector have a lower probability of changing jobs. The results show that workers in the public sector are 2.4 per cent less likely to change jobs and the effect is highly significant. The model results also show that workers in the building and market services sector are 2.5 per cent and 1 per cent respectively more likely to change jobs relative to workers in the nonmarket services sector. Workers in the agricultural, utilities and manufacturing sectors are less likely to change jobs than those in the nonmarket services sector.

The year dummies are used to control for factors that vary over time and that affect all workers. The coefficients on the year dummies are positive and significant (with the exception of the dummies for 1996 and 1997) implying that there is an increase in voluntary mobility in the later part of the period. It is likely that these year dummies are picking up the strong rise in economic and employment growth that took place towards the end of the 1990s. One would expect the mobility rate to be higher when the labour market is tight. Ideally, one would like to include a variable that captures the job offer arrival rate to workers over time. Vacancy rates may be a good proxy for this variable. Unfortunately, vacancy rates are not available for this period. Table 3.13 reports the results for a probit model of voluntary job changes (involuntary changes are dropped from the models reported in the table). The first model in Table 3.13 is a standard probit model of voluntary mobility and the second model includes the unemployment rate, instead of the year dummies, as an indicator of labour market tightness. This variable is included to try to capture the changes in labour market conditions over the period. Essentially, lower unemployment rates may signal to workers that jobs are more plentiful and that job search is likely to result in an alternative to their current job. The marginal effect on the unemployment rate is negative as expected and significant. The second model which includes the unemployment rate has a better fit so it is the preferred model.

The results in Table 3.12 suggest that there are some notable differences in the effects of characteristics on the probability of job mobility when we distinguish between types of mobility. The marginal effect of experience on involuntary mobility is negative, but the effect is more muted than for voluntary mobility. To the extent that experience and tenure are correlated, the negative impact of experience may indicate that employers operate a 'last in first out' policy towards layoffs. However the smaller effect of experience on involuntary mobility may mean that workers undergo involuntary mobility throughout their careers, not just in the earlier years. Education has a significant impact on the probability of involuntary job mobility. Workers with higher levels of education are less likely to experience involuntary mobility. To the extent that education acts as a positive signal of productivity, employers that are shedding jobs may be less likely to layoff better educated workers, as they may be harder to replace if the business recovers. Campbell (1997) also finds a significant negative education gradient for involuntary mobility in the United States. The results for occupation are broadly similar for both types of mobility.

The household/family variables have similar effects on involuntary mobility. Although the marginal effect on the gender dummy changes sign, it is only significant at the 10 per cent level. The results also show a strong positive relationship between involuntary mobility and working part-time. However, women who work part-time are less likely to experience involuntary mobility. The effects of firm size, being overskilled and recent training are all smaller for involuntary moves. The effect of sector is smaller for involuntary changes with the exception of working in the construction sector where the effect is slightly larger than for voluntary mobility. We would expect that workers in the public sector would make fewer involuntary changes as the public sector is relatively sheltered, in the sense that it is less exposed to market forces. Somewhat surprisingly, working in the public sector does not affect the probability of an involuntary change. Finally, the time dummies are smaller and insignificant for involuntary moves. We would expect the impacts to be negative as firms are less likely to layoff workers when demand is high. However, due to the tightness in the labour market at this time employers may have been more tolerant as workers were harder to replace. The marginal effects on the time dummies imply there was no significant change in forced moves over the period.

Table 3.13: Probit Models of Voluntary Job Mobility*

Variable	*Marginal Impact*	*P>|Z|*	*Marginal Impact*	*P>|Z|*
	Specification 1		Specification 2	
Experience	-0.0059	0.00	-0.0060	0.00
Experience squared	0.0001	0.00	0.0001	0.00
Female	-0.0010	0.89	-0.0011	0.89
Children	0.0027	0.65	0.0028	0.64
Living in a Couple	0.0100	0.20	0.0103	0.19
Female*Living in a Couple	-0.0194	0.02	-0.0198	0.02
Education (Ref: low):				
Education- medium	-0.0067	0.19	-0.0068	0.19
Education- high	-0.0065	0.41	-0.0064	0.42
Working Part-Time	-0.0059	0.59	-0.0062	0.58
Female*Working Part-Time	0.0286	0.08	0.0293	0.08
Recent Training	0.0285	0.00	0.0278	0.00
Public Sector	-0.0243	0.00	-0.0243	0.00
Number of Employees > 50	-0.0143	0.00	-0.0146	0.00
Overskilled	0.0171	0.00	0.0172	0.00
Occupation of Origin:				
(Ref: Elementary Occ's)				
Manager	-0.0161	0.03	-0.0167	0.02
Professional	-0.0146	0.03	-0.0148	0.03
Clerk	-0.0116	0.07	-0.0117	0.07
Skilled	-0.0166	0.01	-0.0172	0.01
Sector of Origin:				
(Ref: Non Market Services)				
Agriculture, Mining & Utilities	-0.0278	0.00	-0.0278	0.00
Manufacturing	-0.0100	0.26	-0.0101	0.25
Building	0.0247	0.06	0.0247	0.06
Market Services	0.0090	0.23	0.0087	0.25
Year Dummies:				
(Ref: 1995)				
1996	0.0064	0.48		
1997	0.0077	0.38		
1998	0.0280	0.00		
1999	0.0319	0.00		
2000	0.0506	0.00		
2001	0.0346	0.00		
Unemployment Rate			-0.0036	0.00
N	8,391		8,391	
Wald chi2	430.42		427.44	
Prob > chi2	0.0000		0.0000	
Pseudo R2	0.1572		0.1557	
Log pseudolikelihood	-1611.0752		-1613.7926	

*Notes: Standard errors are adjusted to take account of the fact that there are multiple observations on the same people. The marginal effects are computed at the sample means of the explanatory variables.

3.4 Decomposing the Increase in the Rate of Voluntary Job Change

The voluntary job mobility rate trebled over the period 1995 to 2000. It is useful to ascertain whether this increase is simply driven by changes in the composition of the sample or whether it is due to other factors. One approach to doing this is to group some of the earlier years and some of the later years of the sample together and to decompose the difference in mobility rates between the two groups into the difference attributable to differences in the observable characteristics and the difference due to differences in the effects of characteristics by applying a non-linear Blinder-Oaxaca type decomposition to the estimates. This decomposition is important as it may help our understanding of the extent to which the nature of the Irish labour market itself changed over the period.

3.4.1 Non-Linear Decomposition Technique

I have grouped together the observations for 1995 to 1997 and for 1998 to 2001 as the marginal effects of the time dummies for voluntary mobility from the multinomial probit model are only significant from 1998 on.[24] There are 3,321 observations in the 1995-97 group and the average mobility rate is 3.9 per cent while there are 5,070 observations in the 1998-01 group and the average mobility rate is 7.4 per cent. There is a 3.5 percentage point difference in average mobility rates between the two groups. To decompose the gap between the two mobility rates, a technique developed by Fairlie (2005) is applied. The approach follows that of the Blinder-Oaxaca decomposition technique for linear models.

Consider the general case where the expected value of the dependent variable is a function of a linear combination of independent variables where the function F may or may not be linear:

$$E(Y) = F(X\beta) \tag{3.1}$$

where Y is an N × 1 vector, X is an N × K matrix of independent variables, β is a $K \times 1$ vector of coefficients and N is the sample size.

[24] Involuntary job changes are excluded from this part of the analysis as the rate of involuntary mobility is roughly constant over the period.

From (3.1) the general expression for the mean difference in the expected value of Y between two groups, say A and B, can be written as:

$$\overline{Y}^A - \overline{Y}^B = \left[\overline{F\left(X^A\hat{\beta}^A\right)} - \overline{F\left(X^B\hat{\beta}^A\right)}\right] + \left[\overline{F\left(X^B\hat{\beta}^A\right)} - \overline{F\left(X^B\hat{\beta}^B\right)}\right] \tag{3.2}$$

The first term in the brackets in (3.2) represents the part of the difference in the expected value of Y for the two groups that is due to differences in the distribution of the independent variables between the two groups; this is referred to as the "explained" component. The second term in the brackets represents differences in the processes that determine Y for the two groups.

In a linear regression model $E(Y) = F(X\beta) = X\beta$, the effect of X is constant, so

$$\overline{Y} = \overline{F\left(X\hat{\beta}\right)} = \overline{X\hat{\beta}} = \overline{X_1\hat{\beta}_1 + X_2\hat{\beta}_2 \ldots}$$

$$= \frac{\sum_{i=1}^{N}\left(X_{1i}\hat{\beta}_1 + X_{2i}\hat{\beta}_2 + \ldots\right)}{N} = \overline{X_1}\hat{\beta}_1 + \overline{X_2}\hat{\beta}_2 + \ldots \tag{3.3}$$

where i=1….N is the number of cases.

Using the expression for the general decomposition given in (3.2) yields the standard Blinder-Oaxaca decomposition:

$$\overline{Y}^A - \overline{Y}^B = \left[\left(\overline{X}^A - \overline{X}^B\right)\hat{\beta}^A\right] + \left[\left(\hat{\beta}^A - \hat{\beta}^B\right)\overline{X}^B\right] \tag{3.4}$$

In a non-linear regression model, such as a probit model, the effect of X is not constant i.e. $\frac{dE(Y)}{dX_K} = f(X\beta)\beta_K$, the marginal effect of k varies with the level of X and the other variables in the model so $\overline{Y} = \overline{F\left(X\hat{\beta}\right)} \neq F\left(\overline{X}\hat{\beta}\right)$. In this case:

$$\overline{Y} = \overline{F\left(X\hat{\beta} \right)} = \overline{F\left(X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + ... \right)}$$

$$= \frac{\sum_{i=1}^{N} F\left(X_{1i}\hat{\beta}_1 + X_{2i}\hat{\beta}_2 + ...\right)}{N} \tag{3.5}$$

Therefore we can write:

$$\overline{Y}^A - \overline{Y}^B = \left[\sum_{i=1}^{N^A} \frac{F\left(X_i^A \hat{\beta}^A\right)}{N^A}\right] - \left[\sum_{i=1}^{N^B} \frac{F\left(X_i^B \hat{\beta}^B\right)}{N^B}\right] \tag{3.6}$$

Fairlie suggests a decomposition for a non-linear regression equation, which can be written as:

$$\overline{Y}^A - \overline{Y}^B = \left[\sum_{i=1}^{N^A} \frac{F\left(X_i^A \hat{\beta}^A\right)}{N^A} - \sum_{i=1}^{N^B} \frac{F\left(X_i^B \hat{\beta}^A\right)}{N^B}\right] + \left[\sum_{i=1}^{N^B} \frac{F\left(X_i^B \hat{\beta}^A\right)}{N^B} - \sum_{i=1}^{N^B} \frac{F\left(X_i^B \hat{\beta}^B\right)}{N^B}\right] \tag{3.7}$$

Again, the first term in the brackets provides an estimate of the overall contribution of the independent variables to the gap in mobility rates and the second term represents the unexplained component. As with the standard Blinder-Oaxaca decomposition one can use the coefficients from Group A as weights for the first term in the decomposition or the coefficients from a pooled sample of the two groups or one can re-write the decomposition to use the coefficient estimates from Group B.

Fairlie focuses on the first part of the decomposition, which estimates the overall contribution of the independent variables to the difference in the average value of the dependent variable. The change in the average value of Y is calculated by replacing the distribution of all independent variables from Group A with the distributions of all the independent variables from Group B.

The contribution of each independent variable to the overall change in the average value of the dependent variable is calculated by separately replacing the distribution of each independent variable from Group A with its distribution from Group B while holding the distribution of the other variables constant. Suppose, first of all, that the sample size of both groups is the same. Then the contribution of variable X_1 to the change in the average value of Y is given by:

$$\frac{1}{N^A}\sum_{i=1}^{N^A} F\left(\hat{\beta}_0^A + X_{1i}^A\hat{\beta}_1^A + X_{2i}^A\hat{\beta}_2^A + \ldots\right) - F\left(\beta_0^A + X_{1i}^B\hat{\beta}_1^A + X_{2i}^A\hat{\beta}_2^A + \ldots\right) \qquad (3.8)$$

To calculate the contributions of individual independent variables there needs to be a one-to-one matching of observations from both groups. To generate this matching, each person in Group A is ranked according to their predicted probability and similarly for each person in Group B. Then the person with the highest predicted probability in Group A is matched with the person with the highest predicted probability in Group B and the person with the second highest predicted probability in Group A is matched with the person with the second highest predicted probability in Group B and so on.[25]

In practice, the sample sizes of both groups will seldom be the same so to calculate the contribution of individual independent variables to the gap Fairlie suggests taking a random sample of the larger group that is equal in size to the other group. Each observation in the subsample of the larger group and the full sample of the smaller group is separately ranked by their predictive probabilities and matched by their respective rankings as before. The decomposition estimates will depend on the randomly chosen subsample. Ideally, the results should approximate those from matching all of Group A to Group B. To achieve this, lots of random subsamples from the larger group should be chosen and each of these should be matched to the smaller sample. Then separate decompositions for each subsample should be computed and the average value of the separate decompositions can be used to approximate the results for the whole of the larger group.

[25] As the predicted probabilities are non-linear functions of the parameter estimates standard errors for the estimates are calculated using the delta method.

Table 3.14 presents the results of the non-linear decomposition of the difference in job mobility rates between the two periods.[26] The coefficient estimates from the pooled sample are used to calculate the decomposition.[27] The results are based on mean values of decompositions with 1,000 different subsamples. The table also shows the average values of the independent variables over the two time periods.

The difference in the average value of the independent variables accounts for around 21 per cent of the difference in job mobility rates over the two time periods. This means that the difference in mobility rates between the two time periods would be around 21 per cent lower if the people in the 1995-97 group had the same distribution of characteristics as the people in the 1998-01 group. In terms of individual characteristics, experience, occupation and working in the public sector are important contributors to explaining the difference in mobility rates between the two time periods. The standard errors on practically all of the individual contributions are high so we cannot say with a lot of confidence how important individual variables are. However, the standard error on the overall contribution of the independent variables is low. The results suggest that the changing composition of the sample is only driving around a fifth of the increase in job mobility over the period.[28]

[26] The software to implement the decomposition is from Jann, B. (2006), "Fairlie: Stata module to generate nonlinear decomposition of binary outcome differentials", available at http://ideas.repec.org/c/boc/bocode/s456727.html.

[27] Using the coefficient estimates from 1995-97 or 1998-01 in the decomposition produces similar results.

[28] Including year dummies in both sub-periods in the decomposition leads to the overall contribution of the independent variables rising to 31 per cent; however the standard error is high indicating that the overall contribution of the independent variables is insignificant.

Table 3.14: Non-Linear Decomposition of the Difference in Job Mobility Rates between 1995-97 and 1998-01 using the Fairlie Method

Sample used to estimate coefficients	**Pooled Coefficients**
Average Mobility Rate 1995-97	**0.0388**
Average Mobility Rate 1998-01	**0.0744**
Difference	**0.0355**
All Variables (Amount of Gap Explained)	**0.0075**
Standard Error	**0.0010**
% of Overall Gap Explained	**21.2%**

	Contribution	*P>|Z|*	$\overline{X}9597$	$\overline{X}9801$
Experience	0.0103	0.29	20.42	19.16
Experience squared	-0.0059	0.50	553.12	493.74
Female	0.0000	0.96	0.33	0.37
Children	-0.0002	0.74	0.59	0.55
Living in a Couple	-0.0016	0.30	0.73	0.67
Female*Living in a Couple	0.0005	0.73	0.22	0.23
Education (Ref: low):				
Education- medium	-0.0004	0.52	0.38	0.43
Education- high	0.0001	0.80	0.14	0.14
Working Part-Time	-0.0001	0.88	0.13	0.16
Female*Working Part-Time	0.0004	0.72	0.10	0.12
Recent Training	0.0000	0.97	0.08	0.07
Public Sector	0.0024	0.06	0.31	0.27
Number of Employees > 50	0.0002	0.76	0.36	0.34
Overskilled	-0.0013	0.01	0.48	0.47
Occupation of Origin:				
(Ref: Elementary Occ's)				
Manager	0.0009	0.11	0.11	0.09
Professional	0.0008	0.33	0.26	0.25
Clerk	-0.0006	0.32	0.21	0.25
Skilled	-0.0004	0.56	0.23	0.21
Sector of Origin:				
(Ref: Non Market Services)				
Agric. & Mining & Utilities	0.0010	0.23	0.15	0.12
Manufacturing	0.0000	0.99	0.19	0.20
Building	0.0011	0.26	0.07	0.08
Market Services	0.0002	0.80	0.33	0.36

In Section 3.2.3, the rising proportion of young people in the sample was put forward as a possible explanation for the rise in mobility. Including age and its square in the decomposition instead of the experience variables produces broadly similar results;

the overall contribution of the independent variables increases to 27 per cent (see Appendix Table 3.2). Finally, including the unemployment rate in the model increases the proportion of the gap explained to 70 per cent (see Appendix Table 3.3). However, the fall in the unemployment rate captures the changing labour market conditions facing workers and is not related to the changing composition of the sample.

3.5 Conclusions

This chapter has analysed job mobility in Ireland over the period 1995 to 2001 using data from the Living in Ireland Survey. It finds that there are several factors that determine mobility. Consistent with the theoretical and empirical literature in this area, years of labour market experience is a key determinant of voluntary job change. Workers in the public sector are less likely to change jobs and workers who are overskilled are more likely to change jobs. It finds that gender does not affect the probability of job mobility. Although human capital captured by education does not affect the probability of voluntary mobility, occupational level exerts a negative influence on job mobility. However, human capital captured by both education level and occupation significantly reduces the probability of experiencing involuntary mobility. In addition, somewhat surprisingly, working in the public sector does not reduce the probability of involuntary mobility.

The chapter also finds the rate of voluntary job mobility in Ireland trebled over the period. Estimation results show that workers were more likely to change jobs in the later part of the period. A decomposition analysis shows that around a fifth of this increase is driven by changes in the composition of the sample. The changing labour market conditions facing workers appear to be an important factor driving the increase. Even accounting for compositional changes and changes in the labour market, a substantial part of the increase in job mobility over the period remains unexplained. It may be that there has been an increase in job instability over the period, although this is not necessarily worrying as the increase in mobility was voluntary in nature. At the same time, worker preferences may also have changed over the period, with a decline in the importance of the idea of a “job for life”.

Appendix Table 3.1: Explanatory Variables: Definitions and Summary Statistics

Variable	*Description*	*Mean*	*Std. Dev.*
Experience	Number of years in employment	19.5	11.5
Experience Squared	Number of years in employment squared	513.2	508.6
Female	Dummy variable that takes the value 1 if female and zero if male	0.36	0.48
Child	Dummy variable that takes the value 1 if the person has children and zero otherwise	0.56	0.50
Couple	Dummy variable that takes the value 1 if the person is married or living in a couple and zero otherwise	0.69	0.46
Female*Couple	Dummy variable that takes the value 1 if the person is female and married or living in a couple and zero otherwise	0.22	0.42
Education- low (Reference Category)	Dummy variable that takes the value 1 if highest educational qualification is Junior Certificate and zero otherwise	0.46	0.50
Education- medium	Dummy variable that takes the value 1 if highest educational attainment is above Junior Certificate but below degree level and zero otherwise	0.41	0.49
Education- high	Dummy variable that takes the value 1 if highest educational qualification is a degree or above and zero otherwise	0.13	0.34
Part-Time	Dummy variable that takes the value 1 if the person works less than 30 hours per week and zero otherwise	0.15	0.36
Female*Part-Time	Dummy variable that takes the value 1 if the person is female and works less than 30 hours per week and zero otherwise	0.11	0.31
Recent Training	Dummy variable that takes the value 1 if the person has been in education or training in the past year and zero otherwise	0.08	0.27
Public	Dummy variable that takes the value 1 if the person was working in the public sector in the previous year and zero otherwise	0.28	0.45
Number of Employees	Dummy variable that takes the value 1 if the number of employees in the firm in the previous year is more than 50 and zero otherwise.	0.35	0.48

Overskilled	Dummy variable that takes the value 1 if the worker reported that they felt they had skills and qualifications to do a more demanding job and zero otherwise.	0.48	0.50
Occupation of Origin:			
Manager	Dummy variable that takes the value 1 if occupation of origin is manager, senior official or legislator and zero otherwise	0.10	0.29
Professional	Dummy variable that takes the value 1 if occupation of origin is professional, technician or associated professionals and zero otherwise	0.25	0.43
Clerk	Dummy variable that takes the value 1 if occupation of origin is clerk, service, shop or sale worker and zero otherwise.	0.23	0.42
Skilled	Dummy variable that takes the value 1 if occupation of origin is skilled agricultural or fishery worker or a skilled craft or trades worker and zero otherwise.	0.22	0.41
Elementary (Reference Category)	Dummy variable that takes the value 1 if occupation in the previous year is plant or machine operator or assembler, or elementary occupation and zero otherwise.	0.20	0.40
Sector of Origin:			
Agriculture, Mining & Utilities	Dummy variable that takes the value 1 if sector of origin is agriculture, fishing, mining or quarrying, or utilities and zero otherwise.	0.13	0.34
Manufacturing	Dummy variable that takes the value 1 if sector of origin is manufacturing and zero otherwise.	0.19	0.39
Building	Dummy variable that takes the value 1 if sector of origin is building and zero otherwise.	0.08	0.27
Market Services	Dummy variable that takes the value 1 if sector of origin is distribution, hotels and restaurants, transport, storage and communications, financial intermediation, or real estate, renting and business activities and zero otherwise.	0.35	0.48
Non-Market Services (Reference Category)	Dummy variable that takes the value 1 if sector of origin is education, public administration and defence or health and social work and zero otherwise.	0.25	0.43

Year Dummies:			
1995 (Reference Category)	Dummy variable that takes the value 1 if the year is 1995 and zero otherwise.	0.13	0.34
1996	Dummy variable that takes the value 1 if the year is 1996 and zero otherwise.	0.13	0.34
1997	Dummy variable that takes the value 1 if the year is 1997 and zero otherwise.	0.13	0.34
1998	Dummy variable that takes the value 1 if the year is 1998 and zero otherwise.	0.14	0.35
1999	Dummy variable that takes the value 1 if the year is 1999 and zero otherwise.	0.15	0.36
2000	Dummy variable that takes the value 1 if the year is 2000 and zero otherwise.	0.15	0.36
2001	Dummy variable that takes the value 1 if the year is 2001 and zero otherwise.	0.16	0.37
Unemployment Rate	ILO annual unemployment rate from the CSO	7.72	3.30

Appendix Table 3.2: Non-Linear Decomposition of the Difference in Job Mobility Rates between 1995-97 and 1998-01 using the Fairlie Method, including Age instead of Experience

Sample used to estimate coefficients	**Pooled Coefficients**	
Average Mobility Rate 1995-97	**0.0388**	
Average Mobility Rate 1998-01	**0.0744**	
Difference	**0.0355**	
All Variables (Amount of Gap Explained)	**0.0095**	
Standard Error	**0.0010**	
% of Overall Gap Explained	**26.9%**	
	Contribution	P>|Z|
Experience	0.0102	0.60
Experience squared	-0.0052	0.77
Female	0.0000	0.94
Children	-0.0001	0.91
Living in a Couple	-0.0006	0.69
Female*Living in a Couple	0.0003	0.69
Education (Ref: low):		
Education- medium	-0.0003	0.61
Education- high	0.0000	0.98
Working Part-Time	-0.0001	0.93
Female*Working Part-Time	0.0008	0.64
Recent Training	0.0000	0.89
Public Sector	0.0025	0.07
Number of Employees > 50	0.0002	0.81
Overskilled	-0.0011	0.04
Occupation of Origin:		
(Ref: Elementary Occ's)		
Manager	0.0011	0.07
Professional	0.0008	0.32
Clerk	-0.0006	0.30
Skilled	-0.0003	0.69
Sector of Origin:		
(Ref: Non Market Services)		
Agric. & Mining & Utilities	0.0010	0.36
Manufacturing	0.0000	0.97
Building	0.0010	0.29
Market Services	0.0002	0.87

Appendix Table 3.3: Non-Linear Decomposition of the Difference in Job Mobility Rates between 1995-97 and 1998-01 using the Fairlie Method, including the Unemployment Rate

Sample used to estimate coefficients	**Pooled Coefficients**	
Average Mobility Rate 1995-97	**0.0388**	
Average Mobility Rate 1998-01	**0.0744**	
Difference	**0.0355**	
All Variables (Amount of Gap Explained)	**0.0249**	
Standard Error	**0.0095**	
% of Overall Gap Explained	**70.1%**	
	Contribution	P>|Z|
Experience	0.0087	0.37
Experience squared	-0.0061	0.48
Female	0.0000	0.96
Children	-0.0002	0.72
Living in a Couple	-0.0016	0.29
Female*Living in a Couple	0.0004	0.78
Education (Ref: low):		
Education- medium	-0.0005	0.50
Education- high	0.0001	0.84
Working Part-Time	-0.0001	0.89
Female*Working Part-Time	0.0004	0.72
Recent Training	0.0000	0.91
Public Sector	0.0021	0.10
Number of Employees > 50	0.0002	0.77
Overskilled	-0.0012	0.02
Occupation of Origin:		
(Ref: Elementary Occ's)		
Manager	0.0008	0.13
Professional	0.0008	0.37
Clerk	-0.0007	0.29
Skilled	-0.0003	0.58
Sector of Origin:		
(Ref: Non Market Services)		
Agric. & Mining & Utilities	0.0008	0.28
Manufacturing	0.0000	0.98
Building	0.0010	0.30
Market Services	0.0002	0.79
Unemployment Rate	0.0203	0.07

4. The Performance of the Hausman *et al.* Estimator in Correcting for Misclassification in the Dependent Variable in Binary Choice Models

4.1 Introduction and Motivation

Discrete choice models, such as a probit model, are used when the dependent variable is a binary outcome or choice. Measurement error in a binary variable results in misclassification i.e. an observation is classified as a zero when the variable is truly a one, and vice versa. In a linear regression model classical measurement error in the dependent variable only affects the precision of coefficient estimates; however the same problem leads to estimates that are biased and inconsistent in a nonlinear model. Hausman, Abrevaya and Scott-Morton (1998) use Monte Carlo simulations to demonstrate that even small amounts of misclassification can lead to substantially biased parameter estimates in a probit model. They propose a modified estimator that can be used to correct for misclassification. The estimator provides consistent estimates of the coefficients as well as the extent of misclassification.

This chapter explores the performance of the Hausman *et al.* (1988) estimator and tries to extend their results by examining a range of different models. It analyses instances where the estimator performs well and others where it performs poorly. To do this a range of Monte Carlo simulations are run on models with different specifications and with different levels of misclassification. The coefficient estimates are also compared with those from an ordinary probit model. The aim of the chapter is to provide insights into when it is reasonable to use their estimator to control for misclassification.

The chapter is organised as follows: Section 4.2 begins by reviewing the effect of measurement error in a binary dependent variable. It then describes the estimator developed by Hausman *et al.* (1998) to control for misclassification. It also outlines some of the empirical applications that the estimator has been used in. Section 4.3 illustrates the identification of the model using the results from Monte Carlo simulations. Section 4.4 considers a range of changes and extensions to the basic model presented in Section 4.3 to try to determine situations where the estimator is appropriate to use to control for misclassification. It examines how the estimator performs when the effect of the explanatory variable in the model becomes weaker, the sample size is increased and decreased, the proportion of "1s" in the dependent variable is changed, when misclassification is asymmetric and when the explanatory variable in the model is dichotomous. It also compares the coefficient estimates from the Hausman *et al.* model with those from an ordinary probit model. Section 4.5 considers situations where the estimator performs poorly and explains how this can occur. Section 4.6 concludes and offers some practical suggestions for researchers using the estimator.

4.2 Binary Choice Model with Misclassification

4.2.1 Effect of Measurement Error in the Dependent Variable

In the classical linear regression model classical measurement error in the dependent variable does not have very serious consequences - the standard errors on coefficients will tend to be larger than they would have been if there was no measurement error. Consider the following model:[29]

$$\tilde{y}_i = x_i'\beta + \varepsilon_i \tag{4.1}$$

where i=1,……n and n is sample size, ε_i is an independently and identically distributed error term and all variables are measured as deviations from sample means.

Suppose that $\tilde{y}_i$ is measured with error so what we actually observe is:

$$y_i = \tilde{y}_i + v_i \tag{4.2}$$

where v_i is assumed to be independent of the covariates and ε_i

[29] Hausman (2001) discusses the effects of measurement error in dependent variables.

Inserting (4.2) into (4.1) yields:

$$y_i = x_i'\beta + \omega_i \qquad \text{where } \omega_i = \varepsilon_i + v_i \tag{4.3}$$

The effect of measurement error in y_i is an error term with increased variance since the new error term, ω_i, contains both the original error term, ε_i, and the measurement error, v_i. In this case, the OLS estimates of β will remain unbiased since ω_i is uncorrelated with x_i but will be measured with less precision.

In a non-linear regression model, such as a probit model, the effects of measurement error are more severe. As before, we can write the observed value as the sum of the true value and the measurement error, as follows:

$$z_i = \tilde{z}_i + u_i \tag{4.4}$$

where $\tilde{z}_i$ denotes the correctly measured binary variable, z_i is the mismeasured proxy and u_i is the measurement error. In this case, when $\tilde{z}_i = 1$, the variable z_i can only take on two values. It will be equal to one when it is correctly specified so $u_i = 0$ (i.e. there is no measurement error) or $z_i = 0$ when it is incorrectly specified and so $u_i = -1$. Therefore when $\tilde{z}_i = 1$, the mismeasured variable z_i can never overestimate the true value. Similarly, when $\tilde{z}_i = 0$, the variable z_i can never underestimate the true value; the measurement error u_i is always either 0 or +1. As a result, the measurement error u_i is negatively correlated with the true variable $\tilde{z}_i$. This can lead to coefficient estimates that are biased and inconsistent.

4.2.2 Standard Model of Misclassification

To address the problem of misclassification in discrete dependent variables, Hausman *et al.* (1988) propose a modified maximum likelihood estimator that provides consistent coefficient estimates as well as estimates of the extent of misclassification.[30] Consider the latent variable y_i^*:

$$y_i^* = x_i'\beta + \varepsilon_i \qquad \text{where i=1, 2 …n} \qquad \text{n = sample size} \qquad (4.5)$$

and ε_i is an i.i.d. error term

In later chapters, y_i^* will be a latent variable that represents the potential or tendency for a worker to change jobs. It is a continuous variable that is unobservable and is determined by a set of explanatory variables, x_i, in such a way that the larger the value of y_i^*, the greater the probability of some event occurring, such as changing jobs.

The true response (or what we would observe in the data if there was no measurement error), $\widetilde{y}_i$, is generated by the latent variable crossing the zero threshold i.e.

$$\begin{aligned} \widetilde{y}_i &= 1 \quad \text{if } y_i^* \geq 0 \\ &= 0 \quad \text{otherwise.} \end{aligned} \qquad (4.6)$$

Let F(.) denote the cdf of ε_i. The probability that an observation is truly equal to one is given by:

$$\begin{aligned} \Pr\left(\widetilde{y}_i = 1 \middle| x_i\right) &= \Pr(x_i'\beta + \varepsilon_i > 0) \\ &= \Pr(\varepsilon_i < x_i'\beta) \qquad \text{(if F(.) is a symmetric distribution)} \\ &= F(x_i'\beta) \end{aligned} \qquad (4.7)$$

while the probability that it is truly equal to zero is given by:

$$\Pr\left(\widetilde{y}_i = 0 \middle| x_i\right) = 1 - F\left(x_i'\beta\right) \qquad (4.8)$$

[30] The details of the model come from Hausman *et al.* (1998).

Now suppose the true response, $\tilde{y}_i$, is observed with error. Let y_i denote the actual response that is observed in the data. Let α_0 denote the probability that $y_i = 1$ when truly $\tilde{y}_i = 0$ and α_1 denote the probability that $y_i = 0$ when truly $\tilde{y}_i = 1$. The misclassification probabilities depend on the true value, $\tilde{y}_i$, so the extent of misclassification depends on how good a proxy y_i is of $\tilde{y}_i$. The misclassification probabilities are assumed to be independent of the covariates, x_i, conditional on the true response, more formally:[31]

$$\alpha_0 = \Pr(y_i = 1 | \tilde{y}_i = 0) = \Pr(y_i = 1 | \tilde{y}_i = 0, x_i), \tag{4.9}$$

$$\alpha_1 = \Pr(y_i = 0 | \tilde{y}_i = 1) = \Pr(y_i = 0 | \tilde{y}_i = 1, x_i). \tag{4.10}$$

The assumption that the misclassification probabilities are independent of the covariates is important for identification and this will be discussed later on.

The probability that an observation is classified as being equal to one $(\Pr(y_i = 1 | x_i))$ is given by the probability that it has been correctly classified as being equal to one $(1 - \alpha_1)$ multiplied by the probability that it is truly equal to one $(F(x_i^{'}\beta))$ plus the probability that it has been incorrectly classified as being equal to one (α_0) multiplied by the probability that it truly is not equal to one $(1 - F(x_i^{'}\beta))$ as follows:

$$\Pr(y_i = 1 | x_i) = (1 - \alpha_1) F(x_i^{'}\beta) + \alpha_0 (1 - F(x_i^{'}\beta)) = \alpha_0 + (1 - \alpha_0 - \alpha_1) F(x_i^{'}\beta) \tag{4.11}$$

Likewise, the probability that an observation is classified as being equal to zero is given by:

$$\Pr(y_i = 0 | x_i) = \alpha_1 F(x'_i\beta) + (1 - \alpha_0)(1 - F(x'_i\beta)) = 1 - \alpha_0 - (1 - \alpha_0 - \alpha_1) F(x'_i\beta) \tag{4.12}$$

Therefore we can write the expected value of the observed dependent variable, y_i, as:

---

[31] Hausman *et al.* (1998) show how the model can be extended to allow for covariate dependent misclassification and this is discussed in Chapter 5.

$$E(y_i|x_i) = \Pr(y_i = 1|x_i) = \alpha_0 + (1-\alpha_0-\alpha_1)F(x'_i\beta) \tag{4.13}$$

When there is no misclassification $(\alpha_0 = \alpha_1 = 0)$, this collapses to usual expression $F(x_i'\beta)$.

If we assume that ε_i are normally distributed then we can use equations (4.11) and (4.12) to derive the log-likelihood function for the probit model with misclassification:

$$\ln L = \sum_{i=1}^{n}\{y_i \ln(\Pr(y_i = 1|x_i)) + (1-y_i)\ln(\Pr(y_i = 0|x_i))\} =$$

$$n^{-1}\sum_{i=1}^{n}\{y_i \ln(\alpha_0 + (1-\alpha_0-\alpha_1)\Phi(x_i'\beta)) + (1-y_i)\ln(1-\alpha_0-(1-\alpha_0-\alpha_1)\Phi(x_i'\beta))\} \tag{4.14}$$

where $\Phi(.)$ denotes the cdf of the standard normal distribution

Maximising the log-likelihood function given in (4.14) with respect to α_0, α_1 and β yields consistent and efficient estimates of β as well as the probabilities of misclassification.

Identification

The conditions for identification of α_0, α_1 and β are similar to those for the traditional binary choice model. One additional assumption is needed for identification, namely that the misclassification probabilities are not very large, specifically, $\alpha_0 + \alpha_1 < 1$.[32] The assumption is needed because the normal distribution is symmetric ($\Phi(x_i'\beta)$=1- $\Phi(-x_i'\beta)$) and we can define $\hat{\alpha}_0 = 1-\alpha_1$, $\hat{\alpha}_1 = 1-\alpha_0$ and $\hat{\beta} = -\beta$ so that:

$$\begin{aligned}&\hat{\alpha}_0 + (1-\hat{\alpha}_0-\hat{\alpha}_1)\Phi(x_i'\hat{\beta}) = (1-\alpha_1) + ((1-(1-\alpha_1)-(1-\alpha_0))(1-\Phi(-x_i'\hat{\beta}))\\&= (1-\alpha_1) + ((-1+\alpha_0+\alpha_1)(1-\Phi(x_i'\beta)) = \alpha_0 + (1-\alpha_0-\alpha_1)\Phi(x_i'\beta)\end{aligned} \tag{4.15}$$

[32] Hausman *et al.* (1998) refer to this assumption as the 'monotonicity condition'.

When the assumption, $\alpha_0 + \alpha_1 < 1$, is not imposed the maximum likelihood estimator cannot distinguish between the parameter values $(\alpha_0,\ \alpha_1, \beta)$ and $(1-\alpha_0, 1-\alpha_1, -\beta)$. The assumption that $\alpha_0 + \alpha_1 < 1$ excludes this situation because $\alpha_0 + \alpha_1 < 1$ implies $(1-\alpha_1)+(1-\alpha_0) > 1$. An implication of this assumption is that if $\alpha_0 + \alpha_1 > 1$ but we impose $\alpha_0 + \alpha_1 < 1$ the estimates of β will have the wrong sign. This assumption guarantees that $\alpha_0 + (1-\alpha_0-\alpha_1)\Phi\left(x_i^{'}\beta\right)$ is strictly increasing in $x_i^{'}\beta$ as Φ (.) is strictly increasing.

The model parameters are identified from the nonlinearity of $\Phi(.)$. To see this, consider the linear probability model where $F(x_i^{'}\beta) = (x_i^{'}\beta)$, then the expected value of y_i is given by:

$$
\begin{aligned}
&E\left(y_i | x_i\right) = \Pr(y_i = 1 | x_i) = \alpha_0 + (1-\alpha_0-\alpha_1)\left(x_i^{'}\beta\right) \\
&= (\alpha_0 + \beta_0) + z_i^{'}((1-\alpha_0-\alpha_1)\beta_1)
\end{aligned}
\tag{4.16}
$$

where $x_i = (1, z_i^{'})^{'}$ and $\beta = (\beta_0, \beta^{'}{}_1)^{'}$ (i.e. separating out the constant)

In this case the parameters of the model cannot be separately identified.

Estimating α_0 and α_1 is only possible because they enter (4.14) additively and are then multiplied by the expression with the normal cdf. To identify the misclassification probabilities, α_0 and α_1, $x_i^{'}\beta$ has to get reasonably large in magnitude so as to push $\Pr\left(\tilde{y}_i = 1 | x_i\right)$ close to 0 and 1 for some i. The intuition behind this is that we have assumed that misclassification rates are constant and depend only on the true value, $\tilde{y}_i$, so α_0, the probability that we classify a case to be equal to one when truly it is equal to zero, is identified from the observations that have a near zero probability of truly having $\tilde{y}_i = 1$. These are observations where $x_i^{'}\beta$ is highly negative and which are very unlikely to have $\tilde{y}_i = 1$. So, if we observe some proportion of them as having $y_i = 1$, then these cases are probably misclassified. We can also see this by noting that in equation (4.13) as $x_i^{'}\beta$ becomes more negative $E(y_i|x_i)$ tends to α_0.

In a similar fashion, α_1, the probability of misclassifying a true one being equal to zero i.e. misclassifying $\tilde{y}_i = 1$ as $y_i = 0$, is estimated from the observations that have very large and positive $x_i^{'}\beta$ and so have a very high probability of truly having $\tilde{y}_i = 1$. If we observe some proportion of these observations as having $y_i = 0$, then these cases are probably misclassified. Similar to before, we can see from equation (4.13) that as $x_i^{'}\beta$ becomes more positive $E(y_i|x_i)$ tends to $1-\alpha_1$. The identification of the misclassification rates is discussed further in Section 4.3.2.

Marginal Effects

From equation (4.7) we know that the expected value on the true response, $\tilde{y}_i$, is given by:

$$E(\tilde{y}_i|x_i) = \Pr(\tilde{y}_i = 1|x_i) = F(x_i^{'}\beta) \tag{4.17}$$

In general, we are usually more interested in the marginal effect of a specific variable, k, which is given by:

$$\frac{\partial E(\tilde{y}_i|x_i)}{\partial x_{ik}} = \frac{\partial \Pr(\tilde{y}_i = 1|x_i)}{\partial x_{ik}} = f(x_i^{'}\beta)\beta_k \tag{4.18}$$

Equation (4.13) gives the expected value on the observed response, y_i, and the marginal effect of the variable, k, is given by:

$$\frac{\partial E(y_i|x_i)}{\partial x_{ik}} = \frac{\partial \Pr(y_i = 1|x_i)}{\partial x_{ik}} = (1-\alpha_0-\alpha_1)f(x_i^{'}\beta)\beta_k \tag{4.19}$$

Comparing equations (4.18) and (4.19) shows that when there is misclassification the marginal effect on the observed response (from equation (4.19)) will be biased towards zero (as $(1-\alpha_0-\alpha_1) < 1$). The marginal effect on the observed response will always be less than the marginal effect on the true response by a factor of $(1-\alpha_0-\alpha_1)$. This result only holds when misclassification is independent of the covariates.

To see the intuition behind this result, consider the following simple job change example: suppose you have a sample of 20 people, 10 of whom have a high value of some characteristic that makes them more likely to change jobs and the remaining 10 people have a low value of this characteristic that makes them less likely to change jobs. Further suppose that 8 people from the first group and 4 from the second group are identified as job changers. Then the true marginal effect on the characteristic is 0.4 (0.8-0.4). Now further suppose that we introduce misclassification (that does not depend on the particular characteristic) such that 4 out of the 8 true non job changes are misclassified (i.e. α_0=.5) and 3 out of the 12 true job changes are misclassified (i.e. α_1=.25). As the misclassification probabilities are assumed not to depend on the characteristic, this implies that 1 job stayer from the first group and 3 from the second group are misclassified as job changers and 2 job changers from the first group and 1 from the second group are misclassified as job stayers. Then the marginal effect on the characteristic is 0.1 (0.7-0.6) which is a quarter or $(1-\alpha_0-\alpha_1)$ of the true marginal effect.

4.2.3 Empirical Applications

The Hausman *et al.* (1998) estimator has been used in a wide range of empirical applications. Some studies focus on situations where there is misreporting in data, possibly because respondents in a survey have a psychological or economic incentive to misreport. Alternatively they may misunderstand the question or they may have poor recall. In addition, their responses may be coded incorrectly. For example, Artís *et al.* (2002) estimate models of fraud detection in the insurance industry, where there is uncertainty about whether claims have been correctly classified as honest or fraudulent. Brachet (2008) and Kenkel *et al*. (2004) investigate the misreporting of smoking participation. Caudill and Mixon (2005) are interested in the reliability of the self-reported incidence of undergraduate student cheating. Flathmann and Sheffrin (2003) use a taxpayer compliance survey to assess the reliability of self-reported non-compliance in completing tax returns. Dustman and Van Soest (2004) examine measurement error in the english-speaking fluency of immigrants in the UK (which is measured using an ordinal scale) and they extend the Hausman *et al.* estimator to the ordered probit case. Finally, Dustman and Van Soest (2001) also examine self-reported speaking fluency of migrants and they distinguish between time-varying and time-persistent misclassification errors.

In other studies, a dummy variable is created to serve as a proxy for some true underlying variable. For example, Leece (2000) examines the household choice of fixed versus variable rate mortgages where an algorithm is used to identify fixed rate debt in the data and so the type of debt may be misclassified. Dye and McMillen (2007) investigate redevelopment, where existing housing in established locations are torn down and replaced by new housing, in the Chicago metropolitan area. The authors use demolition permits as a proxy for redevelopment. This is an imperfect measure as a demolition permit may be issued but the house may not be demolished or it could be demolished but the land could be converted to non-residential use. Jensen *et al.* (2008) examine misclassification in patent applications. Patents are awarded on the basis of the size of the inventive step of the innovation and so it is possible that an application with a very high inventive step is incorrectly classified by a patent office as a 'refusal'.

4.3 Identification of the Model

This section uses Monte Carlo simulation techniques to examine the identification of the Hausman *et al*. estimator. The aim of the section is to complement the discussion of the identification of the estimator in Section 4.2.2.

4.3.1 Monte Carlo Simulations

In the first set of Monte Carlo simulations, the latent variable, y_i^*, depends on a single continuous covariate, x_{1i}, as follows:

Model 1: $$y_i^* = -1 - 1.5x_{1i} + \varepsilon_i$$

Where x_{1i} is drawn from a normal distribution with mean zero and standard deviation two and the error term, ε_i, is drawn from a standard normal distribution. The true response, $\tilde{y}_i$, is given by:

$$\tilde{y}_i = 1 \text{ if } y_i^* >= 0$$
$$= 0 \text{ otherwise.}$$

The observed dependent variable is generated using symmetric misclassification i.e. $\alpha_0 = \alpha_1$ and four rates of misclassification, namely 1%, 5%, 10% and 20%, are considered.[33] To generate misclassification the observed dependent variable, y_i, is drawn from a uniform distribution and whenever the value exceeds the cut-off point the observations are recoded to be zeros if they were ones and ones if they were zeros and when the value is less than the cut-off point the observations are equal to the true response, $\tilde{y}_i$.

The average estimates of the misclassification probabilities from a thousand Monte Carlo simulations are reported in Table 4.1. For each level of misclassification considered, the Hausman *et al.* estimates of the misclassification probabilities are very close to the true rates. For each set of simulations, the actual rates of misclassification that have been generated in the data are reported in Table 4.1. On average, the generated rates of misclassification are very close to the desired ones, although these rates will vary somewhat between simulations. The average standard errors increase as the level of misclassification increases although on average the estimates are highly significant. The average standard error of α_1 exceeds that of α_0 and this will be discussed in Section 4.4.3.

[33] Section 4.4.4 considers examples where misclassification is asymmetric (i.e. where $\alpha_0 \neq \alpha_1$).

Table 4.1: Estimates of Misclassification Probabilities for Model 1*

Level of Misclassification	Actual Misclassification Rate	$\hat{\alpha}_0$	Std Error $\hat{\alpha}_0$	MSE $\hat{\alpha}_0$	Min $\hat{\alpha}_0$	Max $\hat{\alpha}_0$	Actual Misclassification Rate	$\hat{\alpha}_1$	Std Error $\hat{\alpha}_1$	MSE $\hat{\alpha}_1$	Min $\hat{\alpha}_1$	Max $\hat{\alpha}_1$	No. of Sims
							Model 1: $y_i^* = -1 - 1.5x_{1i} + \varepsilon_i$						
1%	0.0100	0.0092	0.0018	0.0000	0.0043	0.0148	0.0100	0.0103	0.0029	0.0000	0.0025	0.0204	1,000
5%	0.0500	0.0487	0.0038	0.0000	0.0367	0.0624	0.0501	0.0497	0.0058	0.0000	0.0327	0.0657	1,000
10%	0.1000	0.0983	0.0052	0.0000	0.0832	0.1172	0.1000	0.0988	0.0077	0.0000	0.0760	0.1235	1,000
20%	0.2001	0.1985	0.0068	0.0000	0.1780	0.2162	0.2001	0.1984	0.0102	0.0001	0.1686	0.2306	1,000

*The sample size is 10,000. The results are based on the average values from the Monte Carlo simulations. Specifically, the results in columns 3-7 and 9-13 are based on the average values from the Monte Carlo simulations and columns 2 and 8 are based on the average generated rates of misclassification from the simulations.

4.3.2 Identification of the Model

In a standard probit model without misclassification, we would expect there to be no significant difference between the observed proportion of $y_i = 1$ and the predicted $\Pr(y_i = 1)$ at different values of the explanatory variables, if the model fits the data well. However, we would expect there to be a difference between the observed proportion of $y_i = 1$ and the predicted $\Pr(y_i = 1)$ when some of the observations are misclassified because these misclassified observations cannot be explained by the covariates in the model.

To illustrate this point, Table 4.2 compares the average predicted $\Pr(y_i = 1)$ with the average observed $y_i = 1$ from 1,000 Monte Carlo simulations of Model 1 from a probit model where 10 per cent of the data is misclassified (i.e. $\alpha_0 = 10\%$ and $\alpha_1 = 10\%$) and another probit model when none of the data is misclassified. In each simulation, the data are sorted from the observation with the largest negative value of $x_{1i}^{'}\beta$ to the observation with the largest positive value of $x_{1i}^{'}\beta$. The table shows that when the data is misclassified there is quite a difference between the observed proportion of $y_i = 1$ and the predicted $\Pr(y_i = 1)$, while the two are practically the same when there is no misclassification in the data. The table also shows that when 10 per cent of the data is misclassified, the groups with the most negative characteristics have around 10 per cent of observations with $y_i = 1$ and a very low or zero average predicted $\Pr(y_i = 1)$. This arises because these misclassified observations cannot be explained by the covariates in the model. Similarly, only 90 per cent of the groups with the most positive characteristics have $y_i = 1$ and they essentially have an average predicted $\Pr(y_i = 1)$ of 1.

Table 4.2: Comparison of Probit Models with and without Misclassification*

	Probit Model with 10% Misclassification			***Probit Model with No Misclassification***		
Groups of $x_{1i}^{'}\beta$	Proportion of Observed $y_i = 1$	Predicted $\Pr(y_i = 1)$	Absolute Difference	Proportion of Observed $y_i = 1$	Predicted $\Pr(y_i = 1)$	Absolute Difference
Minimum $x_{1i}^{'}\beta$	0.1004	0.0004	0.1000	0.0000	0.0000	0.0000
2-100	0.1013	0.0112	0.0901	0.0000	0.0000	0.0000
101-500	0.0990	0.0569	0.0422	0.0000	0.0000	0.0000
501-1000	0.1002	0.1299	0.0297	0.0000	0.0007	0.0007
1001-2000	0.1145	0.2389	0.1243	0.0190	0.0239	0.0049
2001-3000	0.2356	0.3644	0.1288	0.1700	0.1592	0.0108
3001-4000	0.4459	0.4773	0.0313	0.4320	0.4313	0.0007
4001-5000	0.6699	0.5784	0.0915	0.7120	0.7121	0.0001
5001-6000	0.8150	0.6756	0.1394	0.8930	0.9029	0.0099
6001-7000	0.8863	0.7657	0.1206	0.9830	0.9806	0.0024
7001-8000	0.8990	0.8467	0.0523	0.9990	0.9982	0.0008
8001-9000	0.8998	0.9182	0.0184	1.0000	0.9999	0.0001
9001-9500	0.8994	0.9647	0.0653	1.0000	1.0000	0.0000
9501-9900	0.8997	0.9878	0.0881	1.0000	1.0000	0.0000
9901-9990	0.8984	0.9979	0.0995	1.0000	1.0000	0.0000
9991-9999	0.9000	0.9999	0.0999	1.0000	1.0000	0.0000
Maximum $x_{1i}^{'}\beta$	0.9082	1.0000	0.0918	1.0000	1.0000	0.0000

* In both models the sample size is 10,000 and the results are based on the average values from 1,000 Monte Carlo simulations

Turning to the Hausman *et al.* estimator, the identification of α_0, or the proportion of observations that are misclassified as $y_i = 1$ when $\tilde{y}_i = 0$, comes from the observations with the most negative characteristics. If we take a group of observations with the most negative characteristics, we would expect them to have $\tilde{y}_i = 0$ but if some of them are observed as having $y_i = 1$ then these cases are probably misclassified. For this group, we would expect the observed proportion of $y_i = 1$ to be higher than the predicted $\Pr(y_i = 1)$ because the observations with $y_i = 1$ will have a very low or zero $\Pr(y_i = 1)$ as they have very low values of $x_{1i}^{'}\beta$. Table 4.3 compares the predicted $\Pr(y_i = 1)$ with the observed $y_i = 1$ by groups of the $x_{1i}^{'}\beta$ distribution for Model 1 from the Hausman *et al.* model when there is 1 per cent misclassification in the data. As before, in each simulation the data are sorted from the observation with the largest negative value of $x_{1i}^{'}\beta$ to the observation with the largest positive value of $x_{1i}^{'}\beta$.

Table 4.3: Comparison of Predicted Probabilities with Observed Proportion of 1's by Groups of $x_{1i}^{'}\beta$ for Model 1 with 1% Symmetric Misclassification, using the Hausman *et al.* Estimator*

Groups of $x_{1i}^{'}\beta$	Proportion of True $\tilde{y}_i = 1$	Proportion of Observed $y_i = 1$	Predicted $\Pr(y_i = 1)$	Proportion of Observed $y_i = 1$ - Predicted $\Pr(y_i = 1)$	Predicted $\Pr(y_i = 1)$- Proportion of Observed $y_i = 1$
Minimum $x_{1i}^{'}\beta$	0.0000	0.0040	0.0000	0.0040	-0.0040
2-100	0.0000	0.0101	0.0000	0.0101	-0.0101
101-500	0.0000	0.0100	0.0000	0.0100	-0.0100
501-1000	0.0000	0.0102	0.0000	0.0102	-0.0102
1001-2000	0.0000	0.0100	0.0001	0.0099	-0.0099
2001-3000	0.0010	0.0110	0.0027	0.0083	-0.0083
3001-4000	0.0200	0.0296	0.0231	0.0065	-0.0065
4001-5000	0.0910	0.0992	0.0977	0.0015	-0.0015
5001-6000	0.3060	0.3097	0.2815	0.0282	-0.0282
6001-7000	0.5530	0.5520	0.5664	-0.0144	0.0144
7001-8000	0.8320	0.8252	0.8377	-0.0125	0.0125
8001-9000	0.9780	0.9685	0.9751	-0.0066	0.0066
9001-9500	0.9960	0.9860	0.9991	-0.0131	0.0131
9501-9900	1.0000	0.9897	1.0000	-0.0103	0.0103
9901-9990	1.0000	0.9921	1.0000	-0.0079	0.0079
9991-9999	1.0000	0.9921	1.0000	-0.0079	0.0079
Maximum $x_{1i}^{'}\beta$	1.0000	0.9921	1.0000	-0.0079	0.0079

* The sample size is 10,000 and the results are based on the average values from 1,000 Monte Carlo simulations

The table shows that around 1 per cent of the observations in the groups with the lowest values of $x_{1i}^{'}\beta$ have $y_i = 1$. However, the associated predicted $\Pr(y_i = 1)$ for these observations is essentially zero because based on their values of $x_{1i}^{'}\beta$ we would expect these cases to have $y_i = 0$. In fact, the table also shows the true proportion of $\tilde{y}_i = 1$ and we can see that it is zero for these groups. The difference between the proportion of $y_i = 1$ and the predicted $\Pr(y_i = 1)$ for the groups with the most negative values of x_{1i} is approximately 1 per cent and this type of measure can be used to identify α_0. As the estimator uses observations in the bottom part of the $x_{1i}^{'}\beta$ distribution to identify α_0, it will only be able to accurately identify it if there are observations with very low values of $x_{1i}^{'}\beta$ where $y_i = 1$.

In an analogous way, we can see how α_1, or the proportion of cases that are misclassified as having $y_i = 0$ when truly $\tilde{y}_i = 1$, is identified. We would expect observations with very high values of $x_{1i}^{'}\beta$ to have $\tilde{y}_i = 1$ so observations with very high values of $x_{1i}^{'}\beta$ that have $y_i = 0$ are likely to be misclassified. In this case, we would expect the predicted $\Pr(y_i = 1)$ to exceed the proportion of $y_i = 1$ for groups with the most positive characteristics. The bottom part of Table 4.3 shows that groups with the most positive values of $x_{1i}^{'}\beta$ have a predicted $\Pr(y_i = 1)$ of essentially 1 yet only around 99 per cent of each of the groups is observed with $y_i = 1$. The difference between the predicted $\Pr(y_i = 1)$ and the proportion of $y_i = 1$ for these groups yields an estimate of α_1. Therefore to identify α_1 there needs to be observations with $y_i = 0$ when $x_{1i}^{'}\beta$ takes on very high values. Appendix Tables 4.1 to 4.3 show a similar comparison of predicted probabilities and observed proportions of $y_i = 1$ for Model 1 when there is 5 per cent, 10 per cent and 20 per cent misclassification respectively, in the data. From these tables, we can see that the misclassification probabilities are identified from the groups of observations with the most negative and most positive values of the $x_{1i}^{'}\beta$ distribution.

The range of $\Pr(\widetilde{y}_i = 1)$ in each of these models is from 0 to 1 meaning that $x_{1i}^{'}\beta$ gets sufficiently large in magnitude in each model to generate such probabilities. This will become more relevant in Section 4.5 where an example is given of when the estimator performs poorly.

4.4 Extensions to the Basic Model Specification

This section considers a range of changes and extensions to the basic model specification outlined in Section 4.3. The aim of the section is to ascertain how the Hausman *et al.* estimator performs in different situations and consequently to determine conditions under which the estimator is appropriate to use to control for misclassification.

4.4.1 Weakening the Impact of the Explanatory Variable in the Model

The next set of Monte Carlo simulations consider two models, identical to Model 1 in all respects except that the effect of the explanatory variable becomes progressively weaker, as follows:[34,35]

Model 2: $y_i^* = -1 - 1.0x_{1i} + \varepsilon_i$

Model 3: $y_i^* = -1 - 0.5x_{1i} + \varepsilon_i$

The results of the Monte Carlo simulations are shown in Table 4.4. The results show that the estimated misclassification probabilities remain close to their true values even when the effect of the explanatory variable becomes weaker. However, there is a loss in estimation efficiency for each given level of misclassification. This is also reflected in the increase in the range of estimates for α_0 and α_1.

Table 4.4 shows that some of the Monte Carlo simulations have generated negative estimates of the misclassification probabilities, particularly for Model 3. For example, in the simulations run on Model 3 when 1 per cent of the data is misclassified there are 10 cases where the estimate of α_0 is negative. In all cases, misclassification is assumed to be independent of the covariates so that each observation with $y_i = 1$ or $y_i = 0$ has an equal chance of being misclassified in any given simulation. A negative α_0 could be generated if none or very few of the observations in the left tail of the $x_{1i}^{'}\beta$ distribution are misclassified as $y_i = 1$. In this case, the proportion of cases with $y_i = 1$ in the group with the most negative values of $x_{1i}^{'}\beta$ would be very low or it could be zero and this value could be below the predicted $\Pr(y_i = 1)$ for the group. This could yield a negative value for α_0. In the case of Model 3 it appears that because the effect of the explanatory variable is considerably weaker and the

[34] In one respect, these models are similar to introducing measurement error in the explanatory variable as that leads to downwards bias in the coefficient estimate.

[35] Reducing the effect of the explanatory variable in the true underlying model also serves to reduce the proportion of $\widetilde{y}_i = 1$ in the data. For example, the proportion of $\widetilde{y}_i = 1$ in Model 1 is 38 per cent and this proportion falls to 33 per cent and 25 per cent in Models 2 and 3 respectively. Appendix Table 4.4 shows the proportion of $\widetilde{y}_i = 1$ in each of the models considered in the chapter.

proportion of $\tilde{y}_i = 1$ is smaller than in Models 1 and 2, that in some simulations, there may not be sufficient support at low values of the index $x_{1i}^{'}\beta$ to accurately identify α_0.[36]

To illustrate this point, I chose one of the simulations generating a negative estimate of α_0 and looked closely at the dataset. Table 4.5 reports results for a simulation where the estimate of α_0 is negative in Model 3 when there is 1 per cent symmetric misclassification. In this case, it appears that essentially none of the observations with very negative values of $x_{1i}^{'}\beta$ have been misclassified as $y_i = 1$ because the observed proportion of $y_i = 1$ is equal to the true proportion $\tilde{y}_i = 1$ for the groups with the lowest values of $x_{1i}^{'}\beta$. For some of the groups with very negative values of $x_{1i}^{'}\beta$ the proportion of observed $y_i = 1$ is smaller than the predicted $\Pr(y_i = 1)$ (highlighted in bold in the table) and this produces a negative estimate of α_0. A similar argument about none or very few cases being misclassified in the right tail can explain how a negative estimate of α_1 could be generated. This highlights how the Hausman *et al.* estimator only uses the observations with strong characteristics to identify the misclassification probabilities. In the simulation reported in Table 4.5, even though 1 per cent of the data is misclassified, none of the data in the left tail of the $x_{1i}^{'}\beta$ distribution has been misclassified and so α_0 is not estimated accurately. This result shows that even when the misclassification probabilities are independent of the covariates, negative estimates of α_0 are possible.

In a real world dataset, a negative misclassification probability could indicate that misclassification is not independent of the covariates. For example, if misclassification depends on x_{1i} in such a way that only observations with intermediate values of x_{1i} are misclassified, then the estimator would not be able to identify the misclassification rates from the groups with the extreme values of $x_{1i}^{'}\beta$.

[36] This idea is explored further in Section 4.4.2 where the sample size in Model 3 is increased to 100,000 in an attempt to ascertain whether increasing the number of cases that are misclassified in absolute terms helps to improve the estimates from the Monte Carlo simulations.

Table 4.4: Estimates of Misclassification Probabilities for Models 1-3*

Level of Misclassification	Actual Misclassification Rate	$\hat{\alpha}_0$	Std Error $\hat{\alpha}_0$	MSE $\hat{\alpha}_0$	Min $\hat{\alpha}_0$	Max $\hat{\alpha}_0$	Actual Misclassification Rate	$\hat{\alpha}_1$	Std Error $\hat{\alpha}_1$	MSE $\hat{\alpha}_1$	Min $\hat{\alpha}_1$	Max $\hat{\alpha}_1$	No. of Sims
Model 1: $y_i^* = -1 - 1.5x_{1i} + \varepsilon_i$ *(Repeated from Table 4.1)*													
1%	0.0100	0.0092	0.0018	0.0000	0.0043	0.0148	0.0100	0.0103	0.0029	0.0000	0.0025	0.0204	1,000
5%	0.0500	0.0487	0.0038	0.0000	0.0367	0.0624	0.0501	0.0497	0.0058	0.0000	0.0327	0.0657	1,000
10%	0.1000	0.0983	0.0052	0.0000	0.0832	0.1172	0.1000	0.0988	0.0077	0.0000	0.0760	0.1235	1,000
20%	0.2001	0.1985	0.0068	0.0000	0.1780	0.2162	0.2001	0.1984	0.0102	0.0001	0.1686	0.2306	1,000
Model 2: $y_i^* = -1 - 1.0x_{1i} + \varepsilon_i$ *(Note: Weaker Effect of* x_{1i}*)*													
1%	0.0100	0.0095	0.0021	0.0000	0.0036	0.0154	0.0100	0.0099	0.0046	0.0000	-0.0023	0.0241	999
5%	0.0500	0.0491	0.0043	0.0000	0.0346	0.0630	0.0501	0.0491	0.0085	0.0000	0.0202	0.0725	1,000
10%	0.1000	0.0989	0.0058	0.0000	0.0825	0.1161	0.1000	0.0980	0.0112	0.0001	0.0674	0.1315	1,000
20%	0.2001	0.1990	0.0076	0.0000	0.1790	0.2219	0.2002	0.1971	0.0145	0.0002	0.1358	0.2387	1,000
Model 3: $y_i^* = -1 - 0.5x_{1i} + \varepsilon_i$ *(Note: Weaker Effect of* x_{1i}*)*													
1%	0.0100	0.0075	0.0035	0.0000	-0.0022	0.0180	0.0100	0.0198	0.0303	0.0003	-0.0233	0.0672	1,000
5%	0.0500	0.0462	0.0068	0.0000	0.0267	0.0641	0.0502	0.0546	0.0371	0.0008	-0.0287	0.1428	1,000
*10%***	*0.1000*	*0.9041*	*0.0089*	0.6467	*0.8757*	*0.9314*	*0.1000*	*0.8997*	*0.0435*	0.6409	*0.8087*	*1.0691*	*1,000*
10%	0.1000	0.0959	0.0089	0.0001	0.0686	0.1243	0.1000	0.1003	0.0435	0.0013	-0.0691	0.1913	1,000
20%	0.2000	0.1959	0.0117	0.0001	0.1593	0.2317	0.2004	0.1940	0.0534	0.0027	-0.0926	0.3119	1,000

* The sample size is 10,000 and the results are based on the average values from the Monte Carlo simulations

** In this case each simulation estimates 1-α_0, 1-α_1 and $-\beta$. This arises because the condition $\alpha_0 + \alpha_1 < 1$ which is needed for identification has not been imposed in the Monte Carlo simulations. When the true values of β from Model 3 are given as starting values the simulation results (given in the following line of the table) are sensible

Table 4.5: Comparison of Predicted Probabilities with Observed Proportion of 1's by Groups of the x_{1i} Distribution for Model 3 with 1% Symmetric Misclassification when $\hat{\alpha}_0$ is -0.0014

Groups of $x_{1i}^{'}\beta$	Proportion of True $\tilde{y}_i = 1$	Proportion of Observed $y_i = 1$	Predicted $\Pr(y_i = 1)$	Proportion of Observed $y_i = 1$ - Predicted $\Pr(y_i = 1)$	Predicted $\Pr(y_i = 1)$- Proportion of Observed $y_i = 1$
Minimum $x_{1i}^{'}\beta$	0.0000	0.0000	0.0000	0.0000	0.0000
2-100	0.0000	0.0000	0.0005	**-0.0005**	0.0005
101-500	0.0025	0.0025	0.0040	**-0.0015**	0.0015
501-1000	0.0020	0.0120	0.0123	-0.0003	0.0003
1001-2000	0.0190	0.0310	0.0310	0.0000	0.0000
2001-3000	0.0550	0.0680	0.0619	0.0061	-0.0061
3001-4000	0.0790	0.0880	0.1009	-0.0129	0.0129
4001-5000	0.1370	0.1420	0.1479	-0.0059	0.0059
5001-6000	0.2260	0.2310	0.2086	0.0224	-0.0224
6001-7000	0.2720	0.2780	0.2857	-0.0077	0.0077
7001-8000	0.3910	0.3860	0.3843	0.0017	-0.0017
8001-9000	0.5320	0.5280	0.5192	0.0088	-0.0088
9001-9500	0.6680	0.6640	0.6640	0.0000	0.0000
9501-9900	0.8325	0.8300	0.7986	0.0314	-0.0314
9901-9990	0.9444	0.9444	0.9202	0.0243	-0.0243
9991-9999	1.0000	1.0000	0.9861	0.0139	-0.0139
Maximum $x_{1i}^{'}\beta$	0.0000	0.0000	0.0000	0.0000	0.0000

4.4.2 Effect of Sample Size

It is important to ascertain how sensitive the results are to changes in sample size. The simulation results presented so far have been based on datasets where the sample size is 10,000. Two additional models are considered. The first uses the same equation as Model 1 (which performed well) but where the sample size is reduced to 1,000 (Model 4) and the second uses the same equation as Model 3 but the sample size is increased to 100,000 (Model 5). The Monte Carlo simulation results are presented in Table 4.6.

The results for Model 4 when there is 5%, 10% and 20% symmetric misclassification show that the average estimates of the misclassification probabilities are quite close to their true values, which is encouraging because of the small sample size. However there is a marked increase in the average standard errors associated with the estimates compared to those in Model 1.

The results for this model when there is 1% symmetric misclassification are disappointing. First, some of the estimates of the misclassification probabilities are excessively large in magnitude and this will be discussed in Section 4.5.1. The results indicate that some of the estimates, particularly for α_0 are being driven by extreme values so it may be more appropriate to consider the median estimates. The median estimate of α_0 is 0.0087 and the median estimate of α_1 is 0.0120, both of which are close to the true misclassification rates. Second, the estimator only converges in 542 out of 1,000 simulations. In this model, the sample size is only 1,000 and the proportion of $\tilde{y}_i = 1$ is .377 (see Appendix Table 4.4). This means that $\tilde{y}_i = 1$ for 377 cases and $\tilde{y}_i = 0$ for the remaining 623 cases. If 1 per cent of each group is misclassified, then on average only around 4 cases where $\tilde{y}_i = 1$ and around 6 cases where $\tilde{y}_i = 0$ will be misclassified. In this case, because the number of misclassified cases in any given simulation is so small in absolute terms the estimator has difficulty in identifying the model in some simulations.

Table 4.6 also reports the results for this model when there is 1 per cent symmetric misclassification when the true parameters from the hypothetical dataset are used as starting values for β_0 and β_1. The simulation results are much better when these starting values are given. The results show that on average the estimates of α_0 and α_1 are close to the true misclassification rates and although some of the estimates are negative, there are no extreme estimates. However, even giving the true population parameters as starting values in each simulation, the estimator only converges in 677 out of 1,000 simulations.

The third panel in Table 4.6 shows the simulation results where the sample size is increased to 100,000 (Model 5). This model uses the same equation as Model 3. The reason for using this equation is that when Model 3 was run with a sample of 10,000 negative probabilities were estimated in some instances. One of the arguments put forward as to why this may happen is that when the effect of the explanatory variable is weaker there may not be a sufficient number of cases in the tails of the distribution to accurately identify the misclassification probabilities. The results should be compared to those of Model 3 in Table 4.4. The results show that the range of estimates for both misclassification probabilities is much tighter than before and no estimates are negative. This indicates that increasing the number of misclassified cases in absolute terms helps with the identification of the estimates. The average estimate of α_1 for each model is further from the true rates of misclassification than is the case with Model 3. This is somewhat surprising but the negative estimates in Model 3 help to pull down the average estimates of the misclassification probabilities. The tables also report the mean square errors (MSE) associated with the estimates of the misclassification probabilities. Although the average estimates of α_1 from Model 5, where the sample size is 100,000, are further from the true values than for Model 3, where the sample size is 10,000, the MSEs of the estimates are lower, indicating the estimates from Model 5 are superior to those from Model 3.

Table 4.6: Estimates of Misclassification Probabilities for Models 4-6*

Level of Misclassification	Actual Misclassification Rate	$\hat{\alpha}_0$	Std Error $\hat{\alpha}_0$	MSE $\hat{\alpha}_0$	Min $\hat{\alpha}_0$	Max $\hat{\alpha}_0$	Actual Misclassification Rate	$\hat{\alpha}_1$	Std Error $\hat{\alpha}_1$	MSE $\hat{\alpha}_1$	Min $\hat{\alpha}_1$	Max $\hat{\alpha}_1$	No. of Sims
Model 1: $y_i^* = -1 - 1.5x_{1i} + \varepsilon_i$ *(Repeated from Table 4.1, note: sample size is 10,000)*													
1%	0.0100	0.0092	0.0018	0.0000	0.0043	0.0148	0.0100	0.0103	0.0029	0.0000	0.0025	0.0204	1,000
5%	0.0500	0.0487	0.0038	0.0000	0.0367	0.0624	0.0501	0.0497	0.0058	0.0000	0.0327	0.0657	1,000
10%	0.1000	0.0983	0.0052	0.0000	0.0832	0.1172	0.1000	0.0988	0.0077	0.0000	0.0760	0.1235	1,000
20%	0.2001	0.1985	0.0068	0.0000	0.1780	0.2162	0.2001	0.1984	0.0102	0.0001	0.1686	0.2306	1,000
Model 4: Effect of Reducing Sample Size, $y_i^* = -1 - 1.5x_{1i} + \varepsilon_i$ *(Note: Sample size is 1,000)*													
1%	0.0104	-3002.15	93.5732	926000000	-535392.6	0.0291	0.0116	0.0025	0.0131	0.0060	-0.7475	0.0386	542
Model 4: 1% Misclassification with the true values of β *given as starting values*													
1%	*0.0102*	*0.0092*	*0.0054*	0.0000	*-0.0103*	*0.0291*	*0.0114*	*0.0108*	*0.0112*	0.0001	*-0.0609*	*0.0386*	*677*
5%	0.0500	0.0473	0.0117	0.0001	0.0139	0.0823	0.0505	0.0476	0.0182	0.0002	-0.0081	0.1084	992
10%	0.1000	0.0965	0.0163	0.0001	0.0452	0.1474	0.1001	0.0951	0.0249	0.0004	0.0151	0.1664	998
20%	0.2000	0.1963	0.0220	0.0002	0.1224	0.2575	0.2000	0.1935	0.0335	0.0008	0.0696	0.2936	999
Model 5: Effect of Increasing Sample Size, $y_i^* = -1 - 0.5x_{1i} + \varepsilon_i$ *(Note: Sample size is 100,000 and the equation is the same as in Model 3)*													
1%	0.0100	0.0087	0.0011	0.0000	0.0057	0.0117	0.0100	0.0251	0.0097	0.0002	0.0095	0.0434	1,000
5%	0.0500	0.0484	0.0021	0.0000	0.0423	0.0558	0.0500	0.0633	0.0118	0.0002	0.0332	0.0878	1,000
*10%***	*0.0999*	*0.0984*	*0.0027*	0.0000	*0.0913*	*0.1065*	*0.1000*	*0.1118*	*0.0136*	0.0003	*0.0740*	*0.1444*	*1,000*
20%	0.2000	0.1988	0.0035	0.0000	0.1870	0.2079	0.2000	0.2093	0.0158	0.0003	0.1467	0.2522	1,000
Model 6: Increasing the Proportion of $\tilde{y}_i = 1$ **in the Sample,** $y_i^* = 1 - 1.5x_{1i} + \varepsilon_i$ *(Note: Change in Intercept)*													
1%	0.0100	0.0092	0.0027	0.0000	0.0023	0.0175	0.0100	0.0096	0.0018	0.0000	0.0045	0.0149	1,000
5%	0.0499	0.0490	0.0058	0.0000	0.0302	0.0663	0.0501	0.0493	0.0038	0.0000	0.0384	0.0603	1,000
10%	0.0998	0.0987	0.0078	0.0000	0.0720	0.1212	0.1001	0.0992	0.0051	0.0000	0.0837	0.1145	1,000
20%	0.1998	0.1987	0.0102	0.0001	0.1641	0.2311	0.2003	0.1994	0.0068	0.0000	0.1757	0.2189	999

* The sample size is 10,000 unless otherwise stated and the results are based on the average values from the Monte Carlo simulations

** In this set of Monte Carlo simulations, the true values of β are used as starting values, otherwise each simulation produces estimates of 1-α_0, 1-α_1 and $-\beta$

4.4.3 Increasing the Proportion of $\tilde{y}_i = 1$ in the Sample

The Monte Carlo simulation results for each model and each level of misclassification reported in the previous tables show that the average standard errors of the estimates for α_1 exceed those of α_0. For example, in Model 1 the average standard errors of α_1 are around 1.5 times the magnitude of those of α_0. This arises because there are relatively fewer observations with $\tilde{y}_i = 1$ than $\tilde{y}_i = 0$. In Model 1 around 38 per cent of observations have $\tilde{y}_i = 1$. Model 6 changes the intercept in Model 1 from –1 to +1 and this has the effect of increasing the proportion of observations with $\tilde{y}_i = 1$ to 62 per cent.[37] The Monte Carlo simulation results (reported in bottom panel of Table 4.6) show that the average standard errors of α_0 are now around 1.5 times those of the estimates of α_1.

4.4.4 Asymmetric Misclassification

This section investigates how the estimator performs when misclassification is asymmetric (i.e. $\alpha_0 \neq \alpha_1$). Table 4.7 shows the simulation results for Model 1 for three different examples of asymmetric misclassification. In Model 7 α_0=5% and α_1=20%, in Model 8 α_0=25% and α_1=1% and in Model 9 α_0=5% and α_1=50%. In each case, the estimator performs well with the average estimates of the misclassification probabilities being close to their true values.

However in Model 8, where α_0=25% and α_1=1%, the estimator only converges in 917 out of 1000 simulations. A further set of simulations were run on Model 8 with α_0=25% but α_1 was increased from 1% to 5%. In these simulations (not reported in the table) the model converges in 999 out of 1,000 simulations.

[37] Appendix Table 4.4 shows the proportion of $\tilde{y}_i = 1$ in each of the models considered in the chapter.

Table 4.7: Asymmetric Misclassification in Model 1*

Level of Misclassification	Actual Misclassification Rate	$\hat{\alpha}_0$	Std Error $\hat{\alpha}_0$	MSE $\hat{\alpha}_0$	Min $\hat{\alpha}_0$	Max $\hat{\alpha}_0$	Actual Misclassification Rate	$\hat{\alpha}_1$	Std Error $\hat{\alpha}_1$	MSE $\hat{\alpha}_1$	Min $\hat{\alpha}_1$	Max $\hat{\alpha}_1$	No. of Sims
Model 7: $y_i^* = -1 - 1.5x_{1i} + \varepsilon_i$ *(Note: α_0 =5% and α_1 =20%)*													
α_0=5% and α_1=20%	0.0500	0.0487	0.0038	0.0000	0.0364	0.0628	0.2001	0.1990	0.0101	0.0001	0.1691	0.2321	1000
Model 8: $y_i^* = -1 - 1.5x_{1i} + \varepsilon_i$ *(Note: α_0 =25% and α_1 =1%)*													
α_0=25% and α_1=1%	0.2510	0.2497	0.0073	0.0000	0.2301	0.2684	0.0100	0.0098	0.0028	0.0000	0.0024	0.0202	917
Model 9: $y_i^* = -1 - 1.5x_{1i} + \varepsilon_i$ *(Note: α_0 =5% and α_1 =50%)*													
α_0=5% and α_1=50%	0.0500	0.0491	0.0038	0.0000	0.0360	0.0633	0.5003	0.4994	0.0123	0.0001	0.4592	0.5379	1000

* The sample size is 10,000 and the results are based on the average values from the Monte Carlo simulations

Table 4.8 compares the predicted $\Pr(y_i = 1)$ with the observed $y_i = 1$ by groups of the $x_{1i}^{'}\beta$ distribution for Model 9 where α_0 is equal to 5 percent and α_1 is equal to 50 per cent. As before, in each simulation the data are sorted from the observation with the largest negative value of $x_{1i}^{'}\beta$ to the observation with the largest positive value of $x_{1i}^{'}\beta$. The table shows that around 5 per cent of the observations in the groups with the lowest values of $x_{1i}^{'}\beta$ have $y_i = 1$ and the associated predicted $\Pr(y_i = 1)$ for these observations are essentially zero. The difference between these two values provides the estimate of α_0. The bottom part of the table shows that groups with the most positive values of $x_{1i}^{'}\beta$ have a predicted $\Pr(y_i = 1)$ of essentially 1 yet only around 50 per cent of each of the groups is observed with $y_i = 1$. The difference between the predicted $\Pr(y_i = 1)$ and the proportion of $y_i = 1$ for these groups is how the estimate of α_1 is identified.

Table 4.8: Comparison of Predicted Probabilities with Observed Proportion of 1's by Groups of the $x_{1i}^{'}\beta$ Distribution for Model 9 where α_0=5% and α_1=50%

Groups of $x_{1i}^{'}\beta$	Proportion of True $\tilde{y}_i = 1$	Proportion of Observed $y_i = 1$	Predicted $\Pr(y_i = 1)$	Proportion of Observed $y_i = 1$ - Predicted $\Pr(y_i = 1)$	Predicted $\Pr(y_i = 1)$- Proportion of Observed $y_i = 1$
Minimum $x_{1i}^{'}\beta$	0.0000	0.0466	0.0000	0.0466	-0.0466
2-100	0.0000	0.0511	0.0000	0.0511	-0.0511
101-500	0.0000	0.0498	0.0000	0.0498	-0.0498
501-1000	0.0000	0.0503	0.0000	0.0503	-0.0503
1001-2000	0.0000	0.0502	0.0002	0.0500	-0.0500
2001-3000	0.0010	0.0506	0.0035	0.0471	-0.0471
3001-4000	0.0200	0.0591	0.0258	0.0333	-0.0333
4001-5000	0.0910	0.0909	0.1022	-0.0113	0.0113
5001-6000	0.3060	0.1875	0.2850	-0.0974	0.0974
6001-7000	0.5530	0.2982	0.5648	-0.2666	0.2666
7001-8000	0.8320	0.4237	0.8319	-0.4082	0.4082
8001-9000	0.9780	0.4899	0.9716	-0.4817	0.4817
9001-9500	0.9960	0.4984	0.9986	-0.5002	0.5002
9501-9900	1.0000	0.4994	1.0000	-0.5006	0.5006
9901-9990	1.0000	0.4939	1.0000	-0.5061	0.5061
9991-9999	1.0000	0.4909	1.0000	-0.5091	0.5091
Maximum $x_{1i}^{'}\beta$	1.0000	0.4909	1.0000	-0.5091	0.5091

4.4.5 When the Explanatory Variable is Dichotomous

Various attempts were made to estimate a model where the explanatory variable is a dummy variable. One example used the same equation as Model 1 only with a binary explanatory variable that is equal to 1 with a 50% probability. A simple probit model on the true (correctly classified) data estimates that $\Pr(\tilde{y}_i = 1)$ is either 0.0006 or 0.1654. When some proportion of the data is misclassified the approach experiences serious convergence problems.[38] In previous sections α_0 and α_1 were identified from subgroups of the data with the most negative and positive characteristics. When the explanatory variable is binary, there are only two subgroups in the data with distinct values of the explanatory variable so the estimator tries to use the entire subgroup where x_i=1 to estimate α_0 and similarly the entire subgroup where x_i=0 to estimate α_1 (i.e. the entire dataset is needed to just estimate the two misclassification probabilities).

Increasing the number of dummy regressors helps the Hausman *et al.* estimator converge to a solution. Model 10 includes 3 dummy variables as follows:

Model 10: $y_i^* = -1 - 2.5x_{1i} + 4x_{2i} - 0.5x_{3i} + \varepsilon_i$

where x_{1i}=1 with probability ½, $x_{2i} = 1$ with probability 1/3 and x_{3i}=1 with probability ¼. In this case the range of $\Pr(\tilde{y}_i = 1)$ is from 0.0001 to 0.7004. The results of the Monte Carlo simulations are shown in Table 4.9. The estimates of α_0 for 1 per cent and 20 per cent misclassification are reasonably close to their true values and the average of the standard errors of the estimates are quite low. At first glance, the estimator of α_0 when there is 5 per cent and 10 per cent misclassification does not perform quite as well. However, it is important to note that when there is 1 per cent and 20 per cent misclassification in the simulations only 814 and 788 respectively of the models converge and this is probably accounts for why some of the results appear superior to the 5 per cent and 10 per cent cases. When the

[38] Different models with a single dummy explanatory variable were simulated; where the effect of β_1 is very strong, with and without a constant term and when the true population parameters are used as starting values. In each case, the iterations do not make any progress from about the third iteration.

misclassification rate is 5 per cent, each simulation estimates $1-\alpha_0$, $1-\alpha_1$ and $-\beta$. Using the true values of β as starting values in the simulations (Model 11 in Table 4.9) results in α_0, α_1 and β being estimated. When there is 10 per cent symmetric misclassification the average of the estimates of α_0 is being driven by some extreme estimates. The median estimates of α_0 are near to their true values; although in the case of 5 per cent misclassification this is only true when the true values of β are used as starting values in the simulations. The estimates for α_1 are inferior to the estimates of α_0. They are further away from their true values, the range of estimates is wider and more negative probabilities are estimated. In addition, when the level of misclassification increases there are more instances of extreme estimates of α_1. While using the true values of β as starting values in the simulations results in there being no extreme estimates of α_0 when there is 10 per cent misclassification, the same is not true for the estimates of α_1 when there is 10 per cent and 20 per cent misclassification.

The third panel of Table 4.9 shows the simulation results when we extend Model 10 to include 5 dummy variables as follows:

$$\text{Model 12: } y_i^* = -1 - 2.5x_{1i} + 4x_{2i} - 0.5x_{3i} + 1.5x_{4i} - 1x_{5i} + \varepsilon_i$$

where x_{1i}=1 with probability ½, $x_{2i}=1$ with probability 1/3, x_{3i}=1 with probability ¼, x_{4i}=1 with probability ¼ and x_{5i}=1 with probability 0.4. In this case the range of $\Pr(\tilde{y}_i = 1)$ if from 6.37e-07 to 0.9817. This model performs better than Models 10 and 11. The average of the estimates are quite close to their true values, the models converge in practically every case and no extreme values are estimated.

Table 4.9: Estimates of Misclassification Probabilities when the Models only include Dummy Variables*

Level of Misc'n	Actual Misc'n Rate	$\hat{\alpha}_0$	Median $\hat{\alpha}_0$	Std Error $\hat{\alpha}_0$	MSE $\hat{\alpha}_0$	Min $\hat{\alpha}_0$	Max $\hat{\alpha}_0$	Actual Misc'n Rate	$\hat{\alpha}_1$	Median $\hat{\alpha}_1$	Std Error $\hat{\alpha}_1$	MSE $\hat{\alpha}_1$	Min $\hat{\alpha}_1$	Max $\hat{\alpha}_1$	No. of Sims
Model 10: $y_i^* = -1 - 2.5x_{1i} + 4x_{2i} - 0.5x_{3i} + \varepsilon_i$ *(Note: 3 binary regressors)*															
1%	0.0100	0.0077	0.0078	0.0041	0.0000	-0.0050	0.0173	0.0100	-0.0068	-0.0023	0.1478	0.0027	-0.2684	0.1027	814
5%	0.0501	0.9554	0.9531	0.0097	0.8196	0.9353	0.9990	0.0500	1.0120	0.9759	0.2261	0.9594	0.7381	4.1972	987
10%	0.1000	16.5083	0.1002	36.9314	242694	-0.0487	14983.15	0.0999	-248063	0.1153	603011.50	5.69E+13	-229000000	1.1132	925
20%	0.2001	0.1794	0.1913	0.0404	0.0038	-1.1008	0.2246	0.2003	-4008068	0.1726	11900000	2.62E+15	-908000000	0.3925	788
Model 11: $y_i^* = -1 - 2.5x_{1i} + 4x_{2i} - 0.5x_{3i} + \varepsilon_i$ *(Note: same 3 binary regressors as Model 10 and the true values of β given as starting values)*															
1%	0.0100	0.0081	0.0088	0.0038	0.0000	-0.0050	0.0173	0.0100	-0.0043	0.0008	0.1460	0.0025	-0.2683	0.1028	975
5%	0.0500	0.0446	0.0469	0.0096	0.0001	0.0010	0.0647	0.0501	-0.0115	0.0241	0.2234	0.0372	-3.1789	0.2619	997
10%	0.1000	0.0907	0.0953	0.0159	0.0004	-0.1133	0.1155	0.1000	-224707.40	0.0661	578306.90	4.87E+13	-217000000	0.3155	966
20%	0.2000	0.1793	0.1910	0.0432	0.0039	-1.1036	0.2246	0.2006	-1681924	0.1750	5202606	5.9E+14	-519000000	0.3925	778
Model 12:: $y_i^* = -1 - 2.5x_{1i} + 4x_{2i} - 0.5x_{3i} + 1.5x_{4i} - 1x_{5i} + \varepsilon_i$ *(Note: 5 binary regressors)*															
*1%***	*0.0100*	*0.0100*	*0.0100*	*0.0023*	*0.0000*	*0.0029*	*0.0177*	*0.0100*	*-0.0064*	*-0.0069*	*0.0089*	*0.0003*	*-0.0210*	*0.0192*	*1,000*
*5%***	*0.0500*	*0.0493*	*0.0494*	*0.0046*	*0.0000*	*0.0361*	*0.0628*	*0.0501*	*0.0313*	*0.0313*	*0.0150*	*0.0005*	*-0.0095*	*0.0743*	*1,000*
10%	0.1000	0.0992	0.0993	0.0062	0.0000	0.0812	0.1165	0.0999	0.0800	0.0804	0.0199	0.0007	0.0208	0.1512	995
20%	0.2001	0.1991	0.1993	0.0081	0.0000	0.1739	0.2206	0.2002	0.1822	0.1836	0.0262	0.0009	0.0734	0.2581	1,000

* In each case the sample size is 10,000 and the results are based on the average values from the Monte Carlo simulations

** In this set of Monte Carlo simulations, the true values of β are used as starting values, otherwise each simulation produces estimates of 1-α_0, 1-α_1 and $-\beta$

4.4.6 Coefficient Estimates from Hausman et al. and Probit Models

Table 4.10 compares the Hausman *et al.* estimates for β from the simulations with those from a standard probit model. In each case the Hausman *et al.* estimator outperforms a standard probit estimator, although there is evidence of bias as the level of misclassification increases and when the sample size is small.[39]

When observations are misclassified we expect the probit estimates to be biased downwards. Intuitively, in a model without misclassification observations with strong positive characteristics are the ones we expect to see with $y_i = 1$, if the model fits the data well. However, if some of these observations are misclassified as having $y_i = 0$, then we will observe observations with these strong characteristics that have $y_i = 1$ as well as some with $y_i = 0$. As the standard probit model ignores misclassification, the estimated effect of this characteristic will be weaker as there are more observations with these strong characteristics with $y_i = 0$. The striking feature of the probit estimates is the extent of their bias and how the bias increases as the level of misclassification increases. For example, in Model 1 when there is 10 per cent misclassification, the average estimate of β_1 is -0.5 compared to the true value of -1.5. Also, their average standard errors are much smaller than the corresponding Hausman *et al.* estimates; in fact the average standard errors of the probit estimates fall as the level of misclassification increases. This implies that not only are the probit estimates biased but that their precision is exaggerated.

---

[39] The average Hausman *et al.* estimate of β_1 for Model 4 when there is 1% symmetric misclassification is superior to the probit estimate despite the fact that the average estimate of α_1 (given in Table 4.6) is quite poor; however the MSE of $\hat{\beta}_0$ from the probit model is below that of the Hausman *et al.* model. Also the Hausman *et al.* estimator only converges in 542 out of 1,000 simulations while probit estimates are available for all simulations. In this set of simulations there are 15 cases where the estimate of α_0 is excessively large in magnitude – these estimates range from -7863.92 to -535392.6. For this subset of cases, the average Hausman *et al.* estimate of β_1 is -0.0623 and the average estimate of β_0 is 4.0998. The comparable probit estimates are -1.1761 for β_1 and -0.8080 for β_0. For this subset of cases, where extreme misclassification probabilities are estimated, the probit estimates are much closer to the true β than the Hausman *et al.* estimates. However, when the true values for β_0 and β_1 are used as starting values in the simulations of the Hausman *et al.* model, the estimates have a lower MSE than the probit estimates, although the Hausman *et al.* estimator only converges in 677 out of 1,000 simulations.

The table also reports the MSEs associated with the estimates of β. For Models 1 to 9, in practically every case, the MSEs of the Hausman *et al.* estimates are below the MSEs of the probit estimates.[40] For Model 10, which includes three dummy regressors, the MSEs associated with the probit estimator are generally below the Hausman *et al.* estimator, apart from when the misclassification rates are high. However, when the true parameters for β are used as starting values in the simulations (see Model 11 in Table 4.10), the Hausman *et al.* estimates outperform the probit estimates in terms of having lower MSEs, although the model does not always converge. In Model 12, where the number of dummy regressors increases to five, the Hausman *et al.* estimator is superior to the probit estimator as their MSEs are lower and the model converges in almost every case.

[40] There are two cases where the MSEs of the probit estimates are below those of the Hausman *et al.* estimates. The first was mentioned in the previous footnote and the second is the estimate of β_0 from Model 9.

Table 4.10: Comparison of Hausman *et al.* Coefficient Estimates with Probit Model*

Level of Misclassification	True β_0	$\hat{\beta}_{0HAUS}$	Standard Error $\hat{\beta}_{0HAUS}$	MSE $\hat{\beta}_{0HAUS}$	$\hat{\beta}_{0PROBIT}$	Standard Error $\hat{\beta}_{0PROBIT}$	MSE $\hat{\beta}_{0PROBIT}$	True β_1	$\hat{\beta}_{1HAUS}$	Standard Error $\hat{\beta}_{1HAUS}$	MSE $\hat{\beta}_{1HAUS}$	$\hat{\beta}_{1PROBIT}$	Standard Error $\hat{\beta}_{1PROBIT}$	MSE $\hat{\beta}_{1PROBIT}$	No. of Sims Hausman
Model 1: $y_i^* = -1 - 1.5x_{1i} + \varepsilon_i$															
1%	-1	-0.9311	0.0311	0.0049	-0.7233	0.0217	0.0771	-1.5	-1.4343	0.0368	0.0046	-1.1020	0.0202	0.1596	1,000
5%	-1	-0.9208	0.0402	0.0069	-0.4621	0.0168	0.2895	-1.5	-1.4190	0.0492	0.0077	-0.6964	0.0120	0.6461	1,000
10%	-1	-0.9131	0.0502	0.0090	-0.3403	0.0150	0.4354	-1.5	-1.4061	0.0623	0.0113	-0.5086	0.0093	0.9830	1,000
20%	-1	-0.9079	0.0754	0.0132	-0.2129	0.0135	0.6197	-1.5	-1.3952	0.0947	0.0186	-0.3150	0.0074	1.4043	1,000
Model 2: $y_i^* = -1 - 1.0x_{1i} + \varepsilon_i$ *(Note: Weaker Effect of* x_{1i}*)*															
1%	-1	-0.9638	0.0285	0.0015	-0.8370	0.0205	0.0269	-1.0	-0.9905	0.0255	0.0003	-0.8533	0.0157	0.0219	999
5%	-1	-0.9599	0.0381	0.0023	-0.5942	0.0165	0.1649	-1.0	-0.9855	0.0351	0.0008	-0.5965	0.0109	0.1630	1,000
10%	-1	-0.9577	0.0486	0.0032	-0.4519	0.0148	0.3005	-1.0	-0.9811	0.0451	0.0017	-0.4477	0.0088	0.3051	1,000
20%	-1	-0.9572	0.0745	0.0066	-0.2895	0.0135	0.5050	-1.0	-0.9758	0.0698	0.0048	-0.2822	0.0073	0.5153	1,000
Model 3: $y_i^* = -1 - 0.5x_{1i} + \varepsilon_i$ *(Note: Weaker Effect of* x_{1i}*)*															
1%	-1	-0.9440	0.0311	0.0033	-0.9111	0.0176	0.0080	-0.5	-0.4975	0.0214	0.0001	-0.4661	0.0101	0.0012	1,000
5%	-1	-0.9388	0.0420	0.0047	-0.7482	0.0156	0.0635	-0.5	-0.4923	0.0293	0.0005	-0.3699	0.0087	0.0170	1,000
10%**	-1	-0.9374	0.0540	0.0060	-0.6104	0.0144	0.1519	-0.5	-0.4894	0.0376	0.0011	-0.2925	0.0078	0.0431	1,000
20%	-1	-0.9387	0.0838	0.0097	-0.4148	0.0133	0.3426	-0.5	-0.4868	0.0585	0.0031	-0.1910	0.0069	0.0955	1,000
Model 4: $y_i^* = -1 - 1.5x_{1i} + \varepsilon_i$ *(Note: Sample size is 1,000)*															
1%	-1	-0.8095	0.0977	0.7277	-0.7466	0.0712	0.0700	-1.5	-1.3719	0.1077	0.0671	-1.1091	0.0641	0.1642	542
Model 4: 1% Misclassification with the true values of β *given as starting values*															
1%	*-1*	*-0.9428*	*0.1002*	*0.0051*	*-0.7466*	*0.0712*	*0.0700*	*-1.5*	*-1.3910*	*0.1157*	*0.0158*	*-1.1091*	*0.0641*	*0.1642*	*677*
5%	-1	-0.9216	0.1280	0.0132	-0.4692	0.0537	0.2841	-1.5	-1.3618	0.1431	0.0285	-0.7014	0.0379	0.6409	992
10%	-1	-0.9160	0.1630	0.0244	-0.3430	0.0476	0.4333	-1.5	-1.3481	0.1839	0.0455	-0.5134	0.0295	0.9747	998
20%	-1	-0.9387	0.2651	0.0738	-0.2137	0.0429	0.6197	-1.5	-1.3738	0.3061	0.1196	-0.3191	0.0234	1.3952	999

* In each case the sample size is 10,000 unless otherwise stated and the results are based on the average values from 1,000 Monte Carlo simulations on the Probit Models and the number of simulations indicated in the table for the Hausman *et al.* estimator

** The true parameters for β are used as starting values in this set of simulations of the Hausman *et al.* model; otherwise each simulation produces estimates of $1-\alpha_0$, $1-\alpha_1$ and $-\beta$

Table 4.10 cont'd: Comparison of Hausman *et al.* Coefficient Estimates with Probit Model*

Level of Misclassification	True β_0	$\hat{\beta}_{0HAUS}$	Standard Error $\hat{\beta}_{0HAUS}$	MSE $\hat{\beta}_{0HAUS}$	$\hat{\beta}_{0PROBIT}$	Standard Error $\hat{\beta}_{0PROBIT}$	MSE $\hat{\beta}_{0PROBIT}$	True β_1	$\hat{\beta}_{1HAUS}$	Standard Error $\hat{\beta}_{1HAUS}$	MSE $\hat{\beta}_{1HAUS}$	$\hat{\beta}_{1PROBIT}$	Standard Error $\hat{\beta}_{1PROBIT}$	MSE $\hat{\beta}_{1PROBIT}$	No. of Sims Hausman
Model 5: $y_i^* = -1 - 0.5x_{1i} + \varepsilon_i$ *(Note: Sample size is 100,000 and the equation is the same as in Model 3)*															
1%	-1	-0.9896	0.0101	0.0001	-0.9487	0.0057	0.0026	-0.5	-0.5039	0.0069	0.0000	-0.4652	0.0032	0.0012	1,000
5%	-1	-0.9869	0.0136	0.0003	-0.7765	0.0050	0.0500	-0.5	-0.5023	0.0096	0.0001	-0.3662	0.0027	0.0179	1,000
10%**	-1	-0.9859	0.0175	0.0004	-0.6324	0.0046	0.1351	-0.5	-0.5020	0.0124	0.0001	-0.2882	0.0024	0.0449	1,000
20%	-1	-0.9857	0.0268	0.0008	-0.4295	0.0042	0.3255	-0.5	-0.5029	0.0193	0.0003	-0.1873	0.0022	0.0978	1,000
Model 6: $y_i^* = 1 - 1.5x_{1i} + \varepsilon_i$ *(Note: Change in Intercept)*															
1%	1	0.9681	0.0324	0.0012	0.7342	0.0220	0.0713	-1.5	-1.4735	0.0380	0.0011	-1.1187	0.0206	0.1468	1,000
5%	1	0.9612	0.0419	0.0023	0.4605	0.0168	0.2913	-1.5	-1.4655	0.0520	0.0026	-0.6982	0.0121	0.6432	1,000
10%	1	0.9594	0.0526	0.0033	0.3367	0.0150	0.4401	-1.5	-1.4626	0.0668	0.0045	-0.5085	0.0094	0.9832	1,000
20%	1	0.9574	0.0795	0.0068	0.2093	0.0135	0.6254	-1.5	-1.4607	0.1033	0.0109	-0.3149	0.0074	1.4046	999
Model 7: $y_i^* = -1 - 1.5x_{1i} + \varepsilon_i$ *(Note: α_0 = 5% and α_1 = 20%)*															
α_0 =5% and α_1 =20%	-1	-0.9174	0.0472	0.0081	-0.6239	0.0162	0.1417	-1.5	-1.4103	0.0658	0.0109	-0.5116	0.0097	0.9771	1,000
Model 8: $y_i^* = -1 - 1.5x_{1i} + \varepsilon_i$ *(Note: α_0 = 25% and α_1 =1%)*															
α_0 =25% and α_1 =1%	-1	-0.9169	0.0620	0.0093	0.1282	0.0142	1.2730	-1.5	-1.4169	0.0608	0.0090	-0.4344	0.0086	1.1355	917
Model 9: $y_i^* = -1 - 1.5x_{1i} + \varepsilon_i$ *(Note: α_0 = 5% and α_1 = 50%)*															
α_0 =5% and α_1 =50%	-1	-0.9149	0.0725	0.0113	-0.9233	0.0164	0.0061	-1.5	-1.4084	0.1099	0.0186	-0.3201	0.0085	1.3922	1,000

* In each case the sample size is 10,000 unless otherwise stated and the results are based on the average values from 1,000 Monte Carlo simulations on the Probit Models and the number of simulations indicated in the table for the Hausman *et al.* estimator

** The true parameters for β are used as starting values in this set of simulations of the Hausman *et al.* model; otherwise each simulation produces estimates of $1-\alpha_0$, $1-\alpha_1$ and $-\beta$

Table 4.10 cont'd: Comparison of Hausman *et al.* Coefficient Estimates with Probit Model*

	True Parameter	Estimate	Standard Error	Mean Square Error	No. of Sims Hausman	Estimate	Standard Error	Mean Square Error	No. of Sims Hausman	Estimate	Standard Error	Mean Square Error
	Hausman *et al.* Estimates					**Hausman *et al.* Estimates**				**Probit Estimates**		
	Model 10: $y_i^* = -1 - 2.5x_{1i} + 4x_{2i} - 0.5x_{3i} + \varepsilon_i$					**Model 11:** $y_i^* = -1 - 2.5x_{1i} + 4x_{2i} - 0.5x_{3i} + \varepsilon_i$				**Model 10:** $y_i^* = -1 - 2.5x_{1i} + 4x_{2i} - 0.5x_{3i} + \varepsilon_i$		
						1% Symmetric Misclassification						
β_0	-1	-0.9609	0.1064	0.0025	814	-0.9610	0.1058	0.0025	975	-0.9326	0.0230	0.0046
β_1	-2.5	-2.6428	68.6906	1.9340		-2.9137	53.5238	2.1231		-1.2982	0.0859	1.4514
β_2	4	4.1042	68.7541	1.9930		4.3795	53.5770	2.1683		2.7378	0.0865	1.6003
β_3	-0.5	-0.5134	0.0819	0.0012		-0.5154	0.0818	0.0013		-0.4854	0.0387	0.0004
						5% Symmetric Misclassification						
β_0	-1	0.9743	0.1403	3.9051	987	-0.9740	0.1399	0.0079	997	-0.8341	0.0221	0.0277
β_1	-2.5	2.8425	64.4951	30.8031		-2.8462	59.0481	2.3431		-0.7288	0.0515	3.1397
β_2	4	-4.2733	64.5896	70.9349		4.2775	59.1415	2.5257		2.0096	0.0527	3.9641
β_3	-0.5	0.5027	0.1108	1.0116		-0.5030	0.1107	0.0061		-0.3826	0.0358	0.0143
						10% Symmetric Misclassification						
β_0	-1	-0.5365	0.2037	1.0278	925	-1.0107	-1.0107	0.0805	966	-0.7201	0.0213	0.0786
β_1	-2.5	-1.3974	80.4521	8.8452		-2.7048	70.9468	2.5065		-0.4938	0.0430	4.0265
β_2	4	2.1128	80.6066	18.5293		4.0995	71.0948	2.9189		1.5987	0.0447	5.7679
β_3	-0.5	-0.2514	0.1497	0.2589		-0.4921	0.1488	0.0156		-0.2997	0.0335	0.0408
						20% Symmetric Misclassification						
β_0	-1	-1.1870	0.3794	0.7897	788	-1.1692	0.3768	0.6880	778	-0.5177	0.0203	0.2329
β_1	-2.5	-2.2521	109.1156	2.9820		-2.2503	123.4358	2.9994		-0.2784	0.0370	4.9369
β_2	4	3.6152	109.5305	4.3640		3.6180	124.0054	4.3703		1.0733	0.0393	8.5668
β_3	-0.5	-0.5047	0.5090	0.0839		-0.5055	0.7908	0.0896		-0.1914	0.0310	0.0960

* In each case the sample size is 10,000 unless otherwise stated and the results are based on the average values from 1,000 Monte Carlo simulations on the Probit Models and the number of simulations indicated in the table for the Hausman *et al.* estsimator

Table 4.10 cont'd: Comparison of Hausman *et al.* Coefficient Estimates with Probit Model*

	True Parameter	Estimate	Standard Error	Mean Square Error	No. of Sims Hausman	Estimate	Standard Error	Mean Square Error
		Hausman *et al.* Estimator				Probit Model		
	Model 12: $y_i^* = -1 - 2.5x_{1i} + 4x_{2i} - 0.5x_{3i} + 1.5x_{4i} - 1x_{5i} + \varepsilon_i$							
	*1% Symmetric Misclassification***							
β_0	-1	-0.9487	0.0325	0.0028	1,000	-0.8860	0.0290	0.0131
β_1	-2.5	-2.3260	0.1864	0.0392	1,000	-1.6264	0.0832	0.7712
β_2	4	3.7980	0.1983	0.0510	1,000	3.0354	0.0870	0.9396
β_3	-0.5	-0.5021	0.0445	0.0002	1,000	-0.4631	0.0404	0.0016
β_4	1.5	1.4772	0.0522	0.0009	1,000	1.3708	0.0396	0.0171
β_5	-1	-1.0662	0.0454	0.0047	1,000	-0.9863	0.0372	0.0005
	*5% Symmetric Misclassification***							
β_0	-1	-0.9444	0.0415	0.0039	1,000	-0.7467	0.0269	0.0644
β_1	-2.5	-2.3690	1.3780	0.1549	1,000	-0.9219	0.0525	2.4936
β_2	4	3.8268	1.3951	0.1753	1,000	2.0790	0.0551	3.6937
β_3	-0.5	-0.5044	0.0519	0.0009	1,000	-0.3465	0.0365	0.0242
β_4	1.5	1.4622	0.0653	0.0031	1,000	1.0530	0.0348	0.2003
β_5	-1	-1.0530	0.0553	0.0038	1,000	-0.7406	0.0328	0.0679
	10% Symmetric Misclassification							
β_0	-1	-0.9445	0.0524	0.0047	995	-0.6267	0.0256	0.1397
β_1	-2.5	-2.5657	10.1775	0.7905	995	-0.6281	0.0435	3.5057
β_2	4	4.0157	10.1989	0.8093	995	1.5813	0.0459	5.8521
β_3	-0.5	-0.5071	0.0618	0.0020	995	-0.2668	0.0339	0.0551
β_4	1.5	1.4539	0.0815	0.0060	995	0.8328	0.0324	0.4458
β_5	-1	-1.0463	0.0672	0.0046	995	-0.5686	0.0301	0.1868
	20% Symmetric Misclassification							
β_0	-1	-0.9495	0.0808	0.0078	1,000	-0.4410	0.0242	0.3129
β_1	-2.5	-2.9918	29.4103	2.2350	1,000	-0.3621	0.0371	4.5718
β_2	4	4.4399	29.4414	2.2325	1,000	1.0192	0.0397	8.8868
β_3	-0.5	-0.5112	0.0890	0.0060	1,000	-0.1695	0.0311	0.1101
β_4	1.5	1.4540	0.1244	0.0146	1,000	0.5510	0.0304	0.9013
β_5	-1	-1.0414	0.0991	0.0089	1,000	-0.3613	0.0275	0.4085

* In each case the sample size is 10,000 unless otherwise stated and the results are based on the average values from 1,000 Monte Carlo simulations on the Probit Models and the number of simulations indicated in the table for the Hausman *et al.* estimator

** The true parameters for β are used as starting values in these sets of simulations of the Hausman *et al.* model; otherwise each simulation produces estimates of $1-\alpha_0$, $1-\alpha_1$ and β

4.5 Extensions of the Model where the Estimator Performs Poorly

This section reports the results of Monte Carlo simulations on models where the results from Hausman *et al.* estimator are disappointing. It also examines under what circumstances the estimates can be improved.

4.5.1 Poor Performance of the Estimator

In practice, when using this estimator, independent variables in a model may be discrete and of limited range (e.g. years of work experience). To show how the Hausman *et al.* estimator performs in this instance Model 13 employs the same basic equation as Model 1 only a different explanatory variable is included. In this case the independent variable, x_{2i}, is drawn from a uniform distribution and transformed to generate 11 distinct equally spaced numbers over the interval –0.5 to 0.5 i.e. $x_{2i} = [-0.5,-0.4,-0.3,-0.2,-0.1,0,0.1,0.2,0.3,0.4,0.5]$. The top panel of Table 4.11 reports the average values from the Monte Carlo simulations for this model. The results show that the Hausman *et al.* estimator performs very poorly in this instance. In some cases, the average estimates of the misclassification probabilities are being driven by one or a few extreme values but even when there are no extreme estimates the resulting average estimates are quite biased. The table also reports the median estimates of the misclassification probabilities and the median estimates of α_1 are also quite far from their true values. In addition, when the amount of misclassification increases the estimator does not always converge. Even when the true values for β_0 and β_1 are used as starting values in the simulations, the results remain very disappointing (see Model 14 reported in the second panel of Table 4.11).

The difficulty with estimating the misclassification probabilities for Model 13, in particular for estimating α_1, appears to be that that the range of $x_i'\beta$ is too limited. The misclassification probabilities are identified off cases that have very high and low $x_i'\beta$ so if the range of $x_i'\beta$ is very limited then the Hausman *et al.* estimator cannot accurately identify the misclassification probabilities. In fact in Model 13, the range of the true $\Pr(\tilde{y}_i = 1)$ is between 0.0496 and 0.3993.

Table 4.11: Estimates of Misclassification Probabilities for Models 13-14*

Level of Misc'n	Actual Misc'n Rate	$\hat{\alpha}_0$	Median $\hat{\alpha}_0$	Std Error $\hat{\alpha}_0$	MSE $\hat{\alpha}_0$	Min $\hat{\alpha}_0$	Max $\hat{\alpha}_0$	Actual Misc'n Rate	$\hat{\alpha}_1$	Median $\hat{\alpha}_1$	Std Error $\hat{\alpha}_1$	MSE $\hat{\alpha}_1$	Min $\hat{\alpha}_1$	Max $\hat{\alpha}_1$	No. of Sims
Model 13: $y_i^* = -1 - 1.5x_{2i} + \varepsilon_i$ *(Note: Range of* x_{2i} *is -0.5 to 0.5)*															
1%	0.0100	0.0278	0.0285	0.0233	0.0004	-0.0065	0.0477	0.0100	0.38	0.3988	0.24	0.1452	-0.12	0.5024	1,000
5%	0.0500	0.0559	0.0620	0.0352	0.0008	-0.1129	0.1024	0.0503	-906293.2	0.3685	2710431	8.06E+14	-898000000	0.5509	1,000
10%	0.1000	0.0914	0.1072	0.0458	0.0024	-0.3922	0.1512	0.1001	-1865653	0.3523	5808773	6.13E+14	-599000000	0.5945	994
20%	0.1999	116389.2	0.2220	185537	1.07E+12	-0.3084	21100000	0.2005	-4837976	0.4923	21600000	1.66E+15	-696000000	1.3035	959
Model 14: $y_i^* = -1 - 1.5x_{2i} + \varepsilon_i$ *(Note: same as Model 13 with the true values of* β *given as starting values)*															
1%	0.0099	0.0402	0.0495	0.0824	0.0022	-0.1193	0.1002	0.0100	-146.3043	0.3638	768.9898	1926722	-24176.33	0.6071	999
5%	0.0501	-0.0413	0.0886	0.9174	8.4999	-91.6524	0.1542	0.0500	-848581.2	0.4285	5291148	2.79E+14	-464000000	0.6666	991
10%	0.1000	-110.9671	0.1368	558.37	11300000	-104336.8	0.2062	0.1000	-625675	0.4585	4327738	6.01E+13	-137000000	0.6585	967
20%	0.2002	-41352.48	0.2276	1439354	4.51E+11	-17300000	0.2964	0.1996	-2185396	0.5187	18100000	4.2E+14	-336000000	0.6635	949

* In each case the sample size is 10,000 and the results are based on the average values from the Monte Carlo simulations

4.5.2 Extending the Range of the Explanatory Variable

One way to increase the range of $x_i'\beta$ is to increase the range of the explanatory variable. This idea is explored in Model 15 where the independent variable, x_{3i}, is drawn from a uniform distribution but this time is transformed to generate 41 distinct equally spaced numbers over the interval –2 to +2. The true coefficients on the constant and explanatory variable are the same as Model 13. The average values from the Monte Carlo simulations are reported in Table 4.12. In this case, the Hausman *et al.* estimator performs much better; for every level of misclassification considered the average estimates of the misclassification probabilities are very close to their true values. Also, the range of estimates for the misclassification probabilities is more plausible. This highlights how important it is to have cases with very low and very high values of $x_i'\beta$ in order to identify the misclassification probabilities. In this case the range of $\Pr(\tilde{y}_i = 1)$ is from 0.00005 to 0.9776. This model also has a higher proportion of observations with $\tilde{y}_i = 1$ than Model 13; 0.344 compared to 0.192 in Model 13 and this also helps with the identification of the model.

In Model 15 both the range and the amount of distinct numbers in the explanatory variable were increased. In many practical applications, researchers cannot affect the range of an explanatory variable in a model but sometimes they can affect the number of categories in, say, a discrete variable. For example, data on education, occupation and sector may be collected at a detailed level but a researcher may collapse some of the categories. To examine the effect this may have on the estimates, Model 16 generates the data in the exact same way as Model 15, however, the model is estimated using an explanatory variable that collapses the 41 distinct values of the explanatory variable in Model 15 to just 4 values. The simulation results are reported in the second panel of Table 4.12. The results indicate that the average estimate of α_0 remains close to its true value for each level of misclassification considered; however the average estimates of α_1 are considerably further away from their true values than in Model 15.

4.5.3 Increasing the Importance of the Explanatory Variable

A second situation that gives rise to an expansion in the range of $x_i'\beta$ is when the importance of the explanatory variable is increased. This is examined in Model 17 where the range of the explanatory variable is the same as in Model 13 and the coefficient on the constant term is also the same but the coefficient on the explanatory variable is four times more negative (i.e. β_1=-6). The idea is to investigate whether the Hausman *et al.* estimator yields more sensible estimates if a variable is more important in explaining y_i, even when its range is more limited. The simulation results are reported in the third panel of Table 4.12. The results show that for each level of misclassification considered the estimates of the misclassification probabilities are far superior to those obtained for Model 13. In Model 17, the range of $\Pr(\tilde{y}_i = 1)$ is from 0.0000357 to 0.9753194 and the proportion of observations with $\tilde{y}_i = 1$ is 0.348, again much higher than in Model 13.

Table 4.12: Estimates of Misclassification Probabilities for Models 15-17*

Level of Misc'n	Actual Misc'n Rate	$\hat{\alpha}_0$	Median $\hat{\alpha}_0$	Std Error $\hat{\alpha}_0$	MSE $\hat{\alpha}_0$	Min $\hat{\alpha}_0$	Max $\hat{\alpha}_0$	Actual Misc'n Rate	$\hat{\alpha}_1$	Median $\hat{\alpha}_1$	Std Error $\hat{\alpha}_1$	MSE $\hat{\alpha}_1$	Min $\hat{\alpha}_1$	Max $\hat{\alpha}_1$	No. of Sims
Model 15: $y_i^* = -1 - 1.5x_{3i} + \varepsilon_i$ *(Note: Same as Model 13 but range of* x_i *is increased to -2 to +2)*															
1%	0.0100	0.0092	0.0091	0.0020	0.0000	0.0038	0.0157	0.0100	0.0095	0.0095	0.0105	0.0000	-0.0054	0.0245	1,000
5%	0.0500	0.0488	0.0488	0.0042	0.0000	0.0340	0.0612	0.0501	0.0469	0.0467	0.0140	0.0001	0.0122	0.0820	1,000
10%	0.1000	0.0988	0.0988	0.0057	0.0000	0.0808	0.1210	0.1000	0.0949	0.0952	0.0169	0.0002	0.0455	0.1472	1,000
20%	0.2001	0.1989	0.1990	0.0075	0.0000	0.1725	0.2211	0.2002	0.1937	0.1947	0.0210	0.0004	0.1158	0.2499	1,000
Model 16: $y_i^* = -1 - 1.5x_{3i} + \varepsilon_i$ *(Note: same as Model 15 but fewer categories in* x_i *are included in estimated model)*															
1%	0.0100	0.0092	0.0091	0.0024	0.0000	0.0033	0.0157	0.0100	-0.0261	-0.0251	0.0382	0.0015	-0.0763	0.0105	1,000
5%	0.0500	0.0493	0.0493	0.0050	0.0000	0.0310	0.0656	0.0501	0.0142	0.0161	0.0455	0.0022	-0.0930	0.0906	1,000
10%	0.1000	0.0994	0.0995	0.0068	0.0000	0.0752	0.1233	0.1000	0.0643	0.0703	0.0539	0.0033	-0.1112	0.1593	1,000
20%	0.2001	0.1989	0.1994	0.0094	0.0001	0.1634	0.2275	0.2002	0.1585	0.1783	0.0812	0.0119	-1.7699	0.2863	1,000
Model 17: $y_i^* = -1 - 6x_{2i} + \varepsilon_i$ *(Note: same as Model 13 but* β_1 *is increased in magnitude to -6)*															
1%	0.0100	0.0093	0.0092	0.0019	0.0000	0.0030	0.0168	0.0100	0.0101	0.0099	0.0087	0.0000	-0.0057	0.0242	1,000
5%	0.0500	0.0486	0.0487	0.0041	0.0000	0.0359	0.0617	0.0501	0.0474	0.0476	0.0119	0.0001	0.0123	0.0730	1,000
10%	0.1000	0.0984	0.0983	0.0055	0.0000	0.0818	0.1137	0.1000	0.0958	0.0964	0.0145	0.0002	0.0477	0.1362	1,000
20%	0.2001	0.1984	0.1982	0.0073	0.0000	0.1774	0.2200	0.2002	0.1947	0.1962	0.0182	0.0003	0.1271	0.2366	1,000

* In each case the sample size is 10,000 and the results are based on the average values from the Monte Carlo simulations

Finally Table 4.13 compares the estimates for the coefficient on the constant, and the coefficient on the explanatory variable, from the simulations of the Hausman *et al*. estimator for Models 13 and 15 to 17 with those from a standard probit model. The probit estimates for Model 13 outperform the Hausman *et al.* estimates as they have lower MSEs associated with them. However, the reverse is true for Models 15 and 17, where the range of the explanatory variable in the models is increased, as the Hausman *et al.* estimates are closer to their true values and have lower MSEs than the probit estimates. In addition, the estimates show that the bias in the probit estimates increases as the level of misclassification increases. The results for Model 16 are more mixed. The probit estimates for β_0 are superior to the Hausman *et al.* estimates while the Hausman *et al.* estimates for β_1 outperform the probit estimates.

Table 4.13: Comparison of Coefficient Estimates from Hausman *et al.* Model and Probit Model for Models 13, 15-17*

Level of Misclassification	True β_0	$\hat{\beta}_{0HAUS}$	Standard Error $\hat{\beta}_{0HAUS}$	MSE $\hat{\beta}_{0HAUS}$	$\hat{\beta}_{0PROBIT}$	Standard Error $\hat{\beta}_{0PROBIT}$	MSE $\hat{\beta}_{0PROBIT}$	True β_1	$\hat{\beta}_{1HAUS}$	Standard Error $\hat{\beta}_{1HAUS}$	MSE $\hat{\beta}_{1HAUS}$	$\hat{\beta}_{1PROBIT}$	Standard Error $\hat{\beta}_{1PROBIT}$	MSE $\hat{\beta}_{1PROBIT}$	No. of Sims Hausman
Model 13: $y_i^* = -1 - 1.5x_{2i} + \varepsilon_i$ *(Note: Range of* x_{2i} *is -0.5 to 0.5)*															
1%	-1	-0.6787	0.2438	0.1069	-0.9405	0.0157	0.0036	-1.5	-2.1282	0.7265	0.4403	-1.4269	0.0503	0.0056	1,000
5%	-1	-0.8345	0.4105	0.3406	-0.8216	0.0148	0.0319	-1.5	-1.9529	0.8797	0.5134	-1.1830	0.0471	0.1013	1,000
10%	-1	-1.2103	0.5553	1.4863	-0.6980	0.0140	0.0913	-1.5	-1.7885	0.9539	0.7857	-0.9581	0.0447	0.2947	994
20%	-1	-0.5885	0.6482	5.5117	-0.4939	0.0132	0.2563	-1.5	-0.6924	1.0955	4.7299	-0.6351	0.0419	0.7494	959
Model 15: $y_i^* = -1 - 1.5x_{3i} + \varepsilon_i$ *(Note: Same as Model 13 but range of* x_i *is increased to -2 to +2)*															
1%	-1	-0.9367	0.0295	0.0042	-0.8221	0.0215	0.0319	-1.5	-1.4650	0.0473	0.0017	-1.3115	0.0228	0.0359	1,000
5%	-1	-0.9332	0.0388	0.0052	-0.5762	0.0170	0.1798	-1.5	-1.4518	0.0628	0.0041	-0.9711	0.0166	0.2800	1,000
10%	-1	-0.9332	0.0491	0.0059	-0.4317	0.0150	0.3231	-1.5	-1.4438	0.0800	0.0071	-0.7549	0.0139	0.5553	1,000
20%	-1	-0.9345	0.0746	0.0087	-0.2715	0.0135	0.5308	-1.5	-1.4376	0.1226	0.0163	-0.4934	0.0119	1.0133	1,000
Model 16: $y_i^* = -1 - 1.5x_{3i} + \varepsilon_i$ *(Note: same as Model 15 but fewer categories in* x_i *are included in estimated model)*															
1%	-1	1.0254	0.1271	4.1043	1.0638	0.0265	4.2596	-1.5	-1.3282	0.0759	0.0305	-1.2697	0.0218	0.0533	1,000
5%	-1	1.0240	0.1626	4.1086	0.8473	0.0240	3.4129	-1.5	-1.3297	0.1054	0.0353	-0.9687	0.0167	0.2825	1,000
10%	-1	1.0240	0.2126	4.1285	0.6861	0.0227	2.8433	-1.5	-1.3338	0.1446	0.0445	-0.7630	0.0143	0.5434	1,000
20%	-1	1.0420	0.7819	4.3527	0.4662	0.0215	2.1502	-1.5	-1.3628	0.6767	0.1389	-0.5042	0.0124	0.9918	1,000
Model 17: $y_i^* = -1 - 6x_{2i} + \varepsilon_i$ *(Note: same as Model 13 but* β_1 *is increased in magnitude to -6)*															
1%	-1	-0.9415	0.0300	0.0036	-0.8134	0.0219	0.0351	-6	-5.8353	0.1819	0.0342	-5.1565	0.0891	0.7179	1,000
5%	-1	-0.9335	0.0391	0.0051	-0.5577	0.0171	0.1958	-6	-5.7630	0.2385	0.0823	-3.7677	0.0634	4.9879	1,000
10%	-1	-0.9303	0.0494	0.0064	-0.4138	0.0151	0.3438	-6	-5.7236	0.3033	0.1334	-2.9173	0.0527	9.5058	1,000
20%	-1	-0.9277	0.0748	0.0095	-0.2582	0.0135	0.5504	-6	-5.6818	0.4634	0.2608	-1.9021	0.0447	16.7949	1,000

* In each case the sample size is 10,000 unless otherwise stated and the results are based on the average values from 1,000 Monte Carlo simulations on the Probit Models and the number of simulations indicated in the table for the Hausman *et al.* procedure

4.6 Conclusions

There are many sources of misclassification in models that employ binary dependent variables, such as coding error, self-reporting, recall error and where a dummy variable is used to serve as a proxy for some true underlying variable. Misclassification in a binary dependent variable results in estimates that are biased and inconsistent.

This chapter uses Monte Carlo simulations to explore the performance of the Hausman *et al.* estimator which controls for misclassification. It shows that identification of the misclassification probabilities comes from misclassified observations in the tails of the $x_i'\beta$ distribution. It finds that the estimator outperforms a probit model even when the effect of the explanatory variable is weak. In the case of asymmetric misclassification the estimator is superior to a probit model although the model does not always converge when one of the rates of misclassification is low. When the explanatory variables in the model are dichotomous the Hausman *et al.* estimator is superior to a probit estimator if there is a considerable range and support in $x_i'\beta$. However, when the range of $x_i'\beta$ is more limited the estimator is only better when misclassification rates are high. The chapter also finds that when the sample size is reduced, although some of the estimates of the misclassification probabilities are not sensible, especially for small amounts of misclassification, the Hausman *et al.* estimator still outperforms a probit model, especially when starting values equal to the true parameters are given. However, the results also show that the Hausman *et al.* model fails to converge in many incidences when the sample size is low and the rates of misclassification are also low.

In cases where the true underlying model has a limited range of $x_i'\beta$, some of the Hausman *et al.* estimates of the misclassification probabilities are not sensible. As the misclassification probabilities are identified off observations with very high and low values of $x_i'\beta$, it is necessary to have sufficient range and variation in a dataset that can have such values. The simulation results indicate that when the range and support of $x_i^{'}\beta$ is limited, the coefficient estimates from the probit model are superior to the Hausman *et al.* estimates.

Given the severity of the bias in the probit estimates and the fact that the bias increases as the level of misclassification increases it is important to control for it in cases where misclassification is likely. From a practitioners' point of view, looking for consistency in data or comparing data to a second source may be useful when misclassification is suspected. In this situation, the Hausman *et al.* estimator can provide a useful alternative to a probit estimator. The simulation results reported in the chapter indicate that it is important to be working with a sizeable sample when using the Hausman *et al.* estimator. It is also imperative to closely consider the variables included in the basic regression, as considerable range and support in $x_i^{'}\beta$ is needed to identify the extent of misclassification in the data. Finally, it may be important to experiment with a range of starting values.

Appendix Table 4.1: Comparison of Predicted Probabilities with Observed Proportion of 1's by Groups of $x_{1i}^{'}\beta$ for Model 1 with 5% Symmetric Misclassification

Groups of $x_{1i}^{'}\beta$	Proportion of True $\tilde{y}_i = 1$	Proportion of Observed $y_i = 1$	Predicted $\Pr(y_i = 1)$	Proportion of Observed $y_i = 1$ - Predicted $\Pr(y_i = 1)$	Predicted $\Pr(y_i = 1)$- Proportion of Observed $y_i = 1$
Minimum $x_{1i}^{'}\beta$	0.0000	0.0434	0.0000	0.0434	-0.0434
2-100	0.0000	0.0511	0.0000	0.0511	-0.0511
101-500	0.0000	0.0498	0.0000	0.0498	-0.0498
501-1000	0.0000	0.0503	0.0000	0.0503	-0.0503
1001-2000	0.0000	0.0502	0.0001	0.0501	-0.0501
2001-3000	0.0010	0.0510	0.0030	0.0480	-0.0480
3001-4000	0.0200	0.0680	0.0244	0.0436	-0.0436
4001-5000	0.0910	0.1317	0.1001	0.0316	-0.0316
5001-6000	0.3060	0.3250	0.2837	0.0413	-0.0413
6001-7000	0.5530	0.5476	0.5659	-0.0183	0.0183
7001-8000	0.8320	0.7987	0.8353	-0.0366	0.0366
8001-9000	0.9780	0.9304	0.9740	-0.0436	0.0436
9001-9500	0.9960	0.9459	0.9990	-0.0531	0.0531
9501-9900	1.0000	0.9496	1.0000	-0.0504	0.0504
9901-9990	1.0000	0.9485	1.0000	-0.0515	0.0515
9991-9999	1.0000	0.9484	1.0000	-0.0516	0.0516
Maximum $x_{1i}^{'}\beta$	1.0000	0.9484	1.0000	-0.0516	0.0516

Appendix Table 4.2: Comparison of Predicted Probabilities with Observed Proportion of 1's by Groups of $x_{1i}^{'}\beta$ for Model 1 with 10% Symmetric Misclassification

Groups of $x_{1i}^{'}\beta$	Proportion of True $\tilde{y}_i = 1$	Proportion of Observed $y_i = 1$	Predicted $\Pr(y_i = 1)$	Proportion of Observed $y_i = 1$ - Predicted $\Pr(y_i = 1)$	Predicted $\Pr(y_i = 1)$- Proportion of Observed $y_i = 1$
Minimum $x_{1i}^{'}\beta$	0.0000	0.0972	0.0000	0.0972	-0.0972
2-100	0.0000	0.1019	0.0000	0.1019	-0.1019
101-500	0.0000	0.0991	0.0000	0.0991	-0.0991
501-1000	0.0000	0.1002	0.0000	0.1002	-0.1002
1001-2000	0.0000	0.1001	0.0002	0.1000	-0.1000
2001-3000	0.0010	0.1009	0.0033	0.0976	-0.0976
3001-4000	0.0200	0.1161	0.0255	0.0906	-0.0906
4001-5000	0.0910	0.1728	0.1022	0.0706	-0.0706
5001-6000	0.3060	0.3444	0.2852	0.0592	-0.0592
6001-7000	0.5530	0.5426	0.5650	-0.0224	0.0224
7001-8000	0.8320	0.7651	0.8329	-0.0678	0.0678
8001-9000	0.9780	0.8829	0.9728	-0.0899	0.0899
9001-9500	0.9960	0.8966	0.9989	-0.1023	0.1023
9501-9900	1.0000	0.8998	1.0000	-0.1002	0.1002
9901-9990	1.0000	0.8989	1.0000	-0.1011	0.1011
9991-9999	1.0000	0.8974	1.0000	-0.1026	0.1026
Maximum $x_{1i}^{'}\beta$	1.0000	0.8974	1.0000	-0.1026	0.1026

Appendix Table 4.3 Comparison of Predicted Probabilities with Observed Proportion of 1's by Groups of $x_{1i}^{'}\beta$ for Model 1 with 20% Symmetric Misclassification

Groups of $x_{1i}^{'}\beta$	Proportion of True $\tilde{y}_i = 1$	Proportion of Observed $y_i = 1$	Predicted $\Pr(y_i = 1)$	Proportion of Observed $y_i = 1$ - Predicted $\Pr(y_i = 1)$	Predicted $\Pr(y_i = 1)$- Proportion of Observed $y_i = 1$
Minimum $x_{1i}^{'}\beta$	0.0000	0.1867	0.0000	0.1867	-0.1867
2-100	0.0000	0.2019	0.0000	0.2019	-0.2019
101-500	0.0000	0.1990	0.0000	0.1990	-0.1990
501-1000	0.0000	0.2010	0.0000	0.2010	-0.2010
1001-2000	0.0000	0.2003	0.0002	0.2001	-0.2001
2001-3000	0.0010	0.2011	0.0037	0.1974	-0.1974
3001-4000	0.0200	0.2116	0.0268	0.1847	-0.1847
4001-5000	0.0910	0.2549	0.1041	0.1508	-0.1508
5001-6000	0.3060	0.3835	0.2862	0.0973	-0.0973
6001-7000	0.5530	0.5317	0.5638	-0.0320	0.0320
7001-8000	0.8320	0.6985	0.8301	-0.1316	0.1316
8001-9000	0.9780	0.7873	0.9712	-0.1838	0.1838
9001-9500	0.9960	0.7970	0.9986	-0.2016	0.2016
9501-9900	1.0000	0.7997	1.0000	-0.2003	0.2003
9901-9990	1.0000	0.7942	1.0000	-0.2058	0.2058
9991-9999	1.0000	0.7947	1.0000	-0.2053	0.2053
Maximum $x_{1i}^{'}\beta$	1.0000	0.7947	1.0000	-0.2053	0.2053

Appendix Table 4.4: Proportion of $\tilde{y}_i = 1$ in Each Model

	Proportion of $\tilde{y}_i = 1$
Model 1	0.3779
Model 2	0.3312
Model 3	0.2475
Model 4	0.3770
Model 5	0.2375
Model 6	0.6208
Model 7	0.3779
Model 8	0.3779
Model 9	0.3779
Model 10	0.2675
Model 11	0.2675
Model 12	0.2816
Model 13	0.1902
Model 14	0.1902
Model 15	0.3440
Model 16	0.3440
Model 17	0.3522

5. Measurement Error in Survey Data: A Model of Job Mobility for Ireland

5.1 Introduction

Many studies of labour market dynamics use survey data. Therefore it is valuable to know about the quality of the data collected. There may be ambiguity in a survey question, respondents may misunderstand the question, they may have an incentive to misreport, they may have poor recall or responses may be coded incorrectly. This chapter investigates job mobility or employment-to-employment transitions in Ireland over the period 1995 to 2001 using the Living in Ireland Survey (LIS), the Irish component of the European Community Household Panel (ECHP). As is common with many surveys, there is no direct question in the LIS about job mobility; instead it is inferred from the responses of individuals to a question about tenure. The chapter highlights a potentially serious measurement error problem in the responses used to determine job changes. As a result, there is a risk of misclassifying cases as being job changes when truly no change took place and vice versa. The extent of measurement error is similar to what has been found in other studies (e.g. Brown and Light (1992)).

In estimating the determinants of job mobility it is important to control for misclassification, otherwise it can lead to estimates that are biased and inconsistent. The estimator developed by Hausman, Abrevaya and Scott-Morton (1998) and discussed in detail in Chapter 4 is used to control for misclassification in the dependent variable. The chapter finds that by ignoring misclassification the true number of job changes may be substantially underestimated. In addition, ignoring misclassification leads to diminished covariate effects in models of job change.

The chapter is organised as follows: Section 5.2 examines the reasons for and prevalence of reporting errors in labour market survey data and, in particular, focuses on studies relevant to job mobility. Section 5.3 explores the extent of measurement error in the LIS data. Hausman *et al.* (1998) show how their estimator can be extended to allow for covariate dependent measurement error and this is presented in Section 5.4. Section 5.5 provides estimation results and Section 5.6 concludes.

5.2 Labour Market Survey Data

Many studies of job mobility use survey data and usually surveys do not contain a direct question asking if the respondent has changed jobs in the past year. Instead job changes are inferred from the length of time an employee reports to have been with their current employer. Therefore questions about tenure play a crucial role in many empirical studies of job mobility. There are several reasons to suspect that responses to questions about tenure are measured with error. Respondents may find it difficult to remember when they started working in their current job. Bound *et al.* (2001) describe studies that categorise the question and answer process in a survey as a four-stage procedure. These stages include understanding the question, recovering the information from memory, considering whether the information matches what was requested and communicating the response. Much of the measurement error literature focuses on the stage where respondents retrieve the information from memory. A general principle from this literature is that the longer the length of the recall period the greater the expected bias due to respondent retrieval error. Therefore we might expect respondents with longer tenure to be most likely to misreport tenure. In one sense, this does not pose a serious problem for calculating job changes as job changes are associated with people who have short tenures; provided those with longer tenures who misreport do not significantly underestimate their tenure. Farber (1999) and Ureta (1992) find a heaping of tenure responses at round counts of years or round calendar years and this rounding indicates that individuals do not provide precise responses about tenure.

There may also be ambiguity in the wording of the question about tenure or there may be changes to the wording of the question in other waves of a survey. Farber (1999) points out how the mobility supplements to the Current Population Survey in the US from 1951 to 1981 asked workers what year they "…started working at their present job or business" while in later years the supplement asked workers how many years they have "…been working continuously for the present employer". The earlier question refers to time on the present job rather than time with the present employer. Workers may experience other types of internal labour mobility (e.g. promotion, reassignment) which means that their tenure on the job will be shorter than their tenure with the employer. The interviewer notes for the LIS provide clarity in distinguishing between employer changes and other types of internal labour mobility as they state that the question refers to when they started working with their present employer even if there have been position changes with that employer. In addition, there were no changes to the wording of the question about tenure in the LIS. The interviewer notes in the LIS do not provide guidance on how to handle interrupted employment spells (in particular when someone returns to a previous employer). Farber (1999) mentions that if no reference is made to the continuity of employment that the natural inclination of workers will be to ignore interruptions of "reasonable" length.

Brown and Light (1992) examine the extent of measurement error in tenure responses in the Panel Study of Income Dynamics (PSID). They find that tenure responses are frequently inconsistent with calendar time.[41] In addition, they perform a validation exercise to gauge the accuracy of their measure of job changes. They adopt various definitions of job mobility (based on tenure responses) and use them to partition the data into distinct jobs. They assess the accuracy of the various definitions by comparing the number of jobs and the number of times each job is observed with those identified by the National Longitudinal Survey (NLS). The NLS contains unique employer codes which can be compared across interviews and so provides a more accurate count of the 'true' number of jobs.

Brown and Light (1992) investigate various measures of job mobility and examine which one performs best when there is measurement error in tenure data. One definition of job mobility they employ is to assume that a job change has taken place whenever reported tenure is less than the time elapsed since the previous interview. If tenure was never misreported and if respondents never returned to previous employers then this method would identify job changes without error. They also adopt another set of definitions of job mobility by assuming that a job change occurs whenever the change in tenure between adjacent interviews varies "too much" or "too little" in either direction. In another definition, a job change is defined whenever the change in tenure is not exactly equal to the change in calendar time between interviews. This permits no inconsistency in tenure responses within jobs.

[41] The level of inconsistencies in reported tenure in the PSID is described in Section 5.3.2 where comparisons are made to the LIS data.

They also adopt more flexible measures that permit various amounts of inconsistency in reported tenure within jobs. They define another four measures of job mobility when the change in tenure differs from the change in calendar time by more than 6, 12, 18 and 24 months in either direction. As these latter definitions identify job changes when tenure changes by “too much” as well as by “too little” they are more likely to separate continuing jobs[42] but less likely to link jobs that are truly separate[43] than when job changes are defined as occurring whenever reported tenure is less than the time elapsed since the previous interview. They find the definition of job mobility that is the most accurate when compared to the NLS data is that a job change has occurred whenever reported tenure is less than the time elapsed between interviews. This is essentially the definition of job change that is adopted in my study.

These types of validation studies are also useful because they provide evidence on the magnitude of the measurement error in tenure data. Bound *et al.* (2001) point to the fact that few studies have investigated the quality of tenure data. Duncan and Hill (1985) present results from a validation study of a large manufacturing company in which administrative records are used to validate survey responses from a sample of workers from the company. Overall they find very little evidence of bias in the interview reports. They find that reported tenure is typically quite accurate; 45 per cent of the sample accurately reported the year they were hired and 90 per cent were able to report year of hire to within one year. However, the unit of analysis in the study is defined in terms of years and these types of error margins in a dataset could be problematic if we were to use the measure of tenure to calculate job changes. As job changes are identified from those who report short tenures the under or over

---

[42] For example, consider an individual who truly hasn’t changed jobs and who reports tenure of 24 months in one interview and exactly a year later misreports their tenure and says they have been in their job for 45 months when their true tenure is 36 months. When we define a job change as having occurred when reported tenure is less than the time between interviews then we conclude this person hasn’t changed jobs. However, if we define job mobility as occurring when the change in reported tenure differs from the change in calendar time by more then, say, 12 months then we classify this person as having changed jobs.

[43] For example, consider an individual who truly has changed jobs and is interviewed 12 months apart. In the first year they report tenure of 5 months and in the subsequent year they report tenure of one month. When we define a job change as having occurred when reported tenure is less then the time between interviews then we conclude this person has changed jobs. Using the other definitions of job mobility we would not classify this person as having changed jobs.

reporting of tenure by a year, in particular by those with short tenures, could lead us to misclassify job changes and vice versa.

Bound *et al.* also cite a study where workers' reported starting dates are compared to employer records. The study by Weiss *et al.* (1961) finds that 71 per cent of jobs in the prior 5 years had reported starting dates within one month of company records. They also find that validity significantly declines as a function of the length of time between the job start date and the date of interview. To capture job mobility, tenure, at least for those who have not been in their jobs long, needs to be reported accurately. These validation studies suggest that the quality of tenure data may not be sufficient to do this.

5.3 Measurement Error

5.3.1 Dataset and Defining Job Changes

This chapter uses the same sample of workers from the LIS and defines job changes in an equivalent way as described in Sections 3.2.1 and 3.2.2. As mentioned in Section 3.2.2, there is no explicit question in the LIS about whether or not a person has changed jobs; instead job mobility is inferred from responses to the question about when they started working with their present employer. If a person is employed in two consecutive years and in the second year they report a starting date that falls between the two interview dates we conclude that this person has changed jobs during that period. Table 5.1 shows the number of workers employed in consecutive two-year periods from the revolving balanced panel and the number of job changes each year.[44]

Table 5.1: Number of Workers and Job Changes

	1995	1996	1997	1998	1999	2000	2001
Number of workers	1,163	1,175	1,211	1,276	1,341	1,376	1,434
No. Job Changes	76	85	102	139	146	184	156
Job Mobility Rate	6.5%	7.2%	8.4%	10.9%	10.9%	13.4%	10.9%

[44] This table replicates Table 3.2 in Chapter 3.

However, in the absence of exogenous job change information we cannot be certain that the number of job changes reported in Table 5.1 is correct.[45] For example, a worker may forget their starting date, they may misunderstand the question or their response could be coded incorrectly. In addition, respondents may consider multiple spells with the same employer differently. These problems could be overcome if the LIS contained a direct question about changing jobs or if it contained unique employer codes that could be compared across interviews.

Responses to questions about tenure are frequently inconsistent. For example, in one year a person may report that they started working with their current employer say in January 1995 while the following year they may report that they started working with their current employer in January 1993. The concern in this chapter is not necessarily that tenure is misreported but rather that if tenure is misreported there is a risk that cases may be misclassified as job changes and vice versa. For example, suppose a worker is interviewed in January 1995 and January 1996 and in January 1996 they report that they started working with their current employer is January 1993. Using the measure of job-to-job mobility defined above, we would conclude that no job change has taken place between the interviews in 1995 and 1996. However, suppose this person cannot accurately recall when they started working with their current employer and they misreport their starting date to be January 1995. Then we would erroneously conclude that this person has changed jobs between their interviews in 1995 and 1996. Now, suppose that their true starting date is January 1995 so that they have truly changed jobs between interviews but they misreport their starting date to be January 1993. In this case we would erroneously conclude that no job change has taken place between interviews. In an attempt to ascertain how reliable the responses to the question about when a worker started working with their current employer are the next section examines the consistency of these responses over time.

Given that we cannot be sure that the number of job changes defined in Table 5.1 is correct, an alternative assumption that exploits the monthly activity reports contained in the LIS was also used to identify job changes. The resulting measure of job

[45] No attempt is made to distinguish between different types of mobility. Respondents are asked to state why their last job ended and usually we would use their answer to determine whether the job change was voluntary or involuntary. However, if we cannot be certain that job changes have been accurately identified, we cannot know which of a worker's previous jobs the response refers to.

mobility is discussed in Appendix 5A; however it appears implausibly low for a booming economy with a flexible labour market and so is not used in the analysis.

5.3.2 Consistency of Starting Dates within Jobs

Given the possibility of measurement error in the data we need to try to ascertain how reliable the information on starting dates is and therefore how useful it is for deducing job changes. If there were no measurement error in the data then starting dates would be constant *within* jobs. By partitioning the dataset into distinct jobs and comparing starting dates across interviews we can investigate how consistent the data is.

To convert this data into separate jobs, we begin with the 1995 data. There are 1,163 workers in 1995 but as 76 workers changed jobs a total of 1,239 distinct jobs are observable in that year (see Table 5.1). For this year alone, the previous jobs of those who changed jobs are excluded from the analysis. We only have one observation on their previous jobs (the starting date in 1994) so we cannot check the consistency of responses whereas we can track the new jobs across subsequent interviews. Therefore, we start the analysis with 1,163 distinct jobs in 1995. In each subsequent year one of four alternatives occurs:

1) A worker can stay in their job so the total number of jobs remains the same and we observe the job surviving an additional year.
2) A worker can drop out of the sample if they enter a period of unemployment, leave the labour force for more than a year or if they are over the age of 60. In this case, the total number of jobs remains the same but we no longer observe that particular job. Workers who are unemployed or leave the labour force may re-enter the analysis in later years.
3) A worker can change jobs and accordingly the total number of jobs increases by one and we stop observing the previous job.
4) There can be a new entrant to the sample of workers. This would be someone from the revolving balanced panel who is now 20 and so was excluded in earlier years. This increases the total number of jobs by one. In addition, a worker who was unemployed or out of the labour force may come back into the analysis and this would increase the total number of jobs observed by one.

This results in 2,529 jobs observable for various durations over the period 1995 to 2001. Of these jobs, there are 1,755 jobs observable for more than one year and it is this set of jobs that is considered in the analysis in this section (so there are at least two starting dates to compare for each job).[46] Table 5.2 shows how many jobs display consistency in starting dates. Of the 1,755 jobs considered in the analysis, only 352 or 20 per cent have the same reported starting date in each year the job is observed.

If we adopt a less stringent definition of consistency such as all starting dates being within 3 months of each other then 37 per cent of jobs meet the criterion. If we relax the criterion further to consider jobs where all starting dates are within 6 months of each other then 42 per cent of jobs display consistent responses. This leaves 1,014 or 58 per cent of jobs that survive for more than one year where all starting dates do not fall within 6 months of each other.

Table 5.2: Consistency of Starting Dates within Jobs

Source: Living in Ireland Survey					
		Jobs with All Starting Dates:			
	Jobs	Equal	Within 3 months	Within 6 months	Remaining jobs
Number of jobs	1,755	352	649	741	1,014
% of Jobs		20%	37%	42%	58%
Source: Panel Study of Income Dynamics, taken from Brown and Light (1992)					
Number of jobs	3,318	246	1,170	1,514	1,804
% of Jobs		7%	35%	46%	54%

This level of inconsistency in starting dates is quite alarming; however, it is in line with what has been found in other datasets. Brown and Light (1992) take a sample from the Panel Study of Income Dynamics (PSID) from 1976 to 1985, partition the data into distinct jobs in an analogous fashion and examine the consistency of reported start dates within jobs. They find that only 7 per cent of the jobs in their

[46] This means there are 774 jobs that only survive for one year. For example, in 2001 there are 257 jobs recorded as surviving for one year. Of this figure, 156 are job changers in 2001, 22 observations are new entrants to the sample of workers (i.e. 20 year olds) and the remaining 79 observations have rejoined the sample of workers having been unemployed or out of the labour force. As the survey ends in 2001, there is no later data to compare the starting dates in 2001 to.

sample have identical starting dates in each year the job is observed, while 54 per cent of jobs have starting dates that do not all fall within 6 months of each other.

Brown and Light (1992) highlight another aspect of this definition of consistency that may be quite restrictive. Suppose a job is observed in every year of the LIS and every starting date given between 1995 and 2001 is equal with the exception of one which is different to the others by 7 months, then this job will not meet any of the measures of consistency define above. They argue most researchers would agree that this outlier could be ‘fixed’ to match up to the other observations for that job. One can further extend the measure of consistency used in Table 5.2 by requiring that only a majority of observations for a given job be in agreement.

Table 5.3 shows how many jobs have a majority of starting dates in agreement. A total of 654 jobs or 37 per cent have a majority of starting dates in agreement, while 84 per cent of all jobs identified have a majority of starting dates that are within 3 months of each other. The bottom panel of the table reports comparable statistics for the PSID taken from Brown and Light (1992). As before, the magnitudes of the consistency measures are broadly comparable with the Irish data. Given that both datasets display similar discrepancies, it is likely that any study using a similar question to deduce job changes contains measurement error.

Table 5.3: Consistency of the Majority of Starting Dates within Jobs

Source: Living in Ireland Survey					
		Jobs with a Majority of Starting Dates:			
	All Jobs	Equal	Within 3 months	Within 6 months	Remaining jobs
Number of jobs	1,755	654	1,471	1,513	242
% of Jobs		37%	84%	86%	14%
Source: Panel Study of Income Dynamics, taken from Brown and Light (1992)					
Number of jobs	3,318	676	2,116	2,471	847
% of Jobs		20%	64%	74%	26%

The method for partitioning the dataset into distinct jobs uses job changes to identify when one job ends and another one begins. The analysis presented in this section implies that the measure of job change may not accurately identify the true number of job changes i.e. there are probably cases identified as job changes when no change in jobs took place and vice versa. This means we may over or underestimate the true number of jobs and therefore the level of inconsistent starting dates within jobs. For example, consider a person who truly hasn't changed jobs and started working in their current job in January 1993 and in 1995 they report their true starting date but in 1996 they report that they started working in January 1996. This case will be erroneously classified as a job change in 1996; two separate jobs will be identified for this person and no inconsistency in starting dates will be recorded. In addition, suppose this person truly changed jobs in January 1996 but they misreport the starting date to be January 1995. This will be recorded as one job with an inconsistent starting date in 1996 when it is truly two jobs with an incorrect starting date in 1996. As the number of job changes (and jobs) is measured imperfectly so too will the amount of inconsistencies evident in the data. If the definition of job change accurately allowed us to define jobs then Table 5.2 would show the true amount of inconsistencies within jobs. As the definition of job changes only provides us with an imperfect measure of the number of jobs the data presented in the table reflects both the inconsistencies in starting dates and the fact that we have not accurately identified the true number of jobs.

Tables 5.2 and 5.3 focus on the extent and magnitude of inconsistencies evident in the data and it is clear that there is the possibility of substantial measurement error. In this study, the main concern about measurement error is not directly that starting dates are misreported but rather that the misreporting of starting dates may cause cases to be misclassified as job changes and vice versa (that truly distinct jobs would be linked or continuing jobs would be separated).

There are cases where it is very unlikely that we will erroneously assign a case to be a job change, even though there are inconsistencies in starting dates. For example, suppose an individual gives the following starting dates in successive interviews for a job observed from 1995 to 1998: January 1980, February 1975, May 1972 and January 1982. It is unlikely that this person has changed jobs at any point between

1995 and 1998 but rather they found it difficult to recall the starting date of their job. Although the responses for this job would fail to meet any of the consistency measures described in Tables 5.2 and 5.3, all of the reported starting dates are sufficiently long ago so as to not cause much concern about erroneously defining recent job changes.

Of particular concern are inconsistencies in short jobs (i.e. where reported tenure is low). For example, suppose we observe a job every year between 1995 and 2001; it is more likely that this person has changed jobs at some point over this period and it has not been captured if the inconsistency in starting dates falls close to or within that period. However, if all inconsistencies in reported starting dates refer to a time period sufficiently far back then it is more likely that this person hasn't changed jobs recently and just cannot accurately recall when they started working in their current job.

Table 5.4 examines the timing of inconsistencies in reported starting dates within jobs. It takes the total number of jobs and reports how many of these jobs have the dates of all inconsistencies occurring at least three years prior to the date that we first start observing the job.[47] There are 722 jobs where all discrepancies fall reasonably far in the past so that these are probably truly continuing jobs. However, there are 681 jobs where the reported inconsistencies are more recent and it is more likely in these cases that we have linked jobs that are distinct or divided continuing jobs.

Table 5.4: Timing of Inconsistencies within Jobs

	No. of Jobs	Equal Starting Dates	All inconsistencies at least 3 years prior to date job is first observed	Remaining jobs
Number of jobs	1,755	352	722	681
% of Jobs		20%	41%	39%

If the true number of jobs has been under or overestimated (and therefore the true number of job changes has been under or overestimated), it may be more likely to do so for certain types of worker. As mentioned above, it may be more likely to under or overestimate the number of jobs where reported inconsistencies are recent (i.e. jobs

[47] For example, if we observe a job for the first time in 1995, this measure counts all jobs where each inconsistency refers to dates earlier than or in 1992.

that have inconsistencies when tenure is low). As tenure and years of work experience are correlated we may expect to see differences in inconsistencies in starting dates by years of experience. There may also be differences by gender; as women experience more interrupted employment spells than men it may be harder for them to accurately report starting dates. There could also be differences by full-time and part-time status because part-time workers are less attached to the labour force and so may be more likely to experience interrupted employment spells.

Table 5.5 examines the 681 jobs that have recent reported inconsistencies (from Table 5.4) by years of work experience, gender and whether a job is part-time or full-time. These jobs are labelled as ‘problematic’ in the table in the sense that it is more likely for these cases that truly distinct jobs would be linked or continuing jobs would be separated (i.e. that job changes are misclassified and vice versa). This is not to say that workers in short jobs are more likely to be misreport when they started working in their job than workers in long jobs, but rather that the misreporting by those in short jobs is more likely to lead to misclassifying cases as job changes and vice versa.

The table shows that the incidence of problematic jobs declines with years of work experience.[48] For example, 60 per cent of jobs that have less than ten years of work experience associated with them are classified as problematic and this percentage declines as years of experience increases so that only 17 per cent of jobs with more than 30 years of work experience associated with them fall into this category. As there are more of these problematic jobs in low experience categories and job mobility is negatively correlated with experience, this may indicate that we are underestimating the true number of job changes. There is also some difference when we look at the incidence of these problematic jobs by gender; 35 per cent of all jobs held by men fall into this category while the comparable figure for women is 10 percentage points higher. Similar figures are observed when the frequency of these problematic jobs is broken down by part-time and full-time employment status. This is unsurprising as women are more likely to be part-time workers.

[48] In assigning years of work experience to a worker in a job we use their experience in the first year that the job is observed. For example, if we observe a job each year between 1995 and 2001 the experience assigned to that job when comparing all combinations of starting dates over the period is the years of experience of that person in 1995. Similarly, a job is assigned as being part-time or full-time using the status in the first year the job is observed.

Table 5.5: 'Problematic' Jobs by Worker Characteristics

	Years of Work Experience				
	Total	<10	10-19	19-29	30+
No. Problematic Jobs	681	315	201	111	54
Total Number of Jobs	1,755	523	530	383	319
Problematic jobs as % of total	39%	60%	47%	29%	17%

	Gender		
	Total	Men	Women
No. Problematic Jobs	681	388	293
Total Number of Jobs	1,755	1,099	656
Problematic jobs as % of total	39%	35%	45%

	Full-Time/Part-Time Status[49]		
	Total	Full-Time	Part-Time
No. Problematic Jobs	678	542	136
Total Number of Jobs	1,746	1,456	290
Problematic jobs as % of total	39%	37%	47%

This section focussed on examining discrepancies in reported starting dates within jobs. It used jobs as the unit of analysis for looking at measurement error. In the remainder of the chapter, the focus will be on how measurement error may lead us to misclassify a worker in a given year as being a job changer and vice versa so the unit of analysis switches to workers.

5.4 Binary Choice Model with Misclassification

A binary choice model can be used to explain the decision to change jobs. Given the level of measurement error in the data, it is likely that incorrect inferences have been made about whether or not a worker has changed jobs so it is essential to control for misclassification. When the dependent variable is dichotomous, misclassification can lead to estimates that are biased and inconsistent. The empirical analysis in Section 5.5 uses the estimator developed by Hausman *et al.* (1998) which was analysed in detail in Chapter 4, to control for misclassification in the dependent variable. Section

[49] In this case, nine jobs are excluded from the analysis because the information on whether the job is full-time or part-time is missing from the LIS Survey.

5.4.1 discusses a simple extension of the estimator that allows for covariate dependent measurement error.

5.4.1 Covariate Dependent Misclassification

The analysis of the Hausman *et al.* estimator in Chapter 4 focussed on the case where misclassification is independent of the covariates or where the probabilities of misclassification are constant across all workers. Assuming that the probabilities of misclassification are constant across all types of workers may be quite restrictive. As indicated in Section 5.3.2 (in particular in Table 5.5) it is possible that the misclassification probabilities vary across different types of workers. Hausman *et al.* consider a simple extension to the model to allow for some covariate dependent misclassification error as follows:[50]

Assume that the misclassification probabilities depend on some or all of the covariates x_i:

$$\alpha_0(x_i) = \Pr(y_i = 1 | \tilde{y}_i = 0, x_i) \tag{5.1}$$

$$\alpha_1(x_i) = \Pr(y_i = 0 | \tilde{y}_i = 1, x_i) \tag{5.2}$$

The expected value of the observed dependent variable is given by:

$$E(y_i|x_i) = \Pr(y_i = 1|x_i) = \alpha_0(x_i) + (1 - (\alpha_0(x_i) - \alpha_1(x_i))F(x_i'\beta) \tag{5.3}$$

For example, suppose misclassification only depends on one covariate x_{1i}, then the expression given in (5.3) becomes:

$$E(y_i|x_i) = \Pr(y_i = 1|x_i) = \alpha_0(x_{1i}) + (1 - (\alpha_0(x_{1i}) - \alpha_1(x_{1i}))F(x_i'\beta) \tag{5.4}$$

The likelihood function is similar to equation (4.14) in Chapter 4 only the two misclassification probabilities appear as functions of x_{1i}. The model can be identified

[50] The notation is the same as that used in Chapter 4.

in a similar manner to what was described in Section 4.2.2 of Chapter 4. To see this, first consider the case where x_{1i} is a dummy variable:

$$\alpha_0 = \gamma_0 + \gamma_1 x_1 \tag{5.5}$$

and

$$\alpha_1 = \partial_0 + \partial_1 x_1 \tag{5.6}$$

Therefore when $x_1 = 1 \Rightarrow \alpha_0 = \gamma_0 + \gamma_1$ and $\alpha_1 = \partial_0 + \partial_1$ (5.7)

and when $x_1 = 0 \Rightarrow \alpha_0 = \gamma_0$ and $\alpha_1 = \partial_0$ (5.8)

In this case, the expected value of the observed dependent variable becomes:

$$E(y_i|x_i) = \Pr(y_i = 1|x_i) = \gamma_0 + \gamma_1 x_1 + (1 - (\gamma_0 + \gamma_1 x_1) - (\delta_0 + \partial_1 x_1))F(x_i'\beta) \tag{5.9}$$

Intuitively if misclassification depends on x_{1i}, then the probabilities of misclassification are constant within the two subgroups where $x_{1i} = 1$ and $x_{1i} = 0$. Then γ_0 is identified from the group of workers who truly have a very low probability of changing jobs and who have $x_{1i} = 0$, while γ_1 is identified from the group of workers who truly have a very low probability of changing jobs and who have $x_{1i} = 1$. Similar to what was described in Section 4.2.2, identification of γ_0 and γ_1 is achieved from the group of workers for whom $x_i'\beta$ is highly negative and who are therefore very unlikely to be job changers but some of them will be misclassified. However, in this case we effectively divide this group of workers, with very negative characteristics, into two subgroups where $x_{1i} = 1$ and $x_{1i} = 0$. A comparable argument can be made for the identification of δ_0 and δ_1 where the group of workers for whom $x_i'\beta$ is highly positive is used to identify the two parameters. When x_{1i} is a dummy variable there are four misclassification probabilities to estimate.

Another way of illustrating the identification of the model, is to note that in equation (5.9) as $x_i'\beta$ becomes more negative (where $x_i'\beta$ excludes x_{1i}) it tends to γ_0 when

$x_{1i}=0$ and to $\gamma_0+\gamma_1$ when $x_{1i}=1$. Similarly as $x_i^{'}\beta$ becomes more positive (where $x_i^{'}\beta$ excludes x_{1i}) it tends to $(1-\delta_0)$ when $x_{1i}=0$ and to $(1-\delta_0-\delta_1)$ when $x_{1i}=1$.

If x_{1i} is a discrete variable, then all observations within the subgroups that have the same values for x_{1i} will have constant misclassification probabilities and identification is achieved in a similar way as described above. However, if we have many subgroups it will be harder to estimate $\gamma_0, \gamma_1, \gamma_2$ etc and $\delta_0, \delta_1, \delta_2$ etc as there may not be sufficient workers in the tails of the index with the full range of values of x_{1i}. Finally, if x_{1i} is a continuous variable, we can define cut-off points c_1, c_2, c_3 etc such that if $x_{i1} < c_1$ the misclassification probabilities are constant (or almost constant) within each group.

As discussed in Section 4.2.2, identification is only possible because the misclassification probabilities enter the likelihood function in an additive way and their sum must be less than one.

5.5 Estimation Results

In this section, we formally control for misclassification in job changes in models of job mobility. The section begins by presenting results from job change models where misclassification is assumed to be independent of the covariates and then in Section 5.5.2 the analysis is extended to allow for covariate dependent measurement error.

5.5.1 Estimation Results: Misclassification Independent of Covariates

Table 5.6 shows the estimates from a standard probit regression of the probability of job change and the Hausman *et al.* estimates that control for misclassification. This provides estimates of the probabilities of misclassification and allows comparisons to be made of the effect of response error on the estimated coefficients. The data for 1995 to 2001 have been pooled so that there are 8,736 observations from which 851 job changes are identified.[51] The explanatory variables are defined in Appendix Table

[51] The total number of observations is 240 lower and the total number of job changes is 37 lower than reported in Table 5.1. These observations are excluded because data is missing or not available for at least one of the explanatory variables.

5.2. All the explanatory variables (apart from the year dummies) are lagged by a year so they refer to the workers' characteristics and situation prior to changing jobs or in the previous year. Zero misclassification probabilities are used as starting values in estimating the model with misclassification.[52]

The estimated probability of misclassification for non-job changers, α_0, is very small at less than one per cent and the estimated probability of misclassification for job changers, α_1, is high at 61 per cent. Significance tests on α_0 and α_1 can be used as tests of misclassification. Although α_0 is not significant, α_1 is highly significant and so we reject the model without misclassification. Workers who have truly changed jobs are more likely to be misclassified, as α_1 exceeds α_0. This means that the measure of job change is likely to undercount the true number of job changes. To put this estimate of α_1 in context, the average conventionally defined mobility rate in the dataset is around 9.7 per cent and the estimate implies that the true mobility rate is around 15.6 per cent.

Hausman *et al.* also apply the estimator to a model of job change using US data from the January 1987 Current Population Survey from the Census Bureau. Their study provides external estimates of the misclassification probabilities. They estimate α_0 to be 6.1 per cent and α_1 to be 30.9 per cent.

When we allow for misclassification, the estimated coefficients have higher standard errors implying that errors in responses lead to a loss in estimation efficiency. The results also indicate that ignoring misclassification leads to diminished covariate effects. It is easier to interpret differences in the estimates from the two regressions if

[52] A range of different starting values for α_0 and α_1 are used to check the robustness of the results. If α_1 is given a starting value of 0, the results are robust for any starting value of α_0 between 0 and 1. If α_0 is given a starting value of 0, the results are robust for any starting value of α_1 between zero and 0.90. Above this value the likelihood function encounters flat or discontinuous regions and the model does not converge. If starting values for β from the probit model are used then the results are robust for any starting value of α_1 up to 1. If α_0 and α_1 are given the same starting values, the results are robust up to starting values of 0.28. For higher starting values the likelihood function encounters flat or discontinuous regions and the model does not converge. If starting values for β from the probit model are used then the results are robust up to and including starting values of 1.

we look at marginal effects instead of coefficient estimates. These are presented in Table 5.7.

Table 5.6: Coefficient Estimates from Models of Job Change*

Variable	*Coefficient Estimates*	*Robust Standard Errors*	*Coefficient Estimates*	*Robust Standard Errors*
	Standard Probit Model		Misclassification Model	
$\hat{\alpha}_0$			0.0063	0.0049
$\hat{\alpha}_1$			0.6061	0.0763
Experience	-0.0655	0.0096	-0.1142	0.0263
Experience squared	0.0009	0.0002	0.0016	0.0005
Female	0.0465	0.0850	0.0637	0.1507
Child	-0.0085	0.0656	-0.0122	0.1083
Living in a Couple	0.0500	0.0923	0.0474	0.1584
Female* Living in a Couple	-0.2637	0.1080	-0.3776	0.1871
Education- medium	-0.1304	0.0576	-0.1953	0.1020
Education- high	-0.2502	0.0986	-0.4533	0.1935
(Ref: Education – low)				
Working Part-Time	0.3704	0.1044	0.8147	0.2697
Female* Working Part-Time	-0.1738	0.1271	-0.4672	0.2621
Recent Training	0.2148	0.0769	0.4722	0.1714
Public Sector	-0.2197	0.0810	-0.3479	0.1355
Number of Employees > 50	-0.1462	0.0521	-0.2555	0.1000
Overskilled	0.2131	0.0421	0.3453	0.0856
Occupation of Origin:				
(Ref: Elementary Occ's)				
Manager	-0.4555	0.1031	-0.7572	0.2440
Professional	-0.3413	0.0832	-0.5421	0.1534
Clerk	-0.2941	0.0749	-0.5239	0.1813
Skilled	-0.3361	0.0733	-0.4954	0.1298
Sector of Origin:				
(Ref: Non Market Services)				
Agriculture, Mining & Utilities	-0.3448	0.1210	-0.5364	0.2272
Manufacturing	-0.1833	0.1053	-0.3099	0.1805
Building	0.3898	0.1188	0.7448	0.2674
Market Services	0.1583	0.0846	0.2869	0.1509
Year Dummies:				
(Ref: 1995)				
1996	0.0297	0.0826	0.0703	0.1394
1997	0.0193	0.0908	0.0494	0.1529
1998	0.2050	0.0980	0.4047	0.2110
1999	0.1443	0.1119	0.2697	0.2046
2000	0.2505	0.1212	0.4783	0.2452
2001	0.1087	0.1327	0.2813	0.2565
Dublin Region	-0.1113	0.0832	-0.1970	0.1521
Regional Unemployment Rate	-0.0296	0.0246	-0.0458	0.0416
Constant	-0.2931	0.2346	0.9320	0.5204
N	8,736		8,736	
Wald chi2	540.67		69.71	
Prob > chi2	0.0000		0.0001	
Log pseudolikelihood	-2416.2094		-2410.1338	

* Note: Standard errors are adjusted to take account of the fact that there are multiple observations on the same people

Table 5.7: Marginal Effects from Models of Job Change*

Variable	*Marginal Effect*	*P>|Z|*	*Marginal Effect*	*P>|Z|*	*Ratio of Marginal Effects*
	Standard Probit Model		Misclassification Model		
$\hat{\alpha}_0$			0.0063	0.20	
$\hat{\alpha}_1$			0.6061	0.00	
Experience	-0.0088	0.00	-0.0427	0.00	4.9
Experience squared	0.0001	0.00	0.0006	0.00	5.1
Female	0.0063	0.58	0.0160	0.67	2.5
Child	-0.0011	0.90	-0.0030	0.91	2.7
Living in a Couple	0.0066	0.59	0.0117	0.77	1.8
Female* Living in a Couple	-0.0317	0.02	-0.0846	0.04	2.7
Education- medium	-0.0172	0.02	-0.0478	0.06	2.8
Education- high	-0.0293	0.01	-0.0950	0.02	3.2
(Ref: Education – low)					
Working Part-Time	0.0599	0.00	0.2532	0.00	4.2
Female* Working Part-Time	-0.0211	0.17	-0.0963	0.08	4.6
Recent Training	0.0329	0.01	0.1395	0.01	4.2
Public Sector	-0.0275	0.01	-0.0804	0.01	2.9
Number of Employees > 50	-0.0190	0.01	-0.0612	0.01	3.2
Overskilled	0.0288	0.00	0.0867	0.00	3.0
Occupation of Origin:					
(Ref: Elementary Occ's)					
Manager	-0.0464	0.00	-0.1356	0.00	2.9
Professional	-0.0404	0.00	-0.1175	0.00	2.9
Clerk	-0.0352	0.00	-0.1132	0.00	3.2
Skilled	-0.0393	0.00	-0.1071	0.00	2.7
Sector of Origin:					
(Ref: Non Market Services)					
Agriculture, Mining & Utilities	-0.0383	0.00	-0.1087	0.02	2.8
Manufacturing	-0.0226	0.08	-0.0701	0.09	3.1
Building	0.0659	0.00	0.2363	0.01	3.6
Market Services	0.0220	0.06	0.0745	0.06	3.4
Year Dummies:					
(Ref: 1995)					
1996	0.0040	0.72	0.0180	0.61	4.4
1997	0.0026	0.83	0.0125	0.75	4.8
1998	0.0306	0.04	0.1146	0.06	3.7
1999	0.0208	0.20	0.0733	0.19	3.5
2000	0.0381	0.04	0.1376	0.05	3.6
2001	0.0154	0.41	0.0765	0.27	5.0
Dublin Region	-0.0140	0.18	-0.0455	0.20	3.2
Regional Unemployment Rate	-0.0040	0.23	-0.0171	0.27	4.3
N	8,736		8,736		
Wald chi2	540.67		69.71		
Prob > chi2	0.0000		0.0001		
Log pseudolikelihood	-2416.2094		-2410.1338		

* Notes: Standard errors are adjusted to take account of the fact that there are multiple observations on the same people. The marginal effects are evaluated at the sample means of the explanatory variables.

Although both models indicate that the same factors determine job mobility the effect of misclassification in the dependent variable on the marginal effects of the various explanatory variables is sizeable. In the theoretical literature on job mobility, years of labour market experience is a key determinant of job change. Workers with less labour market experience are more likely to changes jobs as they have less knowledge of the labour market and their own preferences and abilities for different jobs. Both models provide findings that are consistent with this. However, in the model that does not allow for misclassification, an additional year of experience reduces the probability of changing jobs by 0.9 percentage points, while the marginal effect in the misclassification model is almost five times larger, so an additional year of experience reduces the probability of changing jobs by 4.3 percentage points.

The models of job mobility have a range of individual controls that include household structure and personal characteristics. We may expect women to be more likely to change jobs as they have a weaker attachment to the labour force but the results do not indicate any significant gender difference in the probability of changing jobs. The marginal effect of having children is small and insignificant implying that the presence of children does not affect the probability of changing jobs. Workers may be less likely to change jobs if they are more constrained by non-market variables, such as being married or living in a couple, and we would expect this effect to be stronger for women. When misclassification is controlled for, the results indicate that women who are married or living in a couple are 8.5 per cent less likely to change jobs and the result is significant at the 5 per cent level. This effect is more than double what the probit estimates imply.

The results also indicate that the negative effect of human capital on the probability of changing jobs is more marked in the misclassification model. For example, general human capital is proxied by education level and in the model incorporating misclassification the marginal effect of third level education is more than three times higher than in the probit model. The marginal effect indicates that those with third level education are 9.5 per cent less likely to change jobs than those who have at most Junior Certificate education. In addition, the marginal effects of higher levels of occupational attainment relative to those in elementary occupations are higher in the misclassification model. The results also show that workers who have undergone recent training are more likely to change jobs. This may reflect the fact that, typically, training is undertaken at the beginning of a job and there is a high hazard of new jobs ending early.

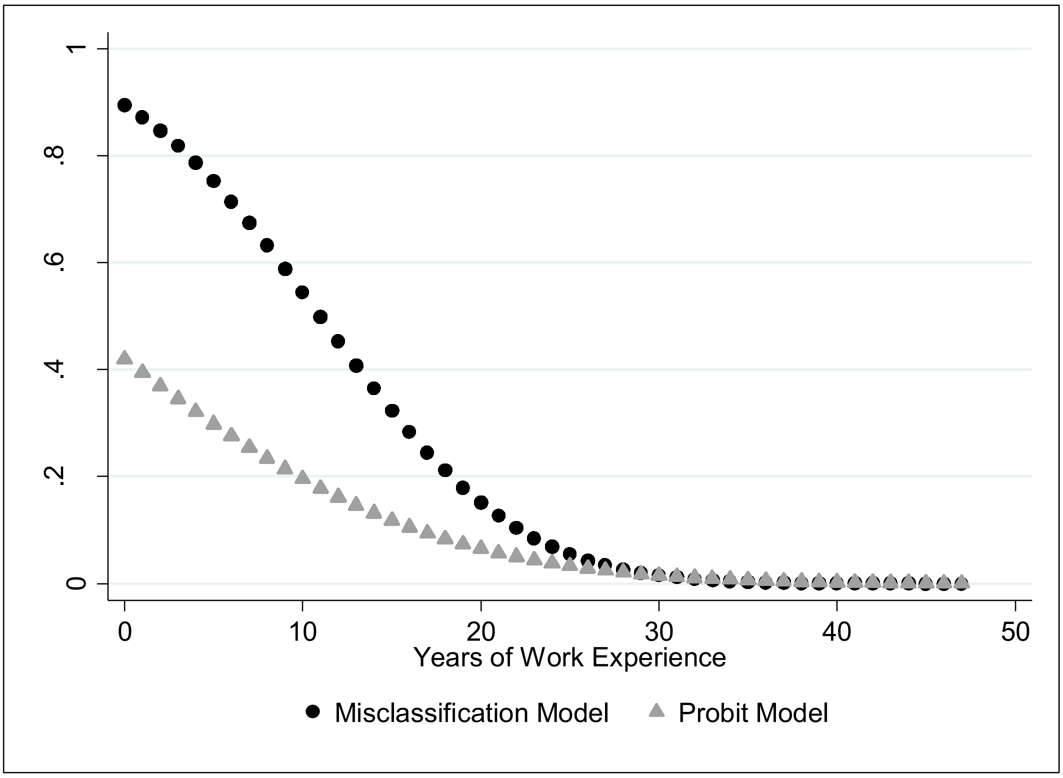
The job mobility models also contain variables that try to capture some job and firm characteristics. We would expect a positive relationship between working part-time and job mobility as part-time workers typically have a lower attachment to the labour force and firms may be less willing to invest in training for them etc. The results provide strong evidence of this relationship. However, women who work part-time are less likely to change jobs. A variable to capture overskilling is included as it may signify a poor job match. The results show a positive relationship between being overskilled and job mobility and the effect from the misclassification model of being overskilled is 3 times the impact from the probit model. A firm size effect is included to capture the fact that those working in a large firm may have more alternative employment opportunities within the firm and so are less likely to change jobs. The results indicate that workers in firms with more than 50 employees have a lower probability of changing jobs and, as before, the impact is more marked in the misclassification model.

Working in the public sector is found to exert a negative effect on the probability of changing jobs and the marginal impact of working in the public sector in the misclassification model is almost treble the impact than in the model without misclassification. The effect of the sector a worker was in the previous year (or for job changers the sector they previously worked in) is similar in both models but again the marginal effects are higher in the misclassification model. The results also show that workers in the construction and market services sector are more likely to change jobs than those in the nonmarket services sector.

A Dublin city dummy variable and regional unemployment rates are included to control for factors such as access to alternative jobs and local labour market conditions. Neither of these effects is significant in either model. We would expect the Dublin city effect to be positive reflecting the fact that living in Dublin city means a worker has more alternative employment opportunities. Also, we would expect the impact on the unemployment rate to be negative as a lower unemployment rate may signal to workers that jobs are more abundant and that job search is likely to result in them finding an alternative job. The impact of the regional unemployment rate is negative, as expected, but insignificant. It is likely that the unemployment rate variable is correlated with the time dummy variables. In fact, when the year dummies are dropped the effect of the unemployment rate is highly significant.

A useful way to demonstrate the differences between the two models is to graph the marginal effects of the variables. Figure 5.1 plots the marginal effect of experience from both models. The curves slope down as the probability of job change decreases as years of experience increases (i.e. the marginal effect on experience is negative). The slopes of the curves are steep at lower values of experience and then flatten out at higher years of experience indicating that an additional year of experience reduces the probability of changing jobs but at a declining rate (i.e. the marginal effect of years of experience squared is positive). Overall, the graph shows that the effect of ignoring misclassification error is large, especially at low values of experience.

Figure 5.1: Marginal Effect of Experience in Models of Job Mobility



5.5.2 Estimation Results: Covariate Dependent Misclassification

The results in Section 5.5.1 show that misclassification has a very big impact on the marginal effects of experience and working part-time.[53] Section 5.3.2 argued that there might be differences in measurement error in the data by experience, gender and working part-time. Table 5.8 reports the results for the misclassification model where the misclassification probabilities depend on experience.[54] Table 5.9 reports the results where the misclassification probabilities depend on gender. Even though gender is not an important determinant of job mobility, Section 5.3.2 argued that there could be gender differences in measurement error. Table 5.10 shows the results when misclassification depends on working part-time.[55]

[53] The effect of misclassification is largest on some of the year dummies although the marginal effects tend to be insignificant.
[54] A categorical experience variable is used in the model when we allow the misclassification probabilities to depend on experience.
[55] A series of models were run where misclassification was allowed to depend separately on each of the covariates but either none of the additional probabilities estimated were significant or the models did not converge. In addition, a model was run where the probability of misclassifying job changes was allowed to depend on all the covariates but the model failed to achieve convergence.

In the model that allows the misclassification probabilities to depend on experience, the estimate of the probability of misclassifying a non-job change remains insignificantly different from zero for each of the experience categories (see Table 5.8). However, there is some variation within experience groups in terms of misclassifying job changes. The probability of misclassifying a job change for a worker with less than 20 years of experience, δ_0, is 58 per cent. The additional effect for someone with more than 20 years experience is given by δ_1 and the estimate indicates that these workers are almost 22 percentage points less likely to be misclassified as not having changed jobs than someone with less than 20 years experience. Although this is the sign we would expect, the estimate is not significant at the 10 per cent level. In general, the marginal effects are similar to the model where misclassification is independent of the covariates.

When the misclassification probabilities depend on gender, the probability of misclassifying a non-job change is not statistically different from zero for men and women. The probability of misclassifying a job change is around 60 per cent for men and the additional effect of misclassifying a job change for women is small and not significant. Table 5.10 reports the results for when misclassification depends on working part-time. The table shows that the probability of misclassifying a non-job change is not significantly different from zero for part-time and full-time workers. The probability of misclassifying a job change for a full-time worker is 53 per cent. The additional effect for someone who works part-time is that they are 17 percentage points more likely to be misclassified as not having changed jobs than someone who works full-time. However, the estimate is not significant. It may be the case the variation in the data is not sufficient to accurately identify covariate dependent misclassification or the results presented in Tables 5.8 to 5.10 may indicate that misclassification is independent of the covariates.

Table 5.8: Model of Job Change where Misclassification Depends on Experience*

Variable	*Marginal Effect*	*Robust Standard Error*	*P>|Z|*
Misclassification Depends on:		**Experience**	
$\hat{\gamma}_0$ if experience < 20 years	0.0086	0.0076	0.26
$\hat{\gamma}_1$ if experience >= 20 years	-0.0049	0.0068	0.47
$\hat{\delta}_0$ if experience < 20 years	0.5807	0.0896	0.00
$\hat{\delta}_1$ if experience >= 20 years	-0.2177	0.1454	0.13
Experience	-0.0361	0.0250	0.00
Experience squared	0.0005	0.0005	0.00
Female	0.0114	0.1393	0.71
Child	-0.0035	0.0974	0.87
Living in a Couple	0.0134	0.1421	0.66
Female* Living in a Couple	-0.0657	0.1755	0.05
Education- medium	-0.0375	0.0962	0.07
Education- high	-0.0745	0.1871	0.03
(Ref: Education – low)			
Working Part-Time	0.1925	0.2271	0.00
Female* Working Part-Time	-0.0688	0.2195	0.09
Recent Training	0.1084	0.1638	0.01
Public Sector	-0.0630	0.1223	0.01
Number of Employees > 50	-0.0459	0.0911	0.02
Overskilled	0.0685	0.0769	0.00
Occupation of Origin:			
(Ref: Elementary Occ's)			
Manager	-0.1078	0.2245	0.00
Professional	-0.0912	0.1379	0.00
Clerk	-0.0837	0.1612	0.01
Skilled	-0.0847	0.1191	0.00
Sector of Origin:			
(Ref: Non Market Services)			
Agriculture, Mining & Utilities	-0.0826	0.1997	0.02
Manufacturing	-0.0558	0.1623	0.08
Building	0.1754	0.2296	0.01
Market Services	0.0546	0.1444	0.09
Year Dummies:			
(Ref: 1995)			
1996	0.0134	0.1338	0.65
1997	0.0095	0.1446	0.77
1998	0.0862	0.1879	0.06
1999	0.0528	0.1872	0.23
2000	0.0991	0.2238	0.08
2001	0.0541	0.2298	0.32
Dublin Region	-0.0397	0.1365	0.14
Regional Unemployment Rate	-0.0140	0.0383	0.23
N		8,736	
Wald chi2		82.11	
Prob > chi2		0.0000	
Log pseudolikelihood		-2407.1317	

* Notes: Standard errors are adjusted to take account of the fact that there are multiple observations on the same people. The marginal effects are evaluated at the sample means of the explanatory variables.

Table 5.9: Model of Job Change where Misclassification Depends on Gender*

Variable	*Marginal Effect*	*Robust Standard Error*	*P>|Z|*
Misclassification Depends on:		**Gender**	
$\hat{\gamma}_0$ if male	0.0052	0.0065	0.43
$\hat{\gamma}_1$ if female	0.0023	0.0078	0.77
$\hat{\delta}_0$ if male	0.6042	0.0880	0.00
$\hat{\delta}_1$ if female	-0.0018	0.0839	0.98
Experience	-0.0422	0.0272	0.00
Experience squared	0.0006	0.0005	0.00
Female	0.0112	0.2910	0.88
Child	-0.0039	0.1083	0.89
Living in a Couple	0.0137	0.1636	0.73
Female* Living in a Couple	-0.0869	0.2025	0.05
Education- medium	-0.0472	0.1032	0.06
Education- high	-0.0946	0.2002	0.02
(Ref: Education – low)			
Working Part-Time	0.2507	0.2751	0.00
Female* Working Part-Time	-0.0949	0.2743	0.09
Recent Training	0.1378	0.1749	0.01
Public Sector	-0.0798	0.1353	0.01
Number of Employees > 50	-0.0607	0.1002	0.01
Overskilled	0.0858	0.0858	0.00
Occupation of Origin:			
(Ref: Elementary Occ's)			
Manager	-0.1342	0.2510	0.00
Professional	-0.1166	0.1534	0.00
Clerk	-0.1112	0.1875	0.01
Skilled	-0.1061	0.1338	0.00
Sector of Origin:			
(Ref: Non Market Services)			
Agriculture, Mining & Utilities	-0.1059	0.2346	0.03
Manufacturing	-0.0672	0.1898	0.12
Building	0.2345	0.2706	0.01
Market Services	0.0749	0.1499	0.05
Year Dummies:			
(Ref: 1995)			
1996	0.0176	0.1388	0.62
1997	0.0123	0.1522	0.75
1998	0.1116	0.2151	0.07
1999	0.0705	0.2096	0.22
2000	0.1328	0.2555	0.07
2001	0.0734	0.2601	0.30
Dublin Region	-0.0445	0.1529	0.21
Regional Unemployment Rate	-0.0177	0.0416	0.26
N		8,736	
Wald chi2		69.84	
Prob > chi2		0.0001	
Log pseudolikelihood		-2410.0709	

* Notes: Standard errors are adjusted to take account of the fact that there are multiple observations on the same people. The marginal effects are evaluated at the sample means of the explanatory variables.

Table 5.10: Model of Job Change where Misclassification Depends on Working Part-Time*

Variable	*Marginal Effect*	*Robust Standard Error*	*P>|Z|*
Misclassification Depends on:		**Working Part-Time**	
$\hat{\gamma}_0$ if working full-time	0.0044	0.0060	0.46
$\hat{\gamma}_1$ if working part-time	0.0090	0.0146	0.54
$\hat{\delta}_0$ if working full-time	0.5254	0.1658	0.00
$\hat{\delta}_1$ if working part-time	0.1650	0.1363	0.23
Experience	-0.0391	0.0322	0.00
Experience squared	0.0006	0.0006	0.01
Female	0.0137	0.1383	0.67
Child	-0.0037	0.1025	0.88
Living in a Couple	0.0155	0.1465	0.65
Female* Living in a Couple	-0.0771	0.1830	0.04
Education- medium	-0.0454	0.1011	0.05
Education- high	-0.0799	0.2082	0.05
(Ref: Education – low)			
Working Part-Time	0.3590	0.3748	0.00
Female* Working Part-Time	-0.0915	0.2887	0.09
Recent Training	0.1191	0.1912	0.02
Public Sector	-0.0742	0.1339	0.01
Number of Employees > 50	-0.0525	0.1064	0.03
Overskilled	0.0760	0.0926	0.00
Occupation of Origin:			
(Ref: Elementary Occ's)			
Manager	-0.1164	0.2614	0.01
Professional	-0.1019	0.1524	0.00
Clerk	-0.0954	0.1923	0.02
Skilled	-0.0952	0.1259	0.00
Sector of Origin:			
(Ref: Non Market Services)			
Agriculture, Mining & Utilities	-0.0915	0.2215	0.03
Manufacturing	-0.0606	0.1751	0.10
Building	0.1865	0.2971	0.03
Market Services	0.0632	0.1565	0.10
Year Dummies:			
(Ref: 1995)			
1996	0.0091	0.1322	0.77
1997	0.0077	0.1429	0.82
1998	0.0832	0.2103	0.13
1999	0.0546	0.1938	0.26
2000	0.1017	0.2456	0.12
2001	0.0513	0.2428	0.40
Dublin Region	-0.0341	0.1417	0.27
Regional Unemployment Rate	-0.0196	0.0391	0.18
N		8,736	
Wald chi2		56.61	
Prob > chi2		0.0023	
Log pseudolikelihood		-2406.5278	

* Notes: Standard errors are adjusted to take account of the fact that there are multiple observations on the same people. The marginal effects are evaluated at the sample means of the explanatory variables.

5.6 Conclusions

This chapter investigates job mobility in Ireland over the period 1995 to 2001. It finds that there are substantial inconsistencies in responses to a question about tenure in the LIS. The extent of the measurement error is similar to what has been found in other studies. Survey data on tenure are very often used to deduce job changes and given the extent of response error evident in the data it is likely that cases are misclassified as job changes when they truly no job change has taken place and vice versa.

The decision to change jobs can be set in a binary choice framework. Misclassification in a binary dependent variable can lead to estimates that are biased and inconsistent so it is important to control for misclassification. An estimator developed by Hausman *et al.* is used to control for misclassification. The results indicate that, by ignoring misclassification, the true number of job changes is underestimated by around 60 per cent. The average mobility rate in the dataset is calculated at around 9.7 per cent and the estimate for α_1, the misclassification rate for job changes, implies that the true mobility rate is around 15.6 per cent.

In addition, the chapter finds that ignoring misclassification leads to diminished covariate effects. The chapter also examined the possibility of misclassification depending on some of the characteristics of workers. However, it does not find strong support for covariate dependent misclassification.

Appendix 5A: An Alternative Measure of Job Mobility

The LIS contains monthly activity reports where respondents are asked to describe their main activity in each month over the period that extends back to January of the previous year. This appendix describes an alternative way of capturing job changes using the responses to these monthly activity reports. We can define a job change to occur when someone is observed to be working in one month, where they were reported to be unemployed or out of the labour force in the previous month given that they were working in an earlier month. Appendix Table 5.1 shows the mapping between the two measures of job mobility for the total number of person-year observations (from Table 5.1).

The monthly activity reports imply that there are a total of 240 job changes observable over the period and this is consistent with an average job mobility rate of around 2.7 per cent over the period.[56] This seems to be an implausibly low mobility rate for a booming economy with a flexible labour market. It is also much lower than the number of job changes defined using reported start dates. In fact, the two measures only agree on 160 cases that a job change has taken place. The table shows that there are 80 cases where a job change has been defined using the monthly activity reports where one has not been defined using the reported job start dates, indicating that perhaps the definition using reported job start dates may undercount the true number of job changes. However, what is striking from the table is that the measure of job mobility using the monthly activity reports classifies 728 cases as being job stays while the measure based on reported start dates defines these cases to be job changes. However, it is important to note that if a worker was not unemployed or out of the labour force between jobs then this case would be classified as a job change using job start dates but not as one using the measure based on monthly activity reports. Also, if the amount of time between jobs was reasonably short a worker may ignore this when reporting their monthly activity status and so this type of case would not be classified as a job change using the monthly activity reports.

[56] The empirical analysis in the chapter uses annual data. The measure of job changes based on monthly activity reports defines a maximum of one job change per person per year (or between interviews). Of course, there are cases where using the monthly activity reports would indicate there is more than one job change within a year.

Appendix Table 5.1: Number of Workers and Job Changes

Job Changes (defined using job start dates)	Job Changes (defined using monthly activity reports)		Total
	Job Stay	Job Change	
Job Stay	8,008	80	8,088
Job Change	728	160	888
Total	8,736	240	8,976

Appendix Table 5.2: Explanatory Variables: Definitions and Summary Statistics

Variable	*Description*	*Mean*	*Std. Dev.*
Experience	Number of years in employment	19.5	11.5
Experience Squared	Number of years in employment squared	510.2	507.8
Female	Dummy variable that takes the value 1 if female and zero if male	0.35	0.48
Child	Dummy variable that takes the value 1 if the person has children and zero otherwise	0.56	0.50
Couple	Dummy variable that takes the value 1 if the person is married or living in a couple and zero otherwise	0.69	0.46
Female*Couple	Dummy variable that takes the value 1 if the person is female and married or living in a couple and zero otherwise	0.22	0.42
Education- low (Reference Category)	Dummy variable that takes the value 1 if highest educational qualification is Junior Certificate and zero otherwise	0.46	0.50
Education- medium	Dummy variable that takes the value 1 if highest educational attainment is above Junior Certificate but below degree level and zero otherwise	0.41	0.49
Education- high	Dummy variable that takes the value 1 if highest educational qualification is a degree or above and zero otherwise	0.13	0.34
Part-Time	Dummy variable that takes the value 1 if the person works less than 30 hours per week and zero otherwise	0.15	0.36
Female*Part-Time	Dummy variable that takes the value 1 if the person is female and works less than 30 hours per week and zero otherwise	0.11	0.31
Recent Training	Dummy variable that takes the value 1 if the person has been in education or training in the past year and zero otherwise	0.08	0.27
Public	Dummy variable that takes the value 1 if the person was working in the public sector in the previous year and zero otherwise	0.28	0.45
Number of Employees	Dummy variable that takes the value 1 if the number of employees in the firm in the previous year is more than 50 and zero otherwise.	0.35	0.48

Overskilled	Dummy variable that takes the value 1 if the worker reported that they felt they had skills and qualifications to do a more demanding job and zero otherwise.	0.48	0.50
Occupation of Origin:			
Manager	Dummy variable that takes the value 1 if occupation of origin is manager, senior official or legislator and zero otherwise	0.09	0.29
Professional	Dummy variable that takes the value 1 if occupation of origin is professional, technician or associated professionals and zero otherwise	0.25	0.43
Clerk	Dummy variable that takes the value 1 if occupation of origin is clerk, service, shop or sale worker and zero otherwise.	0.23	0.42
Skilled	Dummy variable that takes the value 1 if occupation of origin is skilled agricultural or fishery worker or a skilled craft or trades worker and zero otherwise.	0.22	0.41
Elementary (Reference Category)	Dummy variable that takes the value 1 if occupation in the previous year is plant or machine operator or assembler, or elementary occupation and zero otherwise.	0.21	0.40
Sector of Origin:			
Agriculture, Mining and Utilities	Dummy variable that takes the value 1 if sector of origin is agriculture, fishing, mining or quarrying, or utilities and zero otherwise.	0.13	0.34
Manufacturing	Dummy variable that takes the value 1 if sector of origin is manufacturing and zero otherwise.	0.19	0.39
Building	Dummy variable that takes the value 1 if sector of origin is building and zero otherwise.	0.08	0.27
Market Services	Dummy variable that takes the value 1 if sector of origin is distribution, hotels and restaurants, transport, storage and communications, financial intermediation, or real estate, renting and business activities and zero otherwise.	0.35	0.48
Non-Market Services (Reference Category)	Dummy variable that takes the value 1 if sector of origin is education, public administration and defence or health and social work and zero otherwise.	0.25	0.43

Year Dummies:			
1995 (Reference Category)	Dummy variable that takes the value 1 if the year is 1995 and zero otherwise.	0.13	0.34
1996	Dummy variable that takes the value 1 if the year is 1996 and zero otherwise.	0.13	0.34
1997	Dummy variable that takes the value 1 if the year is 1997 and zero otherwise.	0.13	0.34
1998	Dummy variable that takes the value 1 if the year is 1998 and zero otherwise.	0.14	0.35
1999	Dummy variable that takes the value 1 if the year is 1999 and zero otherwise.	0.15	0.36
2000	Dummy variable that takes the value 1 if the year is 2000 and zero otherwise.	0.15	0.36
2001	Dummy variable that takes the value 1 if the year is 2001 and zero otherwise.	0.16	0.37
Dublin	Dummy variable that take the value 1 if the household the person is living in is situated in Dublin city	0.11	0.32
Regional Unemployment Rate	Constructed from the NUTS3 regional data and labour force status available in the LIS	5.14	1.76

6. Measurement Error in Survey Data: Models of Job Mobility Across Europe

6.1 Introduction and Motivation

The previous chapter investigated job mobility in Ireland using the Living in Ireland Survey (LIS), the Irish component of the European Community Household Panel (ECHP). There is no direct question in the LIS about job mobility; instead it is inferred from responses of individuals to a question about tenure. One of the main findings of the chapter was that there is substantial measurement error in the recorded responses which may lead to misclassifying people who have not changed jobs as having changed jobs and vice versa.[57] The chapter also formally controlled for misclassification, using the estimator developed by Hausman, Abrevaya and Scott-Morton (1988) and discussed in detail in Chapter 4, and found that that the true rate of job mobility is being significantly undercounted in Ireland. In addition, the findings show that ignoring misclassification leads to diminished covariate effects in a probit model of job mobility. Given the serious impact misclassification can have, it is imperative to control for it in cases when misclassification is likely. This chapter extends the analysis and explores misclassification in job changes in twelve European countries over the period 1995 to 2001 using the ECHP.

The chapter finds that the true rates of job mobility are undercounted in several countries, typically in the peripheral countries of the EU. In addition, the chapter finds that similar factors are important in determining mobility across countries. Apart from age, personal and household characteristics are generally not important in explaining the probability of job mobility; occupation and sector of origin have some role, while firm and job characteristics have an important role in explaining the decision to change jobs. The effect of these variables on the likelihood of changing jobs is much stronger in the models that control for misclassification.

[57] The reasons for, and prevalence of, reporting errors in labour market survey data and in job mobility studies, in particular, were discussed in Sections 5.2 and 5.3.

The chapter is organised as follows: Section 6.2 describes the dataset. Section 6.3 investigates the extent of measurement error in the ECHP data. Section 6.4 provides estimation results and Section 6.5 concludes.

6.2 Dataset and Definition of Job Mobility

The European Community Household Panel (ECHP) is a harmonised, cross-national annual survey that collects information on several socio-economic aspects in the European Union. It is a longitudinal survey and the data collected on personal and job characteristics are very rich which makes it useful for analysing labour market dynamics. The survey asks when a person started working with their present employer and responses to this question are used to capture job changes. The survey started in 1994 and ended in 2001, however not all countries entered the survey at the same time; it began in 1995 in Austria, in 1996 in Finland and in 1997 in Sweden. In addition, for Germany, the UK and Luxembourg, the original sample has been replaced with harmonised versions of household panels that were already being used in those countries.

The panel dimension of the ECHP is exploited to identify job changes. Similar to the approach for Ireland using the LIS data in Chapters 3 and 5, a revolving balanced panel of people aged 20 to 60 years, roughly the prime working age, is selected for each country. This means that individuals are included in the sample in every year that they meet this age restriction.

As there is no explicit question in the ECHP about changing jobs, job changes are identified using the information about when a worker reports that they started working with their current employer. Job mobility is defined in terms of employment-to-employment transitions so to capture this workers need to be employed in two consecutive waves. Workers are asked to report the month and year that they started working with their current employer. If this date is after their interview in one year but before their interview in the following year, this indicates that the person has changed jobs between the two waves.[58]

[58] This is the same definition of job change as is used in Chapters 3 and 5.

The analysis proceeds with 12 out of the 15 countries in the ECHP, namely Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, The Netherlands, Portugal and Spain.[59] Sweden, Luxembourg and the UK are excluded from the analysis. Sweden is excluded because there is no panel element to the data, it is a series of cross sections. In this chapter, information on individuals over successive waves is used to identify job changes and also all variables are lagged in the models of job mobility (so they refer to a person's characteristics and situation in the previous year or in their previous job) and the Swedish data does not permit this. The data on Luxembourg only records when a person started working with their present employer from 1998 on and also data on the timing of interviews is missing. Finally, for the UK there appears to be significant coding error in the ECHP version of the British Household Panel Survey.[60]

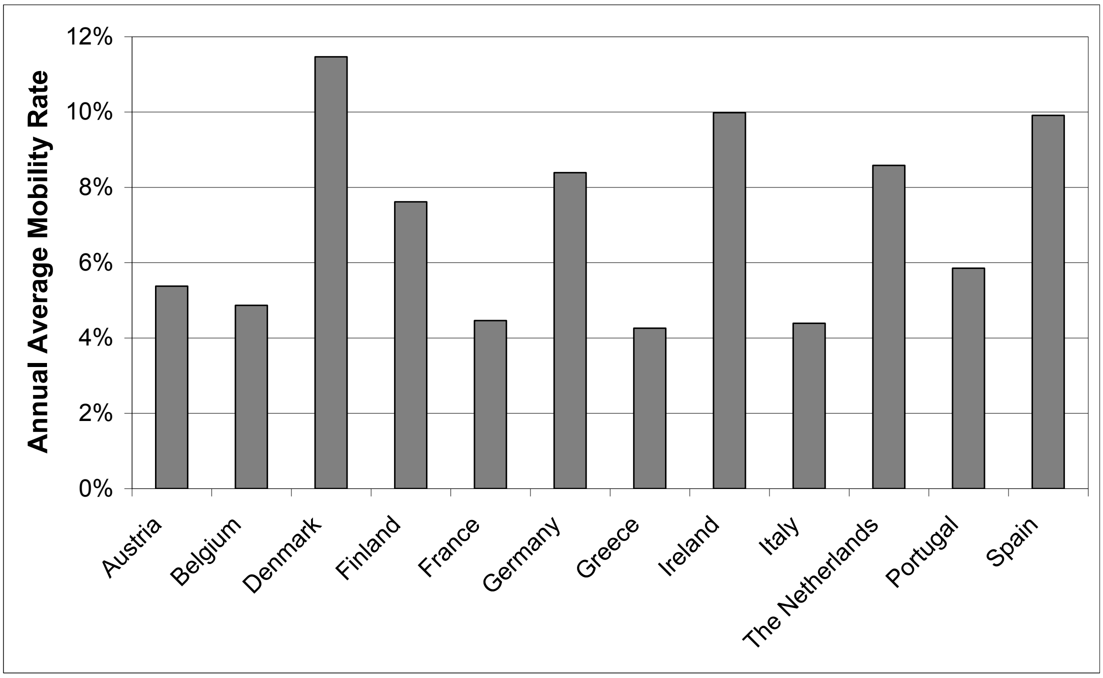
Appendix Table 6.1 shows the number of people in the revolving balanced panel, the number of workers employed in consecutive two-year periods and the rate of job mobility for each country considered in the analysis.[61] Figure 6.1 shows the annual average mobility rate for workers in each country. Across countries, the average mobility rate is around 7 per cent per annum. The figure also shows that countries with more flexible labour markets (e.g. in terms of employment protection legislation) such as Denmark and Ireland have relatively high mobility rates, while countries with more highly regulated labour markets such as France, Italy, Greece and Portugal have relatively low mobility rates.

[59] The ECHP version of the Living in Ireland Survey is used for Ireland in this chapter. The ECHP data for Ireland comes from the LIS but is coded differently by Eurostat to ensure some standardisation in variables across countries.

[60] The implied rates of job mobility in the UK for 1995-1998 and 2001 are, on average, over 20 per cent. Worryingly, the data for 1999 and 2000 seem entirely inconsistent with the other years for the UK as the mobility rate falls to less than 2 per cent. In the data for the 1999 and 2000 waves, there are no recorded reports that anyone started working in their jobs in 1999 or 2000 but there is a massive increase in the number of cases where this variable is missing.

[61] For Austria, the revolving balanced panel covers the period 1996 to 2001 and for Finland it covers the period 1997 to 2001 as the ECHP Survey only began in 1995 in Austria and in 1996 in Finland. Also, a revolving balanced panel over the period 1995 to 1999 is used for Belgium. There was a routing problem in the national questionnaire which resulted in huge amounts of missing data for some of the key variables used in this paper in later years of the survey.

Figure 6.1: Annual Average Mobility Rates derived from the ECHP



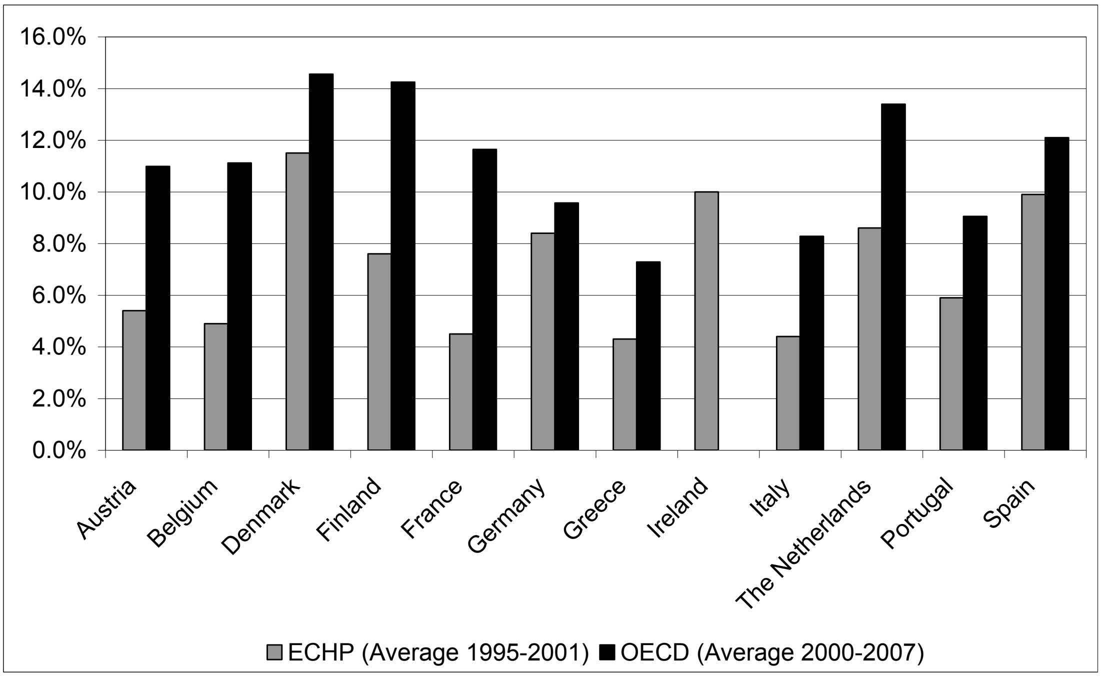
However, without exogenous job change information we cannot be certain that the reported number of job changes is correct. Workers may find it difficult to accurately recall when they started working with their current employer, they may misunderstand the question or their responses could be coded incorrectly.

As in Chapter 5, the concern is not necessarily that job start dates are misreported but rather that if they are misreported there is a risk that cases may be misclassified as job changes and vice versa. One way to assess whether the average mobility rates shown in Figure 6.1 are sensible is to compare them with estimates from other sources. Figure 6.2 compares the job mobility rates from the ECHP with some OECD (2010) cross country estimates. The OECD estimates refer to annual average job-to-job transition rates over the period 2000 to 2007 and the rates are adjusted for industry composition.

OECD (2010) and the related background paper Bassanini *et al.* (2010) use industry level data to construct a measure of job-to-job transitions which they define as the number of workers who are employed at both time *t* and t-*1* but who change employer between these two dates. Separations are calculated as the difference between hirings and employment changes between two years. Hirings rates at the industry level are computed from job tenure data available in European and national labour force surveys. Industry level employment growth rates are obtained from OECD STAN and EU KLEMS databases. Each industry separation rate is further divided into job-to-job and job-to-jobless transition rates. For each industry, the ratio of job-to-job and job-to-jobless transitions is given by the number of workers with an employer in an industry at time *t-1* to the average employment in the same industry in time *t-1* and *t*. Hirings and separations are re-scaled because of discrepancies between labour force surveys and national accounts. This arises because different waves of labour force surveys are hard to compare at a disaggregate industry level because the industry dimension is not taken into account in the labour force survey sampling design.

Although job mobility is calculated differently in OECD (2010) than in this chapter and it considers a later time period, there are some similarities across both measures. For example, both sources indicate that countries like Greece and Italy have the lowest mobility rates while countries like Denmark, Spain and the Netherlands have relatively higher rates of job change. However, what is probably most striking from the graph is the fact that the OECD estimates are above the ECHP ones; of course it has to be remembered the measures refer to two different time periods.

Figure 6.2: Comparison of Cross Country Mobility Rates from ECHP and OECD*



* Note: The OECD estimates come from Bassanini *et al.* (2010), Appendix Table A.1. No estimate is reported for Ireland.

Even though the OECD (2010) measure of job mobility is constructed differently from that employed in this chapter, both definitions involve a self reported measure of tenure which is subject to measurement error. Other studies use administrative data or have matched employee employer data that uses employer records to determine job changes. For example, Ibsen, Trevisan and Westergaard-Nielsen (2008) use registered data to compare the labour markets in Denmark and in Veneto, a large Italian region. They focus on workers aged 25 to 55 and over the period 1995 to 2001 the average mobility rate in their data is around 17 per cent in Denmark and around 10 per cent in Veneto.[62] These estimates compare with average calculated mobility rates for workers aged between 20 and 60 of 11.5 per cent and 4.4 per cent for Denmark and Italy respectively using the ECHP data.

6.3 Measurement Error in Responses

This section examines the consistency of responses given for job start dates over time in order to gauge how reliable the responses are. If workers accurately report the date they started working with their current employer and if this information were recorded without error, then all starting dates would be constant *within* jobs. One way to assess the reliability of the data on job starting dates is to partition the dataset into distinct jobs and to compare the starting dates across interviews.[63]

In order to assess the consistency of reported starting dates within jobs, the information on when the job began must be available. Across the countries considered in this chapter between 3 per cent (in Denmark) and almost 16 per cent (in Germany) of observations that are classified as working in the revolving balanced panels are missing the information on the year the person started working with their current employer in at least one of the waves. These people are excluded from the analysis in this section and Appendix Table 6.2 shows the number of cases that are dropped for each country. The data is converted into separate jobs in an analogous manner to Section 5.3.2.

[62] These average mobility rates for Denmark and Italy are derived from Tables 3.1, 3.2 and 3.3 in Ibsen *et al.* (2008).
[63] Given the likelihood of misclassifying cases as job changes and vice versa, we do not attempt to distinguish between voluntary and involuntary mobility.

Table 6.1 reports the total number of jobs observable for various durations for the 12 countries. We can only assess the consistency of reported job start dates if we observe the job surviving for longer than a year (or more precisely for longer than the amount of time between interviews) so that we will have at least two reported starting dates to compare. The table shows that for each country between 18 and 32 per cent of all jobs are only observed for one year. These jobs are excluded from the analysis in this section as there is only one starting date associated with each of these jobs. For example, all new jobs with a reported start date in 2001 are excluded from the analysis because the survey ended in 2001 so there is no other response to compare to that given in the 2001 survey. Other cases that are excluded from the analysis in this section include workers who start a new job in one wave but who are either unemployed or out of the labour force in the next wave. These excluded jobs are all recent jobs in the sense that they began (and some of them ended) over the period of the survey.

Table 6.1: Number of Jobs Observed over the Sample Period

	Total No. of Jobs	No. of Jobs Observable for one year	No. of Jobs Observable for one year as a % of Total
Austria	2,764	517	19%
Belgium	2,493	633	25%
Denmark	3,017	845	28%
Finland	3,140	819	26%
France	5,466	1,087	20%
Germany	6,789	1,762	26%
Greece	4,087	728	18%
Ireland	2,557	766	30%
Italy	5,579	1,042	19%
The Netherlands	3,864	857	22%
Portugal	5,762	1,197	21%
Spain	5,102	1,611	32%

The ECHP only records the starting month associated with a job if the year associated with the beginning of the job is, at the earliest, two years before the person joined the survey; for other observations only the starting year is recorded. Table 6.2 shows the number of jobs that are reported to be recent (i.e. that are reported to have begun at the earliest two years before the person joined the survey) and that are observed more than once over the survey period (so we have at least two starting dates to compare within jobs).[64] The table shows that between 16 and 35 per cent of all jobs across countries are classified as recent and are observed more than once over the survey period.

The table also reports how many of these recent jobs have inconsistencies in reported start dates in any of the years that the job is observed (i.e. that the reported started date is not the same across waves). The table shows that there are substantial inconsistencies in reported start dates within recent jobs; on average around 9 per cent of all recent jobs have inconsistencies in reported start dates. In addition, there are some marked differences across countries. For example, in France only 2 per cent of recent jobs that are observed at least twice have inconsistencies in reported start dates, whereas the comparable figure in Greece is over 18 per cent.

Table 6.2: Number of 'Recent' Jobs Observed over the Sample Period

	No. of 'Recent' Jobs that are Observed at least twice	No. of 'Recent' Jobs that are Observed at least twice as a % of Total	% of 'Recent' Jobs that have Inconsistencies in Starting Dates
Austria	731	26%	5.5%
Belgium	404	16%	5.2%
Denmark	1,064	35%	6.7%
Finland	730	23%	4.2%
France	1,336	24%	2.0%
Germany	1,817	27%	3.6%
Greece	1,227	30%	18.3%
Ireland[65]	830	32%	16.7%
Italy	1,402	25%	10.6%
The Netherlands	1,164	30%	7.2%
Portugal	1,650	29%	11.0%
Spain	1,512	30%	17.8%

Given the level of inconsistencies in reported start dates in recent jobs, we cannot be confident that all cases classified as job changes are truly job changes, and similarly that cases classified as continuing jobs are truly continuing jobs.[66]

[64] In addition, only jobs where the data on both the month and year that the job started is available are considered. In most countries there are a relatively small number of cases where the information on the month that a person reports to have started their job is missing and these jobs are excluded from the analysis.

[65] The inconsistency rate reported for recent jobs in the ECHP version of the Irish data appears to conflict with the measurement error reported in Chapter 5 for Ireland using the LIS data. This issue is explored in Appendix 6A.

[66] In the analysis, job changes are used to partition the data into distinct jobs. The analysis suggests that the measure of job change may not accurately identify the true number of job changes i.e. it is likely there are cases identified as job changes when truly no change in jobs took place and vice versa. This means we may over or underestimate the true number of jobs and therefore the level of inconsistencies in starting dates within jobs.

This analysis is limited in two ways. Firstly, the analysis only considers jobs that are recorded to be recent. There may be observations where tenure is overestimated, particularly in the earlier years of the survey, and these cases cannot be considered because the month that the person started working in their current job will not be recorded in the survey. Secondly, across countries on average around 24 per cent of jobs are only observed once and these jobs are excluded from the analysis.

6.4 Estimation Results

To consider this more formally, models of job change are estimated using both a standard probit estimator and the Hausman *et al.* (1998) estimator that controls for misclassification. Tables 6.3 to 6.14 show the marginal effects from both estimators for 12 European countries. The annual data for each country have been pooled and year dummies are included to control for factors that affect all workers but that vary over time. Various model specifications were tried for each of the countries; in some instances the models would not converge or misclassification probabilities greater than one in magnitude were estimated. The specifications reported in this chapter exclude workers in any year that they report to be working part-time (i.e. less than 30 hours a week).[67] Appendix Table 6.6 provides some basic descriptive statistics of the final samples used for each country.[68] The same explanatory variables are used in each country model.[69] With the exception of the year dummies, all the explanatory variables have been lagged so they refer to the worker's characteristics and situation in their previous job or in the previous year. Zero misclassification probabilities and estimates from ordinary probit models of job change are used as starting values in estimating the Hausman *et al.* model.

---

[67] In addition, separate models by gender were considered for each country but in many instances the models did not converge.

[68] The number of observations in each country model is lower than those reported in Appendix Table 6.1 as observations are excluded when data for at least one of the explanatory variables is missing or if the worker is reported to be working part-time in a given year.

[69] There are a few exceptions to this. Regional variables are not available for Denmark and the Netherlands so the national unemployment rate is included in those country models rather than the local unemployment rate. The variable indicating firm size is excluded from the models for France and Denmark, as it would reduce the final samples for those countries by 67 per cent and 28 per cent respectively. The overeducation and satisfaction variables are not available for Germany because either the relevant questions were not asked or the information is not available for other reasons.

Table 6.3: Marginal Effects from Models of Job Change for Austria*

	Standard Probit Model		Misclassification Model	
	Marginal Effect	***P>|Z|***	***Marginal Effect***	***P>|Z|***
$\hat{\alpha}_0$			0.0000	1.00
$\hat{\alpha}_1$			0.5559	0.23
Age	-0.0046	0.00	-0.0246	0.12
Age Squared	0.0000	0.10	0.0002	0.27
Female	-0.0060	0.39	-0.0144	0.43
Child	-0.0024	0.64	-0.0035	0.79
Female*Child	0.0121	0.18	0.0275	0.22
Living in a Couple	-0.0037	0.54	-0.0100	0.53
Female*Living in a Couple	-0.0072	0.37	-0.0153	0.46
Education: (Ref: Education – low (ISCED 0-2))				
Education – medium (ISCED 3)	0.0049	0.32	0.0115	0.35
Education – high (ISCED 5-7)	0.0074	0.52	0.0155	0.57
Recent Training	0.0028	0.47	0.0066	0.50
Satisfied with Number of Working Hours	-0.0022	0.69	-0.0046	0.73
Satisfied with Working Times	-0.0226	0.00	-0.0522	0.01
Satisfied with Working Conditions/Environment	-0.0252	0.00	-0.0631	0.04
Satisfied with Distance to Job/Commuting	0.0055	0.23	0.0135	0.29
Overeducated	0.0003	0.93	0.0019	0.84
Public	-0.0280	0.00	-0.0685	0.02
Number of Employees > 50	-0.0106	0.01	-0.0273	0.08
Occupation of Origin: (Ref: Elementary Occ's)				
Manager	-0.0145	0.09	-0.0359	0.14
Professional	-0.0098	0.39	-0.0247	0.38
Technicians	-0.0104	0.17	-0.0275	0.23
Clerks	-0.0170	0.01	-0.0434	0.09
Service	-0.0132	0.06	-0.0348	0.19
Skilled Ag. & Fishery	-0.0136	0.30	-0.0346	0.31
Craft & Trades	-0.0073	0.30	-0.0184	0.32
Plant & Machine Operators	-0.0083	0.32	-0.0186	0.37
Sector of Origin: (Ref: Non Market Services)				
Agriculture, Mining and Utilities	-0.0191	0.12	-0.0488	0.20
Manufacturing	0.0042	0.62	0.0077	0.71
Building	0.0106	0.30	0.0204	0.41
Market Services	0.0132	0.09	0.0280	0.12
Year Dummies: (Ref: 1996)				
1997	-0.0033	0.53	-0.0074	0.57
1998	-0.0054	0.31	-0.0135	0.34
1999	-0.0048	0.35	-0.0129	0.37
2000	-0.0065	0.42	-0.0172	0.43
2001	-0.0042	0.58	-0.0114	0.57
Region: (Ref: Ostösterreich)				
Südösterreich	-0.0054	0.30	-0.0160	0.37
Westösterreich	-0.0065	0.56	-0.0178	0.55
Local Unemployment Rate	0.0020	0.68	0.0096	0.68
N	9,925		9,925	
Wald chi2	391.13		40.76	
Prob > chi2	0.0000		0.3085	
Log pseudolikelihood	-1733.2982		-1732.9355	
Pseudo R^2	0.1226			

* Notes: Standard Errors are adjusted to take account of the fact that there are multiple observations on the same people. The marginal effects are evaluated at the sample means of the explanatory variables. The square root of α_0 and α_1 are estimated to ensure that the estimates are positive.

Table 6.4: Marginal Effects from Models of Job Change for Belgium*

	Standard Probit Model		Misclassification Model	
	Marginal Effect	***P>|Z|***	***Marginal Effect***	***P>|Z|***
$\hat{\alpha}_0$			0.0022	0.84
$\hat{\alpha}_1$			0.3462	0.88
Age	-0.0089	0.00	-0.0222	0.48
Age Squared	0.0001	0.07	0.0002	0.53
Female	-0.0072	0.56	-0.0115	0.69
Child	-0.0222	0.02	-0.0349	0.49
Female*Child	0.0437	0.01	0.0704	0.53
Living in a Couple	0.0113	0.26	0.0172	0.44
Female*Living in a Couple	-0.0218	0.13	-0.0352	0.57
Education: (Ref: Education –low (ISCED 0-2))				
Education – medium (ISCED 3)	0.0009	0.92	0.0020	0.90
Education – high (ISCED 5-7)	0.0003	0.98	0.0002	0.99
Recent Training	0.0145	0.03	0.0234	0.52
Satisfied with Number of Working Hours	-0.0101	0.20	-0.0156	0.39
Satisfied with Working Times	-0.0078	0.38	-0.0119	0.55
Satisfied with Working Conditions/Environment	-0.0141	0.08	-0.0219	0.39
Satisfied with Distance to Job/Commuting	-0.0162	0.06	-0.0261	0.54
Overeducated	0.0092	0.16	0.0144	0.38
Public	-0.0066	0.49	-0.0118	0.68
Number of Employees > 50	-0.0182	0.00	-0.0279	0.41
Occupation of Origin: (Ref: Elementary Occ's)				
Manager	-0.0053	0.73	-0.0083	0.75
Professional	-0.0104	0.45	-0.0160	0.56
Technicians	-0.0138	0.27	-0.0221	0.51
Clerks	-0.0131	0.25	-0.0208	0.50
Service	0.0161	0.27	0.0253	0.57
Skilled Ag. & Fishery	0.0015	0.96	0.0018	0.97
Craft & Trades	0.0022	0.86	0.0042	0.87
Plant & Machine Operators	-0.0159	0.20	-0.0241	0.41
Sector of Origin: (Ref: Non Market Services)				
Agriculture, Mining and Utilities	-0.0031	0.87	-0.0055	0.87
Manufacturing	0.0099	0.39	0.0151	0.63
Building	0.0289	0.11	0.0450	0.50
Market Services	0.0030	0.74	0.0051	0.78
Year Dummies: (Ref: 1995)				
1996	-0.0007	0.96	-0.0030	0.92
1997	0.0179	0.15	0.0265	0.38
1998	0.0018	0.90	0.0006	0.98
1999	-0.0093	0.67	-0.0168	0.74
Region: (Ref: Région Bruxelles-capitale/Brussels hoofdstad gewest)				
Vlaams Gewest	-0.0283	0.03	-0.0451	0.53
Région Wallonne	0.0182	0.47	0.0309	0.64
Local Unemployment Rate	-0.0114	0.14	-0.0300	0.54
N	4,686		4,686	
Wald chi2	196.89		8.40	
Prob > chi2	0.0000		1.000	
Log pseudolikelihood	-940.1600		-940.1238	
Pseudo R^2	0.0944			

* Notes: Standard Errors are adjusted to take account of the fact that there are multiple observations on the same people. The marginal effects are evaluated at the sample means of the explanatory variables.

Table 6.5: Marginal Effects from Models of Job Change for Denmark*

	Standard Probit Model		Misclassification Model	
	Marginal Effect	***P>|Z|***	***Marginal Effect***	***P>|Z|***
$\hat{\alpha}_0$			0.0098	0.88
$\hat{\alpha}_1$			0.4980	0.41
Age	-0.0057	0.06	-0.0209	0.34
Age Squared	0.0000	0.98	0.0000	0.90
Female	-0.0285	0.07	-0.0560	0.12
Child	-0.0050	0.64	-0.0081	0.73
Female*Child	0.0117	0.45	0.0228	0.52
Living in a Couple	-0.0111	0.38	-0.0238	0.48
Female*Living in a Couple	0.0175	0.34	0.0372	0.40
Education: (Ref: Education –low (ISCED 0-2))				
Education – medium (ISCED 3)	0.0083	0.46	0.0143	0.60
Education – high (ISCED 5-7)	0.0068	0.60	0.0134	0.72
Recent Training	0.0102	0.15	0.0225	0.58
Satisfied with Number of Working Hours	0.0046	0.62	0.0107	0.72
Satisfied with Working Times	-0.0102	0.32	-0.0232	0.58
Satisfied with Working Conditions/Environment	-0.0348	0.00	-0.0785	0.40
Satisfied with Distance to Job/Commuting	-0.0173	0.05	-0.0372	0.44
Overeducated	0.0169	0.01	0.0381	0.53
Public	-0.0176	0.11	-0.0361	0.54
Occupation of Origin: (Ref: Elementary Occ's)				
Manager	-0.0277	0.13	-0.0642	0.65
Professional	-0.0350	0.03	-0.0771	0.43
Technicians	-0.0191	0.23	-0.0467	0.64
Clerks	-0.0431	0.00	-0.0957	0.49
Service	-0.0094	0.58	-0.0280	0.71
Skilled Ag. & Fishery	0.0162	0.62	0.0446	0.79
Craft & Trades	-0.0032	0.86	-0.0076	0.86
Plant & Machine Operators	-0.0044	0.80	-0.0148	0.75
Sector of Origin: (Ref: Non Market Services)				
Agriculture, Mining and Utilities	-0.0333	0.10	-0.0741	0.67
Manufacturing	-0.0332	0.02	-0.0677	0.44
Building	-0.0205	0.25	-0.0397	0.59
Market Services	-0.0116	0.32	-0.0218	0.65
Year Dummies: (Ref: 1996)				
1997	0.0068	0.52	0.0150	0.67
1998	0.0551	0.00	0.1398	0.48
1999	-0.0241	0.05	-0.0501	0.34
2000	-0.0289	0.05	-0.0617	0.47
2001	-0.0044	0.78	-0.0083	0.82
National Unemployment Rate	-0.0062	0.09	-0.0188	0.46
N	8,106		8,106	
Wald chi2	391.68		14.22	
Prob > chi2	0.0000		0.9989	
Log pseudolikelihood	-2507.0117		-2506.7382	
Pseudo R^2	0.0864			

* Notes: Standard Errors are adjusted to take account of the fact that there are multiple observations on the same people. The marginal effects are evaluated at the sample means of the explanatory variables. A regional breakdown is not available in the ECHP for Denmark so the national unemployment rate is included in the model. Also, the firm size variable is not included in the model as it would reduce the sample size by 28 per cent.

Table 6.6: Marginal Effects from Models of Job Change for Finland*

	Standard Probit Model		Misclassification Model	
	Marginal Effect	***P>|Z|***	***Marginal Effect***	***P>|Z|***
$\hat{\alpha}_0$			0.0000	0.98
$\hat{\alpha}_1$			0.0000	0.95
Age	-0.0078	0.00	-0.0078	0.00
Age Squared	0.0001	0.13	0.0001	0.13
Female	0.0310	0.02	0.0310	0.02
Child	0.0096	0.30	0.0096	0.30
Female*Child	-0.0188	0.12	-0.0188	0.12
Living in a Couple	0.0109	0.29	0.0109	0.29
Female*Living in a Couple	-0.0261	0.08	-0.0261	0.08
Education: (Ref: Education –low (ISCED 0-2))				
Education – medium (ISCED 3)	-0.0063	0.51	-0.0063	0.51
Education – high (ISCED 5-7)	0.0038	0.73	0.0038	0.73
Recent Training	-0.0031	0.62	-0.0031	0.62
Satisfied with Number of Working Hours	-0.0059	0.42	-0.0059	0.42
Satisfied with Working Times	0.0052	0.51	0.0052	0.51
Satisfied with Working Conditions/Environment	-0.0136	0.08	-0.0136	0.08
Satisfied with Distance to Job/Commuting	-0.0289	0.00	-0.0289	0.00
Overeducated	0.0228	0.00	0.0228	0.00
Public	-0.0126	0.12	-0.0126	0.12
Number of Employees > 50	-0.0201	0.00	-0.0201	0.00
Occupation of Origin: (Ref: Elementary Occ's)				
Manager	-0.0140	0.38	-0.0140	0.38
Professional	0.0079	0.64	0.0079	0.64
Technicians	-0.0076	0.61	-0.0076	0.61
Clerks	-0.0157	0.30	-0.0157	0.30
Service	-0.0173	0.23	-0.0173	0.23
Skilled Ag. & Fishery	-0.0397	0.03	-0.0397	0.03
Craft & Trades	0.0032	0.85	0.0032	0.85
Plant & Machine Operators	-0.0043	0.80	-0.0043	0.80
Sector of Origin: (Ref: Non Market Services)				
Agriculture, Mining and Utilities	-0.0190	0.25	-0.0190	0.25
Manufacturing	-0.0301	0.01	-0.0301	0.01
Building	0.0099	0.56	0.0099	0.56
Market Services	-0.0075	0.43	-0.0075	0.43
Year Dummies: (Ref: 1997)				
1998	0.0314	0.00	0.0314	0.00
1999	-0.0090	0.66	-0.0090	0.66
2000	0.0142	0.60	0.0142	0.60
2001	0.0072	0.80	0.0072	0.80
Region: (Ref: Uusimaa)				
Etelä-Suomi (incl. Åland)	-0.0005	0.97	-0.0005	0.97
Itä-Suomi	0.0347	0.36	0.0347	0.36
Väli-Suomi	0.0246	0.32	0.0246	0.32
Pohjois-Suomi	0.0174	0.54	0.0174	0.54
Local Unemployment Rate	-0.0046	0.33	-0.0046	0.33
N	6,979		6,979	
Wald chi2	348.15		348.15	
Prob > chi2	0.0000		0.0000	
Log pseudolikelihood	-1679.567		-1679.567	
Pseudo R^2	0.1057			

* Notes: Standard Errors are adjusted to take account of the fact that there are multiple observations on the same people. The marginal effects are evaluated at the sample means of the explanatory variables. The square root of α_0 and α_1 are estimated to ensure that the estimates are positive.

Table 6.7: Marginal Effects from Models of Job Change for France*

	Standard Probit Model		Misclassification Model	
	Marginal Effect	*P>|Z|*	*Marginal Effect*	*P>|Z|*
$\hat{\alpha}_0$			0.0050	0.07
$\hat{\alpha}_1$			0.2607	0.74
Age	-0.0045	0.00	-0.0075	0.06
Age Squared	0.0000	0.02	0.0000	0.23
Female	-0.0011	0.81	-0.0024	0.66
Child	0.0028	0.38	0.0022	0.57
Female*Child	-0.0030	0.54	0.0000	1.00
Living in a Couple	-0.0002	0.97	0.0014	0.74
Female*Living in a Couple	-0.0037	0.50	-0.0083	0.26
Education: (Ref: Education –low (ISCED 0-2))				
Education – medium (ISCED 3)	0.0014	0.64	0.0014	0.70
Education – high (ISCED 5-7)	0.0001	0.97	0.0001	0.98
Recent Training	-0.0022	0.42	-0.0013	0.69
Satisfied with Number of Working Hours	-0.0041	0.17	-0.0042	0.27
Satisfied with Working Times	-0.0057	0.05	-0.0072	0.14
Satisfied with Working Conditions/Environment	-0.0089	0.00	-0.0106	0.03
Satisfied with Distance to Job/Commuting	-0.0064	0.03	-0.0072	0.13
Overeducated	0.0037	0.08	0.0040	0.16
Public	-0.0192	0.00	-0.0224	0.00
Occupation of Origin: (Ref: Elementary Occ's)				
Manager	0.0203	0.02	0.0215	0.08
Professional	0.0083	0.25	0.0111	0.27
Technicians	0.0074	0.21	0.0074	0.31
Clerks	0.0029	0.62	0.0038	0.58
Service	0.0061	0.34	0.0069	0.38
Skilled Ag. & Fishery	0.0122	0.36	0.0228	0.31
Craft & Trades	0.0014	0.81	0.0000	1.00
Plant & Machine Operators	0.0034	0.57	0.0029	0.68
Sector of Origin: (Ref: Non Market Services)				
Agriculture, Mining and Utilities	-0.0062	0.43	-0.0085	0.43
Manufacturing	0.0031	0.48	0.0057	0.35
Building	0.0213	0.00	0.0286	0.05
Market Services	0.0078	0.03	0.0109	0.09
Year Dummies: (Ref: 1995)				
1996	-0.0089	0.07	-0.0116	0.11
1997	-0.0115	0.02	-0.0137	0.06
1998	0.0067	0.17	0.0062	0.31
1999	0.0045	0.46	0.0036	0.61
2000	0.0141	0.06	0.0143	0.14
2001	0.0179	0.05	0.0181	0.17
Region: (Ref: Île de France)				
Bassin Parisien	-0.0081	0.03	-0.0083	0.07
Nord - Pas-de-Calais	-0.0031	0.68	-0.0043	0.62
Est	-0.0106	0.01	-0.0111	0.02
Ouest	-0.0081	0.02	-0.0083	0.05
Sud-Ouest	-0.0089	0.03	-0.0100	0.06
Centre-Est	-0.0120	0.00	-0.0129	0.03
Méditerranée	-0.0058	0.33	-0.0054	0.44
Local Unemployment Rate	-0.0011	0.40	-0.0025	0.33
N	17,673		17,673	
Wald chi2	519.76		27.58	
Prob > chi2	0.0000		0.9579	
Log pseudolikelihood	-2458.722		-2454.819	
Pseudo R^2	0.1323			

* Notes: Standard Errors are adjusted to take account of the fact that there are multiple observations on the same people. The marginal effects are evaluated at the sample means of the explanatory variables. The firm size variable is not included in the model as it would reduce the sample size by 67 per cent.

Table 6.8: Marginal Effects from Models of Job Change for Germany*

	Standard Probit Model		Misclassification Model	
	Marginal Effect	***P>|Z|***	***Marginal Effect***	***P>|Z|***
$\hat{\alpha}_0$			0.0000	1.00
$\hat{\alpha}_1$			0.7518	0.00
Age	-0.0005	0.74	-0.0127	0.17
Age Squared	0.0000	0.09	0.0000	0.69
Female	-0.0037	0.64	-0.0268	0.53
Child	-0.0089	0.11	-0.0268	0.30
Female*Child	0.0137	0.13	0.0510	0.22
Living in a Couple	-0.0039	0.57	-0.0464	0.24
Female*Living in a Couple	-0.0091	0.31	-0.0247	0.60
Education: (Ref: Education –low (ISCED 0-2))				
Education – medium (ISCED 3)	0.0106	0.07	0.0434	0.13
Education – high (ISCED 5-7)	0.0174	0.03	0.0585	0.11
Recent Training	0.0344	0.00	0.1627	0.00
Public	-0.0109	0.11	-0.0511	0.12
Number of Employees > 50	-0.0325	0.00	-0.1533	0.00
Occupation of Origin: (Ref: Elementary Occ's)				
Manager	-0.0067	0.53	-0.0060	0.91
Professional	-0.0124	0.21	-0.0245	0.61
Technicians	-0.0128	0.14	-0.0331	0.43
Clerks	-0.0150	0.10	-0.0476	0.28
Service	-0.0155	0.12	-0.0463	0.35
Skilled Ag. & Fishery	-0.0328	0.04	-0.1323	0.09
Craft & Trades	-0.0174	0.03	-0.0695	0.07
Plant & Machine Operators	-0.0208	0.02	-0.0901	0.03
Sector of Origin: (Ref: Non Market Services)				
Agriculture, Mining and Utilities	0.0030	0.81	0.0351	0.56
Manufacturing	-0.0058	0.49	-0.0072	0.86
Building	0.0199	0.06	0.1166	0.04
Market Services	0.0141	0.06	0.0924	0.03
Year Dummies: (Ref: 1995)				
1996	-0.0093	0.18	-0.0371	0.27
1997	-0.0039	0.60	-0.0140	0.71
1998	-0.0208	0.00	-0.0914	0.00
1999	-0.0171	0.01	-0.0813	0.01
2000	-0.0164	0.03	-0.0769	0.04
2001	-0.0157	0.05	-0.0676	0.08
Region: (Ref: Baden-Württemberg)				
Bayern	0.0158	0.05	0.0696	0.06
Berlin	0.0384	0.01	0.1685	0.01
Brandenburg	0.0116	0.39	0.0752	0.26
Bremen	0.0166	0.54	0.1359	0.32
Hamburg	-0.0073	0.73	-0.0064	0.95
Hessen	0.0152	0.14	0.0526	0.25
Mecklenburg-Vorpommern	0.0058	0.70	0.0345	0.64
Niedersachsen	0.0096	0.31	0.0256	0.54
Nordrhein-Westfalen	-0.0015	0.84	-0.0001	1.00
Sachsen	0.0040	0.72	0.0427	0.44
Sachsen-Anhalt	0.0240	0.08	0.1442	0.05
Schleswig-Holstein	0.0122	0.51	0.0574	0.46
Thüringen	0.0235	0.10	0.1171	0.08
Rheinland-Pfalz + Saarland	-0.0120	0.26	-0.0413	0.42
Local Unemployment Rate	0.0000	0.98	0.0002	0.98
N	21.394		21.394	
Wald chi2	636.79		125.42	
Prob > chi2	0.0000		0.0000	
Log pseudolikelihood	-5717.58		-5711.49	
Pseudo R^2	0.0614			

* Notes: Standard Errors are adjusted to take account of the fact that there are multiple observations on the same people. The marginal effects are evaluated at the sample means of the explanatory variables. The square root of α_0 and α_1 are estimated to ensure that the estimates are positive. The overeducation and satisfaction variables are not available for Germany as either the relevant questions were not asked or the information is not available for other reasons.

Table 6.9: Marginal Effects from Models of Job Change for Greece*

	Standard Probit Model		Misclassification Model	
	Marginal Effect	***P>|Z|***	***Marginal Effect***	***P>|Z|***
$\hat{\alpha}_0$			0.0047	0.14
$\hat{\alpha}_1$			0.8112	0.00
Age	-0.0017	0.15	-0.0186	0.19
Age Squared	0.0000	0.92	0.0000	0.99
Female	-0.0047	0.37	-0.0341	0.40
Child	-0.0008	0.85	-0.0171	0.60
Female*Child	-0.0139	0.02	-0.0795	0.05
Living in a Couple	-0.0050	0.35	-0.0250	0.49
Female*Living in a Couple	0.0130	0.09	0.0795	0.16
Education: (Ref: Education –low (ISCED 0-2))				
Education – medium (ISCED 3)	-0.0045	0.21	-0.0185	0.43
Education – high (ISCED 5-7)	0.0011	0.83	0.0237	0.52
Recent Training	-0.0087	0.09	-0.0560	0.12
Satisfied with Number of Working Hours	-0.0007	0.83	-0.0200	0.41
Satisfied with Working Times	-0.0008	0.82	-0.0030	0.89
Satisfied with Working Conditions/Environment	-0.0075	0.03	-0.0666	0.05
Satisfied with Distance to Job/Commuting	-0.0057	0.07	-0.0417	0.11
Overeducated	0.0072	0.01	0.0366	0.05
Public	-0.0090	0.11	-0.0576	0.16
Number of Employees > 50	-0.0039	0.38	-0.0345	0.29
Occupation of Origin: (Ref: Elementary Occ's)				
Manager	-0.0077	0.23	-0.0534	0.25
Professional	0.0059	0.49	0.0319	0.58
Technicians	0.0052	0.56	0.0297	0.62
Clerks	-0.0122	0.06	-0.0707	0.12
Service	-0.0029	0.67	-0.0044	0.93
Skilled Ag. & Fishery	-0.0256	0.00	-0.1620	0.03
Craft & Trades	-0.0092	0.12	-0.0619	0.18
Plant & Machine Operators	0.0000	1.00	-0.0068	0.88
Sector of Origin: (Ref: Non Market Services)				
Agriculture, Mining and Utilities	0.0267	0.03	0.1411	0.06
Manufacturing	0.0153	0.06	0.0718	0.17
Building	0.0224	0.02	0.1266	0.06
Market Services	0.0130	0.05	0.0646	0.13
Year Dummies: (Ref: 1995)				
1996	0.0191	0.01	0.1174	0.02
1997	0.0558	0.00	0.3791	0.00
1998	-0.0053	0.41	-0.0364	0.43
1999	0.0015	0.90	-0.0082	0.92
2000	-0.0109	0.28	-0.0693	0.33
2001	-0.0152	0.09	-0.1034	0.19
Region: (Ref: Voreia Ellada)				
Kentriki Ellada	-0.0081	0.02	-0.0535	0.04
Attiki	0.0053	0.18	0.0358	0.23
Nisia Aigaiou, Kriti	0.0041	0.40	0.0427	0.30
Local Unemployment Rate	0.0009	0.55	0.0073	0.69
N	16,050		16,050	
Wald chi2	579.7700		50.57	
Prob > chi2	0.0000		0.1016	
Log pseudolikelihood	-2561.0613		-2556.9619	
Pseudo R^2	0.1012			

* Notes: Standard Errors are adjusted to take account of the fact that there are multiple observations on the same people. The marginal effects are evaluated at the sample means of the explanatory variables. The square root of α_0 and α_1 are estimated to ensure that the estimates are positive.

Table 6.10: Marginal Effects from Models of Job Change for Ireland*

	Standard Probit Model		Misclassification Model	
	Marginal Effect	*P>|Z|*	*Marginal Effect*	*P>|Z|*
$\hat{\alpha}_0$			0.0047	0.48
$\hat{\alpha}_1$			0.5791	0.00
Age	-0.0086	0.00	-0.0469	0.08
Age Squared	0.0001	0.03	0.0004	0.12
Female	0.0147	0.35	0.0433	0.41
Child	0.0151	0.22	0.0254	0.54
Female*Child	-0.0178	0.28	-0.0359	0.49
Living in a Couple	-0.0128	0.38	-0.0242	0.57
Female*Living in a Couple	-0.0056	0.75	-0.0221	0.69
Education: (Ref: Education –low (ISCED 0-2))				
Education – medium (ISCED 3)	0.0024	0.80	0.0035	0.90
Education – high (ISCED 5-7)	0.0050	0.73	0.0123	0.76
Recent Training	0.0304	0.00	0.0927	0.04
Satisfied with Number of Working Hours	0.0018	0.85	0.0008	0.98
Satisfied with Working Times	-0.0352	0.00	-0.0938	0.01
Satisfied with Working Conditions/Environment	-0.0306	0.01	-0.0972	0.04
Satisfied with Distance to Job/Commuting	-0.0326	0.01	-0.1062	0.08
Overeducated	0.0135	0.06	0.0359	0.11
Public	-0.0485	0.00	-0.1267	0.01
Number of Employees > 50	-0.0202	0.02	-0.0590	0.07
Occupation of Origin: (Ref: Elementary Occ's)				
Manager	-0.0565	0.00	-0.1451	0.01
Professional	-0.0466	0.00	-0.1223	0.02
Technicians	-0.0416	0.00	-0.1078	0.01
Clerks	-0.0541	0.00	-0.1430	0.02
Service	-0.0352	0.01	-0.0993	0.06
Skilled Ag. & Fishery	-0.0429	0.01	-0.1054	0.06
Craft & Trades	-0.0426	0.00	-0.1052	0.00
Plant & Machine Operators	-0.0240	0.06	-0.0623	0.10
Sector of Origin: (Ref: Non Market Services)				
Agriculture, Mining and Utilities	-0.0483	0.01	-0.1252	0.05
Manufacturing	-0.0333	0.04	-0.0971	0.11
Building	0.0504	0.04	0.1624	0.11
Market Services	0.0101	0.51	0.0278	0.52
Year Dummies: (Ref: 1995)				
1996	-0.0101	0.62	-0.0192	0.73
1997	-0.0168	0.47	-0.0391	0.54
1998	0.0082	0.79	0.0565	0.55
1999	-0.0133	0.69	-0.0173	0.85
2000	-0.0050	0.90	0.0157	0.89
2001	-0.0148	0.75	-0.0068	0.96
Region: (Ref: Ireland, excluding Dublin)				
Dublin	-0.0135	0.24	-0.0366	0.32
Local Unemployment Rate	-0.0059	0.29	-0.0229	0.35
N	4,990		4,990	
Wald chi2	350.89		35.58	
Prob > chi2	0.0000		0.5354	
Log pseudolikelihood	-1277.068		-1274.975	
Pseudo R^2	0.1447			

* Notes: Standard Errors are adjusted to take account of the fact that there are multiple observations on the same people. The marginal effects are evaluated at the sample means of the explanatory variables.

Table 6.11: Marginal Effects from Models of Job Change for Italy*

	Standard Probit Model		Misclassification Model	
	Marginal Effect	***P>|Z|***	***Marginal Effect***	***P>|Z|***
$\hat{\alpha}_0$			0.0000	1.00
$\hat{\alpha}_1$			0.4532	0.69
Age	-0.0065	0.00	-0.0230	0.22
Age Squared	0.0001	0.00	0.0002	0.26
Female	0.0045	0.39	0.0085	0.45
Child	-0.0017	0.69	-0.0035	0.70
Female*Child	0.0077	0.31	0.0149	0.39
Living in a Couple	-0.0011	0.83	-0.0016	0.88
Female*Living in a Couple	-0.0081	0.24	-0.0156	0.34
Education: (Ref: Education –low (ISCED 0-2))				
Education – medium (ISCED 3)	-0.0039	0.25	-0.0088	0.55
Education – high (ISCED 5-7)	0.0104	0.12	0.0173	0.17
Recent Training	0.0044	0.29	0.0076	0.34
Satisfied with Number of Working Hours	-0.0005	0.88	-0.0005	0.94
Satisfied with Working Times	-0.0006	0.86	-0.0019	0.82
Satisfied with Working Conditions/Environment	-0.0053	0.09	-0.0099	0.20
Satisfied with Distance to Job/Commuting	-0.0080	0.01	-0.0151	0.14
Overeducated	0.0100	0.00	0.0188	0.08
Public	0.0034	0.55	0.0060	0.56
Number of Employees > 50	-0.0057	0.10	-0.0107	0.21
Occupation of Origin: (Ref: Elementary Occ's)				
Manager	-0.0256	0.00	-0.0479	0.06
Professional	-0.0241	0.00	-0.0448	0.04
Technicians	-0.0112	0.03	-0.0211	0.14
Clerks	-0.0242	0.00	-0.0456	0.06
Service	-0.0133	0.01	-0.0255	0.17
Skilled Ag. & Fishery	-0.0208	0.00	-0.0402	0.18
Craft & Trades	-0.0155	0.00	-0.0299	0.17
Plant & Machine Operators	-0.0103	0.05	-0.0202	0.26
Sector of Origin: (Ref: Non Market Services)				
Agriculture, Mining and Utilities	0.0347	0.00	0.0649	0.12
Manufacturing	0.0105	0.13	0.0205	0.31
Building	0.0404	0.00	0.0762	0.15
Market Services	0.0156	0.01	0.0287	0.09
Year Dummies: (Ref: 1995)				
1996	0.0108	0.05	0.0191	0.09
1997	0.0311	0.00	0.0583	0.09
1998	0.0136	0.02	0.0247	0.08
1999	0.0074	0.24	0.0136	0.30
2000	0.0120	0.05	0.0222	0.15
2001	0.0158	0.03	0.0290	0.11
Region: (Ref: Nord Ovest)				
Lombardia	0.0060	0.41	0.0114	0.46
Nord Est	0.0094	0.23	0.0181	0.33
Emilia-Romagna	0.0106	0.24	0.0198	0.31
Centro	-0.0028	0.66	-0.0060	0.67
Lazio	-0.0016	0.87	-0.0025	0.90
Abruzzo-Molise	-0.0193	0.02	-0.0366	0.15
Campania	-0.0177	0.27	-0.0331	0.33
Sud	-0.0204	0.10	-0.0384	0.19
Sicilia	-0.0156	0.34	-0.0296	0.39
Sardegna	-0.0006	0.98	-0.0296	0.98
Local Unemployment Rate	0.0008	0.42	0.0027	0.45
N	21,497		21,497	
Wald chi2	500.63		34.05	
Prob > chi2	0.0000		0.9036	
Log pseudolikelihood	-3689.2749		-3689.1539	
Pseudo R^2	0.0830			

* Notes: Standard Errors are adjusted to take account of the fact that there are multiple observations on the same people. The marginal effects are evaluated at the sample means of the explanatory variables. The square root of α_0 and α_1 are estimated to ensure that the estimates are positive.

Table 6.12: Marginal Effects from Models of Job Change for The Netherlands*

	Standard Probit Model		Misclassification Model	
	Marginal Effect	***P>|Z|***	***Marginal Effect***	***P>|Z|***
$\hat{\alpha}_0$			0.0022	0.85
$\hat{\alpha}_1$			0.0000	0.99
Age	-0.0049	0.09	-0.0046	0.14
Age Squared	0.0000	0.93	0.0000	0.99
Female	0.0069	0.61	0.0068	0.61
Child	0.0020	0.82	0.0019	0.83
Female*Child	-0.0123	0.43	-0.0118	0.46
Living in a Couple	-0.0036	0.76	-0.0035	0.77
Female*Living in a Couple	-0.0013	0.94	-0.0017	0.92
Education: (Ref: Education –low (ISCED 0-2))				
Education – medium (ISCED 3)	-0.0068	0.50	-0.0071	0.50
Education – high (ISCED 5-7)	-0.0046	0.68	-0.0050	0.67
Recent Training	0.0192	0.03	0.0190	0.03
Satisfied with Number of Working Hours	-0.0032	0.70	-0.0031	0.70
Satisfied with Working Times	-0.0104	0.26	-0.0104	0.27
Satisfied with Working Conditions/Environment	-0.0508	0.00	-0.0508	0.00
Satisfied with Distance to Job/Commuting	-0.0170	0.05	-0.0167	0.05
Overeducated	0.0356	0.00	0.0353	0.00
Public	-0.0013	0.90	-0.0009	0.93
Number of Employees > 50	-0.0240	0.00	-0.0239	0.00
Occupation of Origin: (Ref: Elementary Occ's)				
Manager	-0.0004	0.99	-0.0005	0.98
Professional	0.0197	0.31	0.0198	0.31
Technicians	0.0132	0.48	0.0132	0.48
Clerks	0.0022	0.91	0.0020	0.92
Service	0.0191	0.38	0.0186	0.39
Skilled Ag. & Fishery	0.0153	0.61	0.0153	0.61
Craft & Trades	-0.0061	0.75	-0.0063	0.74
Plant & Machine Operators	0.0147	0.49	0.0141	0.51
Sector of Origin: (Ref: Non Market Services)				
Agriculture, Mining and Utilities	-0.0015	0.94	-0.0007	0.97
Manufacturing	-0.0034	0.79	-0.0027	0.84
Building	0.0237	0.15	0.0239	0.16
Market Services	0.0239	0.02	0.0242	0.04
Year Dummies: (Ref: 1997)				
1998	0.0321	0.00	0.0324	0.01
1999	0.0251	0.12	0.0252	0.13
2000	0.0561	0.01	0.0565	0.01
2001	0.0542	0.03	0.0547	0.04
National Unemployment Rate	0.0069	0.36	0.0071	0.36
N	9,251		9,251	
Wald chi2	379.81		127.16	
Prob > chi2	0.0000		0.000	
Log pseudolikelihood	-2691.562		-2691.535	
Pseudo R^2	0.0812			

* Notes: Standard Errors are adjusted to take account of the fact that there are multiple observations on the same people. The marginal effects are evaluated at the sample means of the explanatory variables. A regional breakdown is not available in the ECHP for The Netherlands so the national unemployment rate is included in the model. The square root of α_0 and α_1 are estimated to ensure that the estimates are positive.

Table 6.13: Marginal Effects from Models of Job Change for Portugal*

	Standard Probit Model		Misclassification Model	
	Marginal Effect	***P>|Z|***	***Marginal Effect***	***P>|Z|***
$\hat{\alpha}_0$			0.0054	0.30
$\hat{\alpha}_1$			0.7413	0.00
Age	-0.0050	0.00	-0.0419	0.00
Age Squared	0.0000	0.06	0.0003	0.03
Female	-0.0005	0.94	0.0016	0.96
Child	-0.0036	0.43	-0.0150	0.51
Female*Child	0.0077	0.31	0.0232	0.50
Living in a Couple	0.0082	0.11	0.0317	0.19
Female*Living in a Couple	-0.0147	0.05	-0.0637	0.09
Education: (Ref: Education –low (ISCED 0-2))				
Education – medium (ISCED 3)	-0.0090	0.08	-0.0491	0.09
Education – high (ISCED 5-7)	0.0030	0.77	0.0116	0.79
Recent Training	0.0205	0.00	0.0777	0.01
Satisfied with Number of Working Hours	-0.0019	0.66	-0.0152	0.47
Satisfied with Working Times	0.0010	0.84	0.0078	0.73
Satisfied with Working Conditions/Environment	0.0034	0.46	0.0201	0.37
Satisfied with Distance to Job/Commuting	-0.0111	0.01	-0.0553	0.02
Overeducated	0.0052	0.09	0.0268	0.10
Public	-0.0153	0.03	-0.0629	0.04
Number of Employees > 50	-0.0157	0.00	-0.0730	0.00
Occupation of Origin: (Ref: Elementary Occ's)				
Manager	-0.0132	0.06	-0.0681	0.11
Professional	-0.0066	0.53	-0.0422	0.36
Technicians	-0.0144	0.05	-0.0694	0.06
Clerks	-0.0219	0.00	-0.1005	0.00
Service	-0.0120	0.03	-0.0621	0.05
Skilled Ag. & Fishery	-0.0153	0.06	-0.0735	0.06
Craft & Trades	-0.0151	0.00	-0.0825	0.02
Plant & Machine Operators	-0.0057	0.36	-0.0404	0.22
Sector of Origin: (Ref: Non Market Services)				
Agriculture, Mining and Utilities	0.0171	0.13	0.0796	0.14
Manufacturing	0.0050	0.56	0.0392	0.37
Building	0.0280	0.01	0.1584	0.02
Market Services	0.0119	0.11	0.0602	0.10
Year Dummies: (Ref: 1995)				
1996	0.0223	0.00	0.1009	0.01
1997	0.0253	0.00	0.1125	0.01
1998	0.0203	0.00	0.0857	0.01
1999	0.0222	0.01	0.0987	0.02
2000	0.0422	0.00	0.1919	0.00
2001	0.0337	0.00	0.1442	0.00
Region: (Ref: Norte)				
Centro	0.0081	0.10	0.0367	0.15
Lisboa e Vale do Tejo	-0.0110	0.12	-0.0452	0.19
Alentejo	0.0162	0.13	0.1046	0.10
Algarve	0.0070	0.28	0.0331	0.31
Açores	-0.0255	0.00	-0.1141	0.00
Madeira	-0.0344	0.00	-0.1517	0.01
Local Unemployment Rate	0.0039	0.03	0.0248	0.06
N	21,105		21,105	
Wald chi2	618.19		70.88	
Prob > chi2	0.0000		0.0035	
Log pseudolikelihood	-4456.2236		-4450.5016	
Pseudo R^2	0.0868			

* Notes: Standard Errors are adjusted to take account of the fact that there are multiple observations on the same people. The marginal effects are evaluated at the sample means of the explanatory variables.

Table 6.14: Marginal Effects from Models of Job Change for Spain*

	Standard Probit Model		Misclassification Model	
	Marginal Effect	***P>|Z|***	***Marginal Effect***	***P>|Z|***
$\hat{\alpha}_0$			0.0000	1.00
$\hat{\alpha}_1$			0.5816	0.00
Age	-0.0040	0.05	-0.0276	0.01
Age Squared	0.0000	0.90	0.0001	0.26
Female	-0.0010	0.91	-0.0138	0.61
Child	-0.0135	0.04	-0.0289	0.14
Female*Child	0.0349	0.01	0.0806	0.02
Living in a Couple	-0.0085	0.29	-0.0345	0.19
Female*Living in a Couple	-0.0144	0.21	-0.0246	0.48
Education: (Ref: Education –low (ISCED 0-2))				
Education – medium (ISCED 3)	-0.0225	0.00	-0.0642	0.00
Education – high (ISCED 5-7)	-0.0260	0.00	-0.0795	0.00
Recent Training	0.0138	0.02	0.0314	0.06
Satisfied with Number of Working Hours	-0.0022	0.67	-0.0060	0.68
Satisfied with Working Times	0.0033	0.53	0.0117	0.43
Satisfied with Working Conditions/Environment	-0.0081	0.12	-0.0245	0.11
Satisfied with Distance to Job/Commuting	-0.0266	0.00	-0.0798	0.00
Overeducated	0.0141	0.00	0.0422	0.00
Public	-0.0084	0.43	-0.0247	0.39
Number of Employees > 50	-0.0229	0.00	-0.0692	0.00
Occupation of Origin: (Ref: Elementary Occ's)				
Manager	-0.0636	0.00	-0.1773	0.00
Professional	-0.0452	0.00	-0.1340	0.00
Technicians	-0.0449	0.00	-0.1367	0.00
Clerks	-0.0552	0.00	-0.1601	0.00
Service	-0.0341	0.00	-0.1162	0.00
Skilled Ag. & Fishery	-0.0674	0.00	-0.1849	0.00
Craft & Trades	-0.0287	0.00	-0.0984	0.00
Plant & Machine Operators	-0.0208	0.01	-0.0774	0.01
Sector of Origin: (Ref: Non Market Services)				
Agriculture, Mining and Utilities	0.0648	0.00	0.1241	0.01
Manufacturing	0.0092	0.47	0.0113	0.74
Building	0.0739	0.00	0.2215	0.00
Market Services	0.0115	0.30	0.0171	0.56
Year Dummies: (Ref: 1995)				
1996	-0.0011	0.90	-0.0030	0.91
1997	-0.0047	0.59	-0.0129	0.60
1998	0.0295	0.01	0.0785	0.01
1999	0.0355	0.02	0.0905	0.03
2000	0.0413	0.05	0.1050	0.07
2001	0.0600	0.02	0.1531	0.03
Region: (Ref: Noroeste)				
Noreste	0.0124	0.30	0.0244	0.44
Comunidad de Madrid	0.0212	0.09	0.0547	0.11
Centro	0.0070	0.55	0.0144	0.66
Este	0.0251	0.03	0.0629	0.04
Sur	0.0223	0.29	0.0674	0.25
Canarias	0.0436	0.01	0.1061	0.02
Local Unemployment Rate	0.0018	0.33	0.0057	0.44
N	16,337		16,337	
Wald chi2	993.26		193.30	
Prob > chi2	0.0000		0.0000	
Log pseudolikelihood	-4665.5915		-4655.6772	
Pseudo R^2	0.1288			

* Notes: Standard Errors are adjusted to take account of the fact that there are multiple observations on the same people. The marginal effects are evaluated at the sample means of the explanatory variables. The square root of α_0 and α_1 are estimated to ensure that the estimates are positive.

Tables 6.3 to 6.14 provide estimates of the misclassification probabilities for each country. The estimated probability of misclassification for job stays, α_0, is very similar across countries and for all countries the estimate is less than 1 per cent. In contrast to the estimates of α_0, the estimated probability of misclassification for job changes, α_1, is far greater and is highly significant in several countries.[70] Significance tests on α_0 and α_1 can be used as tests of misclassification. For Germany, Greece, Ireland, Portugal and Spain the Hausman *et al.* estimator indicates significant misclassification. The range of estimates for α_1, where it is significant, is from 58 per cent in Ireland and Spain to 81 per cent in Greece. The estimate of α_1 for Ireland in the previous chapter using the LIS data was 61 per cent (see Table 5.6) which is very close to the estimate using the ECHP version of the LIS data. Although there is a substantial range in the estimates of α_1, the results indicate that for several countries the true number of job changes is being dramatically undercounted.

To put these estimates of α_1 in context, Table 6.15 shows the average (calculated) mobility rates for each country that were used in estimating the individual country models and the implied true mobility rates.

Table 6.15: True Mobility Rates Derived from Estimates of α_1

	Average Mobility Rates	Implied True Mobility Rate
Austria	5.0%	same
Belgium	5.8%	same
Denmark	10.6%	same
Finland	7.6%	same
France	3.8%	same
Germany	8.2%	14.4%
Greece	4.3%	7.8%
Ireland	8.9%	14.0%
Italy	4.6%	same
The Netherlands	9.6%	same
Portugal	6.2%	10.7%
Spain	10.1%	16.0%

* Note: Estimates for Austria, Belgium, Denmark, Finland, France, Italy and The Netherlands are not significant so we reject the model that controls for misclassification for these countries.

Table 6.16 ranks each of the countries from those with the lowest to the highest inconsistency rates observed in the data (taken from Table 6.2). From the table,

[70] The estimates of α_1 are not significant for Austria, Belgium, Denmark, Finland, France, The Netherlands and Italy.

countries such as Greece, Ireland, Portugal and Spain display the highest level of inconsistencies in reported start dates for recent jobs and these countries also have high and significant estimates of α_1. The table also shows that the estimates of α_1 for countries with lower inconsistency rates observed in the data such as France, Finland, Belgium and Austria are not significant.[71] However for Germany, the level of recorded inconsistencies within recent jobs is low and yet the estimate of α_1 is large and highly significant. This is surprising given the descriptive analysis in Section 6.3. For Germany some of the explanatory variables, such as overeducation, that are used in the country models and that are generally important in explaining job changes are missing. Also, the effect of age on job mobility is small and insignificant in the model for Germany in contrast to the models for the other countries. As shown in Section 4.3.2, the misclassification probabilities are identified off cases with strong negative and strong positive values of the $x_i'\beta$ distribution and it may be the case that the true index does not reach sufficiently large values to capture the rate of misclassification in the data.[72]

Table 6.16: Inconsistencies in Job Start Dates and Estimates of α_1 ^

	% of 'Recent' Jobs that have Inconsistencies in Starting Dates	$\hat{\alpha}_1$ (Misclassification Rate for Job Changes)
France	2.00%	Not significant
Germany	3.60%	0.752***
Finland	4.20%	Not significant
Belgium	5.20%	Not significant
Austria	5.50%	Not significant
Denmark	6.70%	Not significant
The Netherlands	7.20%	Not significant
Italy	10.60%	Not significant
Portugal	11.00%	0.741***
Ireland	16.70%	0.579***
Spain	17.80%	0.582***
Greece	18.30%	0.811***

^ Note: In the table *** denotes significant at the 1% level

[71] The results for both the probit and misclassification models for Finland are identical, as the estimates of the misclassification probabilities are zero. It seems likely that this is a case of lack of identification of the misclassification probabilities. The analysis of the inconsistencies in reported job starting dates for Finland revealed that the rate of inconsistencies were relatively low so there may not be a sufficient number of misclassified cases with strong negative and strong positive values of the $x_i'\beta$ distribution to identify the misclassification rates.

[72] A range of alternative models were run for Germany. A series of models were run that allowed the misclassification probabilities to depend on different age groups and also on gender. In addition, separate models were run for men and women and for younger and older workers. In each case, either the models failed to achieve convergence or the results were very similar to the estimated misclassification rates reported in Table 6.8.

For the countries where misclassification is been found to be significant, comparing the estimates from the probit and Hausman *et al.* models shows the impact of misclassification on the marginal effects of each variable. The results indicate that ignoring misclassification leads to diminished covariate effects. The estimates from the probit model have much lower standard errors, implying that not only are the probit estimates biased, but also that the precision of the estimates is overstated.

The results show that the same factors tend to be important in explaining job mobility, although there are some differences across countries. In the theoretical and empirical literature on job mobility, age is an important determinant of job change. Younger workers are more likely to change jobs as they have less knowledge of the labour market and their own preferences and abilities for different jobs. As age and tenure are correlated, they are also more likely to have lower tenure. Older workers with long tenures are more likely to have acquired job specific human capital, which they may be rewarded for, and so younger workers may have less to lose in monetary terms by changing jobs. The results show a strong negative relationship between age and the probability of job change across countries. When we control for misclassification, this negative relationship is more marked. For example, in the probit model of job change for Portugal, the marginal effect of age tells us that each additional year reduces the probability of changing jobs by 0.5 percentage points. However, the results from the model that controls for misclassification indicate that an additional year reduces the probability of changing jobs by 4.2 percentage points. The square of age is also included in the country models to capture the fact that age may have a non-linear effect on the probability of job mobility. The positive estimates of the squared term imply that as age increases its negative effect on the predicted probability of changing jobs diminishes.

The models of job mobility have a range of individual controls that include household structure and personal characteristics. We may expect women to be more likely to change jobs as they have a weaker attachment to the labour force but the results do not indicate any significant gender difference in the probability of changing jobs. The results indicate that Finland is only country where there are gender differences in the probability of changing jobs. Workers may be less likely to change jobs if they are more constrained by non-market variables, such as having children or living in a couple, and we would expect these effects to be stronger for women. However, the evidence for these effects is weak. The marginal effect of having children in the household is small and insignificant (apart from in Belgium), implying that the presence of children in the household does not affect the probability of changing jobs. A gender and children interaction term is included to capture the possible differential effect of children on mobility. The effect is not significant across countries with the exceptions of Belgium, Spain and Greece. In Belgium and Spain, women with children are more likely to change jobs while in Greece the effect is negative. The results also indicate that being married, or living in a couple does not affect the likelihood of changing jobs. Overall, household structure does not play an important role in explaining job mobility.

The relationship between general human capital which is proxied by the education level is weak across countries, although a negative relationship is found in Spain and a positive one in Germany. For both countries, the impact is more marked in models that control for misclassification although for Germany, the Hausman *et al.* estimates are not significant. The occupation variables may also capture human capital and we expect a negative relationship between higher occupational attainment and job mobility, although this result does not hold across all countries and occupations. Workers who have undergone recent training are more likely to change jobs in many countries apart from Greece, Finland and France but the effects for these latter two countries are not significant.

The job mobility models also contain variables that try to capture some job and firm characteristics. A variable to capture overskilling, meaning that workers report they have skills and qualifications necessary to do a more demanding job, is included in the models as it may signify a poor job match. The results show a positive relationship between being overeducated and job mobility. For example, in Spain workers who report that they are overeducated are 1.4 per cent more likely to change jobs using the probit estimates. In the model that controls for misclassification, the effect of being overskilled is 3 times greater. A firm size effect is included to capture the fact that those working in a large firm may have more alternative employment opportunities within the firm and so are less likely to change employers. The results indicate that workers in large firms have a lower probability of changing jobs. This negative effect is consistent across countries and, as before, is more marked in the misclassification models.

A range of satisfaction variables are included to try to capture aspects of the job and work place such as the organisation of work, conditions in the workplace etc. The satisfaction variables are dummy variables constructed from responses to a range of questions about how satisfied a worker is with various aspects of their present job. These factors are typically unobserved so these subjective measures may provide some useful information. Some of these variables may also reveal the workers' assessment of the quality of the match. Satisfaction is measured on a scale of 1 to 6, where 1 denotes not satisfied and 6 denotes fully satisfied. The dummy variables are equal to one when the rating is 4 or over. Although the marginal effects associated with these variables do not always have the expected sign, a negative relationship between these job satisfaction variables and mobility is evident across countries.

Working in the public sector has a negative effect on job mobility in all countries, apart from Italy where the effect is positive but insignificant. As before, the marginal effects from the models that control for misclassification are much bigger in magnitude. The results also generally show that workers in the construction and market services sector are more likely to change jobs than those in the nonmarket services sector.

Regional dummy variables and regional unemployment rates are included to control for factors such as access to alternative jobs and local labour market conditions. The unemployment rates are only significant in Portugal and Denmark and the sign of the impact differs across countries. We would expect the unemployment rate to have a negative effect on voluntary mobility and a positive effect on involuntary mobility. However, as these models do not distinguish between different types of mobility, the effect of local labour market conditions on overall mobility is ambiguous.

Overall, apart from age, personal and household characteristics are generally not important in explaining the probability of job mobility, occupation and sector of origin have some role while firm and job characteristics have an important role in explaining decisions to change jobs. The effects of these variables are much stronger in the models that control for misclassification.

6.5 Conclusions

This chapter finds that there is measurement error in responses to a question about tenure in the ECHP. Survey data on tenure are very often used to capture job changes. Given the extent of measurement error evident in the data it is likely that cases are misclassified as being job changes when truly no change has taken place and vice versa. An estimator developed by Hausman *et al.* (1998) is used to control for misclassification. The results indicate that, by ignoring misclassification, the true number of job changes is underestimated in several countries.

The range of estimates for this undercounting of job changes varies across countries. The results indicate that the true number of job changes is underestimated by around 58 per cent in Ireland and Spain while the comparable figure for Greece is 81 per cent. Countries such as Greece, Ireland, Portugal and Spain display a high level of inconsistencies in responses to the question about tenure and these countries also have high and significant estimates of misclassification.

In addition, the chapter finds that similar factors are important in determining job mobility across countries. Apart from age, personal and household characteristics are generally not important in explaining the probability of job mobility; occupation and sector of origin have some role while firm and job characteristics have an important role in explaining decisions to change jobs. The effect of these variables on the probability of changing jobs is much stronger in the models that control for misclassification.

Appendix Table 6.1: Revolving Balanced Panels, Number of Workers and Job Changes*

	1995	1996	1997	1998	1999	2000	2001
				AUSTRIA			
Revolving Balanced Panel	3,045	3,029	2,992	2,998	2,985	2,963	2,950
No. of Workers		1,888	1,945	1,947	1,956	2,012	2,039
No. of Job Changes		108	104	107	105	98	112
Job Mobility Rate		5.7%	5.3%	5.5%	5.4%	4.9%	5.5%
				BELGIUM			
Revolving Balanced Panel	3,121	3,098	3,095	3,088	3,107		
No. of Workers	1,680	1,722	1,712	1,688	1,702		
No. of Job Changes	54	90	83	92	95		
Job Mobility Rate	3.2%	5.2%	4.8%	5.5%	5.6%		
				DENMARK			
Revolving Balanced Panel	2,191	2,149	2,106	2,084	2,082	2,081	2,078
No. of Workers	1,590	1,613	1,579	1,578	1,621	1,639	1,640
No. of Job Changes	160	180	203	201	155	199	193
Job Mobility Rate	10.1%	11.2%	12.9%	12.7%	9.6%	12.1%	11.8%
				FINLAND			
Revolving Balanced Panel		3,189	3,084	3,154	3,147	3,208	3,251
No. of Workers			2,090	2,088	2,131	2,161	2,196
No. of Job Changes			115	204	149	173	171
Job Mobility Rate			5.5%	9.8%	7.0%	8.0%	7.8%
				FRANCE			
Revolving Balanced Panel	5,812	5,738	5,690	5,663	5,671	5,704	5,780
No. of Workers	3,592	3,804	3,500	3,468	3,679	3,695	3,780
No. of Job Changes	127	136	113	159	172	203	229
Job Mobility Rate	3.5%	3.6%	3.2%	4.6%	4.7%	5.5%	6.1%
				GERMANY			
Revolving Balanced Panel	6,876	6,812	6,741	6,674	6,613	6,611	6,607
No. of Workers	4,169	4,372	4,307	4,250	4,281	4,434	4,442
No. of Job Changes	388	366	355	305	343	360	422
Job Mobility Rate	9.3%	8.4%	8.2%	7.2%	8.0%	8.1%	9.5%
				GREECE			
Revolving Balanced Panel	5,013	4,942	4,870	4,850	4,834	4,907	4,939
No. of Workers	2,637	2,801	2,757	2,801	2,803	2,816	2,878
No. of Job Changes	99	146	267	95	100	69	54
Job Mobility Rate	3.8%	5.2%	9.7%	3.4%	3.6%	2.5%	1.9%
				IRELAND			
Revolving Balanced Panel	2,470	2,430	2,387	2,355	2,365	2,365	2,422
No. of Workers	1,217	1,264	1,297	1,366	1,433	1,466	1,511
No. of Job Changes	84	90	111	149	157	202	161
Job Mobility Rate	6.9%	7.1%	8.6%	10.9%	11.0%	13.8%	10.7%
				ITALY			
Revolving Balanced Panel	8,062	8,045	8,028	8,008	7,984	7,982	7,980
No. of Workers	3,941	4,068	4,031	4,010	4,127	4,218	4,273
No. of Job Changes	140	155	216	193	162	183	210
Job Mobility Rate	3.6%	3.8%	5.4%	4.8%	3.9%	4.3%	4.9%
				THE NETHERLANDS			
Revolving Balanced Panel	4,012	3,949	3,896	3,843	3,808	3,806	3,862
No. of Workers	2,210	2,392	2,384	2,402	2,455	2,476	2,553
No. of Job Changes	153	148	163	232	200	283	269
Job Mobility Rate	6.9%	6.2%	6.8%	9.7%	8.1%	11.4%	10.5%
				PORTUGAL			
Revolving Balanced Panel	5,582	5,563	5,549	5,574	5,605	5,651	5,724
No. of Workers	3,379	3,635	3,670	3,691	3,794	3,836	3,923
No. of Job Changes	150	206	211	216	209	282	243
Job Mobility Rate	4.4%	5.7%	5.7%	5.9%	5.5%	7.4%	6.2%
				SPAIN			
Revolving Balanced Panel	5,956	5,900	5,884	5,892	5,968	6,057	6,084
No. of Workers	2,556	2,619	2,629	2,752	2,918	3,096	3,236
No. of Job Changes	179	201	187	307	328	349	412
Job Mobility Rate	7.0%	7.7%	7.1%	11.2%	11.2%	11.3%	12.7%

*Note: Number of workers refers to number of workers employed in consecutive two-year periods

Appendix Table 6.2: Revolving Balanced Panels and the Number of Cases where the Year that a Worker's Job Started is Missing

	1995	1996	1997	1998	1999	2000	2001
				AUSTRIA			
Revolving Balanced Panel	3,045	3,029	2,992	2,998	2,985	2,963	2,950
less cases where starting year with employer is missing for workers	169	180	187	194	197	199	196
	2,876	2,849	2,805	2,804	2,788	2,764	2,754
				BELGIUM			
Revolving Balanced Panel	3,121	3,098	3,095	3,088	3,107		
less cases where starting year with employer is missing for workers	194	194	192	191	191		
	2,927	2,904	2,903	2,897	2,916		
				DENMARK			
Revolving Balanced Panel	2,191	2,149	2,106	2,084	2,082	2,081	2,078
less cases where starting year with employer is missing for workers	67	70	71	72	72	74	73
	2,124	2,079	2,035	2,012	2,010	2,007	2,005
				FINLAND			
Revolving Balanced Panel		3,189	3,084	3,154	3,147	3,208	3,251
less cases where starting year with employer is missing for workers		156	156	166	171	172	171
		3,033	2,928	2,988	2,976	3,036	3,080
				FRANCE			
Revolving Balanced Panel	5,812	5,738	5,690	5,663	5,671	5,704	5,780
less cases where starting year with employer is missing for workers	486	495	490	489	475	466	461
	5,326	5,243	5,200	5,174	5,196	5,238	5,319
				GERMANY			
Revolving Balanced Panel	6,876	6,812	6,741	6,674	6,613	6,611	6,607
less cases where starting year with employer is missing for workers	1,021	1,045	1,050	1,051	1,046	1,050	1,045
	5,855	5,767	5,691	5,623	5,567	5,561	5,562
				GREECE			
Revolving Balanced Panel	5,013	4,942	4,870	4,850	4,834	4,907	4,939
less cases where starting year with employer is missing for workers	487	486	484	481	465	462	451
	4,526	4,456	4,386	4,369	4,369	4,445	4,488
				IRELAND			
Revolving Balanced Panel	2,470	2,430	2,387	2,355	2,365	2,365	2,422
less cases where starting year with employer is missing for workers	132	133	133	133	128	124	126
	2,338	2,297	2,254	2,222	2,237	2,241	2,296
				ITALY			
Revolving Balanced Panel	8,062	8,045	8,028	8,008	7,984	7,982	7,980
less cases where starting year with employer is missing for workers	938	959	963	969	963	957	944
	7,124	7,086	7,065	7,039	7,021	7,025	7,036
				THE NETHERLANDS			
Revolving Balanced Panel	4,012	3,949	3,896	3,843	3,808	3,806	3,862
less cases where starting year with employer is missing for workers	381	384	387	388	388	389	389
	3,631	3,565	3,509	3,455	3,420	3,417	3,473
				PORTUGAL			
Revolving Balanced Panel	5,582	5,563	5,549	5,574	5,605	5,651	5,724
less cases where starting year with employer is missing for workers	446	445	437	433	423	408	401
	5,136	5,118	5,112	5,141	5,182	5,243	5,323
				SPAIN			
Revolving Balanced Panel	5,956	5,900	5,884	5,892	5,968	6,057	6,084
less cases where starting year with employer is missing for workers	528	534	541	540	542	540	530
	5,428	5,366	5,343	5,352	5,426	5,517	5,554

Appendix 6A: Differences in Irish Data between LIS and ECHP

As noted in Section 6.2, the ECHP data for Ireland is generated from the LIS data but is coded differently by Eurostat to ensure some standardisation in variables across countries. Appendix Table 6.3 shows the number of jobs that are reported to be recent (i.e. that are reported to have begun, at the earliest, two years before the person joined the survey) and that are observed more than once over the survey period (so we have at least two starting dates to compare within jobs).[73] This data comes from the ECHP version of the Irish data.[74] The table shows that 16.7 per cent of these recent jobs have an inconsistency in reported starting dates.

Appendix Table 6.3: Number of Recent Jobs Observed over the Sample Period using <u>ECHP</u> data

	No. of 'Recent' Jobs that are Observed at least twice	% of 'Recent' Jobs that have Inconsistencies in Starting Dates
Ireland	830	16.7%

Appendix Table 6.4 shows the number of recent jobs using the LIS data. It uses two definitions of a 'recent' job. The first counts the number of jobs where the starting year associated with the job the first time it is observed is at least 1994 (e.g. this implies that a job observed for the first time in 1995 is counted as recent if the starting date associated with the job is 1994 or later). The second measure counts the number of observations where the starting year associated with the job the first time it is observed is at least 1993. Both measures produce a similar estimate of the number of recent jobs to what is found in the ECHP data. However, there is a dramatic difference in the percentage of these recent jobs that have an inconsistency in reported starting dates. Using the LIS data, around 70 per cent of recent jobs have an inconsistency in reported starting dates, whereas the ECHP version of the Irish data indicates that the inconsistency rate is closer to 17 per cent.

---

[73] The reason 'recent' jobs in the ECHP data are examined is that full starting dates (i.e. month and year) are only recorded if the person reports that they started working with their current employer at the earliest two years before they joined the survey.

[74] This table repeats the data for Ireland from Table 6.2 in the body of the chapter.

Appendix Table 6.4: Number of Recent Jobs Observed over the Sample Period using LIS data

	No. of 'Recent' Jobs that are Observed at least twice	% of 'Recent' Jobs that have Inconsistencies in Starting Dates
No. jobs where the start year in the first year the job is observed **is at least 1994**	801	69.8% (or 559 jobs)
No. jobs where the start year in the first year the job is observed **is at least 1993**	858	70.7% (or 607 jobs)

Part of the difference in inconsistency rates between the two datasets can be explained. The ECHP only records full starting dates when the date is at the earliest two years before the person joined the survey (which for most people will be 1993/1994). If a person were to dramatically underestimate the year that they started working in their current job, then this starting date would not be recorded in the ECHP.[75] Appendix Table 6.5 shows the number of recent jobs in the LIS data that have an inconsistency in starting dates, where in at least one year the starting date is significantly underestimated (i.e. it is before 1994 or 1993). Around 17 per cent of recent jobs with an inconsistency (or 12-13 per cent of all recent jobs) have an inconsistency where tenure is significantly underestimated in at least one year where the job is observed.

Appendix Table 6.5: Number of Recent Jobs Observed over the Sample Period using LIS data that have an Inconsistency in Starting Dates

	No. of 'Recent' Jobs that have an Inconsistency in Starting Dates	Number of 'Recent' Jobs that have an inconsistency in starting dates where tenure is significantly underestimated in at least one year
No. jobs where the start year in the first year the job is observed **is at least 1994**	559	105
No. jobs where the start year in the first year the job is observed **is at least 1993**	607	102

This type of inconsistency does not happen in the ECHP data (the earliest full starting date is January 1993 for jobs in the ECHP dataset) which may indicate that the country data is 'fixed'/'smoothed' when it is being coded for the ECHP. Assuming that the LIS accurately records the answers that people give, this may indicate that the inconsistency rate in the ECHP is being underestimated.

[75] For example, suppose we observe someone in 1995 and 1996 and in 1995 they report that they started working in their job in 1994 (i.e. it is a recent job). However in 1996 they could report that they started working in their job in 1990. This value will be recorded in LIS but not in the ECHP dataset (or the response may be 'fixed' in the ECHP dataset).

Appendix Table 6.6: Descriptive Statistics*

	Austria		Belgium		Denmark		Finland		France		Germany	
	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.
Age	38.6	10.5	38.8	8.7	41.1	9.7	41.1	9.5	40.2	9.2	39.2	10.4
Age Squared	1602.2	810.9	1579.4	692.2	1780.5	793.7	1781.1	773.2	1701.8	739.8	1642.9	827.0
Female	0.35	0.48	0.37	0.48	0.44	0.50	0.48	0.50	0.41	0.49	0.35	0.48
Child	0.50	0.50	0.55	0.50	0.49	0.50	0.51	0.50	0.53	0.50	0.44	0.50
Female*Child	0.16	0.37	0.20	0.40	0.22	0.42	0.25	0.43	0.21	0.41	0.12	0.32
Living in a Couple	0.70	0.46	0.81	0.40	0.82	0.38	0.82	0.39	0.79	0.41	0.77	0.42
Female*Living in a Couple	0.23	0.42	0.28	0.45	0.36	0.48	0.39	0.49	0.30	0.46	0.25	0.43
Education – low (ISCED 0-2)	0.19	0.39	0.23	0.42	0.16	0.36	0.19	0.40	0.35	0.48	0.18	0.38
Education – medium (ISCED 3)	0.74	0.44	0.33	0.47	0.45	0.50	0.41	0.49	0.36	0.48	0.57	0.50
Education – high (ISCED 5-7)	0.07	0.25	0.44	0.50	0.39	0.49	0.40	0.49	0.29	0.46	0.25	0.44
Recent Training	0.31	0.46	0.26	0.44	0.63	0.48	0.59	0.49	0.14	0.34	0.16	0.36
Satisfied with Number of Working Hours	0.83	0.37	0.78	0.42	0.85	0.35	0.76	0.42	0.54	0.50		
Satisfied with Working Times	0.90	0.30	0.83	0.38	0.88	0.32	0.82	0.38	0.82	0.38		
Satisfied with Working	0.93	0.26	0.79	0.41	0.87	0.33	0.82	0.38	0.78	0.41		
Satisfied with Distance to Job/Commuting	0.85	0.35	0.81	0.39	0.85	0.35	0.81	0.39	0.84	0.37		
Overeducated	0.59	0.49	0.63	0.48	0.60	0.49	0.63	0.48	0.50	0.50		
Public	0.25	0.43	0.20	0.40	0.40	0.49	0.34	0.47	0.34	0.48	0.18	0.39
Number of Employees > 50	0.40	0.49	0.52	0.50			0.34	0.47			0.45	0.50
Occupation of Origin:												
Manager	0.07	0.26	0.06	0.24	0.08	0.26	0.10	0.30	0.06	0.23	0.06	0.23
Professional	0.04	0.20	0.19	0.39	0.20	0.40	0.20	0.40	0.10	0.30	0.12	0.33
Technicians	0.16	0.37	0.17	0.37	0.21	0.41	0.17	0.38	0.22	0.41	0.20	0.40
Clerks	0.13	0.33	0.20	0.40	0.13	0.33	0.09	0.28	0.17	0.37	0.10	0.31
Service	0.13	0.33	0.08	0.27	0.10	0.30	0.11	0.32	0.11	0.31	0.08	0.27
Skilled Ag. & Fishery	0.14	0.34	0.01	0.11	0.02	0.14	0.10	0.30	0.02	0.15	0.01	0.12
Craft & Trades	0.19	0.39	0.11	0.31	0.11	0.32	0.12	0.33	0.14	0.35	0.25	0.43
Plant & Machine Operators	0.08	0.27	0.08	0.26	0.09	0.28	0.07	0.25	0.12	0.33	0.12	0.32
Elementary Occuoations	0.07	0.25	0.10	0.30	0.06	0.23	0.04	0.19	0.06	0.23	0.05	0.23
Sector of Origin:												
Agriculture, Mining and Utilities	0.17	0.37	0.04	0.19	0.05	0.21	0.13	0.33	0.04	0.20	0.04	0.21
Manufacturing	0.22	0.41	0.26	0.44	0.20	0.40	0.19	0.40	0.22	0.41	0.35	0.48
Building	0.09	0.28	0.06	0.23	0.06	0.24	0.05	0.22	0.07	0.25	0.10	0.31
Market Services	0.32	0.47	0.37	0.48	0.35	0.48	0.36	0.48	0.37	0.48	0.32	0.46
Non Market Services	0.21	0.41	0.27	0.45	0.35	0.48	0.27	0.44	0.30	0.46	0.18	0.38
Local Unemployment Rate	3.73	1.16	6.35	2.14	5.75	1.67	9.35	3.05	10.24	2.36	6.32	2.73

* Note: A regional breakdown is not available in the ECHP for Denmark or The Netherlands so the figures in the table refer to the national unemployment rate. The overeducation and satisfaction variables are not available for Germany as either the relevant questions were not asked or the information is not available for other reasons. Data on firm size for Denmark and France are not included in the table as these variables are not used in the country models because they would dramatically reduce the sample size.

Appendix Table 6.6: Descriptive Statistics cont'd*

	Greece		Ireland		Italy		The Netherlands		Portugal		Spain	
	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.
Age	39.9	10.4	38.6	10.9	39.3	10.0	39.3	10.0	38.5	11.0	38.9	10.4
Age Squared	1697.8	839.0	1609.1	859.3	1644.8	799.0	1644.	799.0	1601.7	864.4	1619.6	819.3
Female	0.33	0.47	0.31	0.46	0.32	0.47	0.32	0.47	0.40	0.49	0.29	0.46
Child	0.51	0.50	0.58	0.49	0.47	0.50	0.47	0.50	0.53	0.50	0.51	0.50
Female*Child	0.16	0.36	0.17	0.38	0.14	0.35	0.14	0.35	0.21	0.41	0.13	0.33
Living in a Couple	0.75	0.43	0.66	0.47	0.72	0.45	0.72	0.45	0.73	0.44	0.70	0.46
Female*Living in a Couple	0.24	0.43	0.18	0.39	0.22	0.41	0.22	0.41	0.29	0.45	0.17	0.38
Education – low (ISCED 0-2)	0.51	0.50	0.36	0.48	0.47	0.50	0.47	0.50	0.82	0.38	0.54	0.50
Education – medium (ISCED 3)	0.31	0.46	0.42	0.49	0.43	0.49	0.43	0.49	0.11	0.32	0.20	0.40
Education – high (ISCED 5-7)	0.19	0.39	0.21	0.41	0.10	0.30	0.10	0.30	0.06	0.25	0.26	0.44
Recent Training	0.05	0.22	0.21	0.41	0.11	0.31	0.11	0.31	0.06	0.24	0.20	0.40
Satisfied with Number of Working Hours	0.52	0.50	0.82	0.39	0.61	0.49	0.61	0.49	0.76	0.43	0.60	0.49
Satisfied with Working Times	0.54	0.50	0.87	0.34	0.65	0.48	0.65	0.48	0.82	0.38	0.70	0.46
Satisfied with Working	0.54	0.50	0.88	0.33	0.67	0.47	0.67	0.47	0.85	0.36	0.74	0.44
Satisfied with Distance to Job/Commuting	0.71	0.46	0.88	0.32	0.68	0.47	0.68	0.47	0.82	0.38	0.73	0.45
Overeducated	0.49	0.50	0.48	0.50	0.46	0.50	0.46	0.50	0.42	0.49	0.52	0.50
Public	0.15	0.36	0.20	0.40	0.20	0.40	0.20	0.40	0.13	0.34	0.14	0.34
Number of Employees > 50	0.12	0.32	0.33	0.47	0.26	0.44	0.26	0.44	0.22	0.42	0.31	0.46
Occupation of Origin:												
Manager	0.12	0.33	0.11	0.32	0.03	0.18	0.03	0.18	0.06	0.24	0.09	0.29
Professional	0.10	0.30	0.13	0.34	0.08	0.28	0.08	0.28	0.04	0.21	0.11	0.31
Technicians	0.04	0.20	0.09	0.29	0.11	0.31	0.11	0.31	0.07	0.25	0.10	0.30
Clerks	0.10	0.30	0.10	0.30	0.17	0.38	0.17	0.38	0.09	0.29	0.09	0.28
Service	0.11	0.31	0.10	0.30	0.14	0.35	0.14	0.35	0.14	0.35	0.12	0.33
Skilled Ag. & Fishery	0.21	0.41	0.13	0.34	0.05	0.21	0.05	0.21	0.12	0.33	0.07	0.25
Craft & Trades	0.18	0.38	0.13	0.33	0.23	0.42	0.23	0.42	0.23	0.42	0.21	0.41
Plant & Machine Operators	0.09	0.29	0.13	0.33	0.08	0.27	0.08	0.27	0.10	0.29	0.11	0.31
Elementary Occuoations	0.05	0.23	0.08	0.27	0.11	0.31	0.11	0.31	0.14	0.35	0.11	0.31
Sector of Origin:												
Agriculture, Mining and Utilities	0.23	0.42	0.17	0.38	0.09	0.29	0.09	0.29	0.16	0.37	0.11	0.31
Manufacturing	0.16	0.37	0.23	0.42	0.25	0.43	0.25	0.43	0.21	0.41	0.22	0.41
Building	0.10	0.30	0.07	0.26	0.09	0.28	0.09	0.28	0.12	0.33	0.11	0.32
Market Services	0.38	0.49	0.34	0.48	0.39	0.49	0.39	0.49	0.36	0.48	0.42	0.49
Non Market Services	0.13	0.33	0.18	0.38	0.19	0.39	0.19	0.39	0.15	0.36	0.14	0.35
Local Unemployment Rate	10.19	2.51	7.13	2.73	14.50	8.75	14.50	8.75	4.47	1.87	17.74	5.84

* Note: A regional breakdown is not available in the ECHP for Denmark or The Netherlands so the figures in the table refer to the national unemployment rate. The overeducation and satisfaction variables are not available for Germany as either the relevant questions were not asked or the information is not available for other reasons. Data on firm size for Denmark and France are not included in the table as these variables are not used in the country models because they would dramatically reduce the sample size.

7. Wage Changes and Job Changes:

Estimation with Measurement Error in a Binary Regressor

7.1 Introduction

Having considered models of job mobility in earlier chapters, we now turn to examining the impact of job mobility on wage growth in Ireland. A number of theoretical models and empirical studies suggest that job mobility makes an important contribution to wage growth, particularly for younger workers. In the empirical literature on job mobility, researchers often rely on self-reported accounts of tenure to determine whether or not a job change has taken place. In chapters 5 and 6, I raised the possibility of substantial inconsistencies or measurement error in these responses.[76]

The main contribution of this chapter is to control for misclassification in job changes in estimating the impact of job mobility on wage growth. The main results are that the probability of undercounting the true number of job changes is high and that ignoring misclassification leads to a significant downwards bias when estimating the wage effects of job mobility.

The chapter also investigates whether heterogeneity in some unobserved individual characteristic can account for the effect of mobility on wages. In addition, it assesses whether there are differential wage impacts depending on the reason for job separation. The chapter also addresses the possible two-way causation between job mobility and wage growth using an instrumental variables approach. It finds that the impact of job mobility on wage growth persists when we control for the endogeneity of job mobility.

This chapter is organised as follows: Section 7.2 describes the econometric problems associated with estimating the impact of job mobility on wage growth. It also reviews

[76] See also Brown and Light (1992).

the effect measurement error in a binary explanatory variable has on OLS estimates and, in particular, discusses an approach to control for misclassification in a binary explanatory variable. Section 7.3 provides some descriptive statistics on wage growth and job mobility. Section 7.4 presents estimation results and Section 7.5 concludes.

7.2 Econometric Approach

The focus of this chapter is to estimate the impact of changing jobs on wage growth. Assuming a linear relationship, the standard model for estimation is:

$$\Delta \log(w_{it}) = \beta m_{it}^{*} + \delta x_{it} + \varepsilon_{it} \tag{7.1}$$

where w_{it} is the natural logarithm of the wage of individual *i* at time *t*, m_{it}^{*} is a dummy variable indicating whether a job change has taken place between *t-1* and *t*, x_{it} is a vector of personal, job and firm characteristics and ε_{it} is a random component that is mean zero and is uncorrelated with m_{it}^{*} and x_{it}. The key parameter of interest, β, captures the average percent difference in wage growth between job changers and job stayers adjusted for worker and job characteristics. Pooled OLS estimation of (7.1) is likely to produce biased estimates. There are three main sources of bias namely unobserved heterogeneity across workers, the endogeneity of job mobility and measurement error in capturing actual job changes. The first two issues have been tackled in the empirical literature and the main contribution of this chapter is to control for measurement error in job mobility status when estimating the impact of job mobility on wage growth. This section describes each of these three problems and also the empirical strategy adopted in the chapter.

7.2.1 Unobserved Heterogeneity

There may be unobserved factors that affect both wage growth and the decision to change jobs and this can lead to bias in the parameter estimates. Intuitively, we want to compare the wage growth of a job changer with what they would have received had they stayed in their job. The estimate of β from (7.1) captures the difference in wage growth between job changers and job stayers. This is unlikely to provide an accurate measure of the effect of mobility, if the average wage growth of stayers does not accurately reflect the average wage growth job changers would have received if they had stayed in their jobs For example, in the mover-stayer model (Blumen *et al*. (1955)) described in Chapter 2, we may expect stayers to experience higher wage growth than job changers because they have some underlying personal characteristic that makes them more likely to stay in their job which also makes them more productive, which leads to higher wage growth. Therefore, the mover-stayer model suggests the estimated coefficient of the effect of mobility on wage growth from (7.1) may be biased downwards.

Several techniques are used in the empirical literature to overcome this problem. One approach suggested by Bartel and Borjas (1981) and developed by Mincer (1986) is to use a proxy for the wage growth job changers would have obtained had they not changed jobs. Mincer proposed using the wage growth of those who do not change jobs in the current period but who change jobs in the following period as the proxy. The returns to mobility are then measured as the difference between the wage growth of workers who change jobs in the current period and the wage growth of workers who do not change jobs in the current period but do change jobs in the following period. This approach has been used by Abbott and Beach (1994), Campbell (2001) and Keith and McWilliams (1999). The key assumption of this approach is that workers who stay in their job in the initial period and who change jobs in the subsequent period are more similar, in terms of unobservable characteristics, to those who change jobs in the initial period than workers who stay in their jobs in both periods.[77]

In more recent empirical work, because of the availability of panel data, the issue of unobserved heterogeneity has been dealt with in a fixed effects estimation framework. This involves explicitly including unobserved heterogeneity into the regression model as follows:

[77] Another approach in the literature is to estimate separate equations for job movers and stayers, usually with a correction for sample selection bias associated with job mobility status. This approach has been used by Borjas and Rosen (1980), Holmlund (1984), Kidd (1991), Marshall and Zarkin (1987) and Simpson (1990).

$$\Delta \log(w_{it}) = \beta m_{it}^{*} + \delta x_{it} + \varepsilon_{it}$$

$$\text{where } \varepsilon_{it} = \eta_i + \upsilon_{it} \tag{7.2}$$

The error component has two distinct parts: the first part, η_i, captures unobservable individual-specific effects that can vary across individuals but are constant over time for each individual and the second component, υ_{it}, is assumed to be uncorrelated with the observed and unobserved characteristics across individuals and time.[78,79] This specification takes into account the fact that there could be some unobserved individual effect that may be correlated with both wage growth and the decision to change jobs and this can lead to bias in the pooled OLS estimate of job mobility on wage growth.

With panel data, unobserved heterogeneity is usually handled using a fixed effects or random effects model. The fixed effects model allows the individual constant term, η_i to be correlated with other regressors in the model. The estimator transforms all variables to deviations from their sample means for all time periods $\left(x_{it} - \overline{x_i}\right)$, which implies that η_i drops out of the equation because it is constant over time. As a result, it does not generate coefficient estimates for any variable that is constant over time. This approach to dealing with unobserved heterogeneity has been used by Davia (2005), Le Grand and Tahlin (2002), Light and McGarry (1998), Naticchiono and Panigo (2004), Munasinghe and Sigman (2004) and Pavlopoulos *et al.* (2007). The key difference between the fixed and random effects models is that the random effects model assumes that the η_i are uncorrelated with the other regressors in the model; a quite restrictive assumption. In the empirical analysis, individual effects are tested for and we also test to discriminate between the fixed and random effects models.

---

[78] In contrast, the pooled OLS model assumes that the intercept is common across all individuals.
[79] This specification assumes that the only source of unobservable heterogeneity is at the individual level. However, the error term can be expanded to include unobservable job-specific effects which would capture, say, the quality of the job match e.g. Light and McGarry (1998) control for both individual fixed effects and job specific effects.

7.2.2 Reverse Causality

One possible source of endogeneity in model (7.1) is two-way causation; not only is wage growth affected by job mobility but also job changes may occur in anticipation of higher wage growth. If this feedback from wages to job mobility occurs then m_{it}^* will be correlated with ε_{it} in (7.1) as m_{it}^* depends on $\Delta\log(w_{it})$ which directly depends on ε_{it} (i.e. $\mathrm{E}(\varepsilon_{it}m_{it}^*)\neq 0$). In addition, unobserved factors that affect both the decision to change jobs and wage growth can cause endogeneity bias.[80]

One approach in the literature to overcome this problem is to use an instrument for mobility status. The idea is to replace the job mobility variable with another variable that is highly correlated with job mobility but uncorrelated with the error term in the wage growth equation. Possible instruments such as housing tenure status, job satisfaction (in particular the components of job satisfaction that do not refer to satisfaction with wages) and dummies for the region in which a person lives have been suggested in the literature. Davia (2005) uses the predicted probabilities from a probit model of job change as an instrument for job mobility.

7.2.3 Misclassification

Measurement Error in Binary Regressors

Measurement error in binary variables takes the form of misclassification (a true 1 can be classified as a 0, or a true 0 can be classified as a 1). Define m_{it} to be a noisy indicator of the binary variable m_{it}^*. More specifically, we can write the observed value, m_{it}, as the sum of the true value, m_{it}^*, plus a measurement error, u_{it}, as follows:

$$m_{it} = m_{it}^* + u_{it} \tag{7.3}$$

[80] A related literature focuses on estimating the returns to tenure (as tenure can be viewed as a series of previous quit and layoff decisions). For example, various studies use an instrumental variable approach to control for the endogeneity of tenure (e.g. Altonji and Shakotko (1987), Topel (1991) and Dustmann and Meghir (2005)).

where u_{it} is mean zero. When $m_{it}^* = 1$, the variable m_{it} can only take on two values; 1 if it is correctly classified so $u_{it} = 0$ (i.e. there is no measurement error), or $m_{it} = 0$ so $u_{it} = -1$. When $m_{it}^* = 1$, the variable m_{it} can never overestimate/over-report the true value. Likewise, when $m_{it}^* = 0$, the variable m_{it} can never underestimate/under-report the true value; u_{it} is either 0 or +1. Therefore the measurement error, u_{it}, is negatively correlated with the true variable, m_{it}^*, so misclassification in a dummy variable leads to non-classical measurement error.

Aigner (1973) and others have shown that when a binary regressor is misclassified the least squares coefficient estimates are biased towards zero and that additional assumptions or knowledge about the extent of misclassification in the data is needed to correct the estimates. To illustrate this point, consider the model given in (7.1) where m_{it}^* denotes true job changes. Suppose we do not observe m_{it}^* but rather we observe m_{it} (as defined in (7.3)), which misclassifies some of the observations. Let α_0 denote the probability that a true job stay is misclassified as a job change i.e. $\alpha_0 = \Pr(m_{it} = 1 | m_{it}^* = 0) = \Pr(u_{it} = 1 | m_{it}^* = 0)$ and α_1 denote the probability that a job change is misclassified as a job stay i.e. $\alpha_1 = \Pr(m_{it} = 0 | m_{it}^* = 1) = \Pr(u_{it} = -1 | m_{it}^* = 1)$.[81] Let μ denote the mean of m_{it}^*. Since m_{it}^* is a binary variable, μ corresponds to the probability of truly changing jobs; the probability that m_{it}^* is equal to 1. It follows that $\Pr(m_{it} = 1) = (1-\alpha_1)\mu + \alpha_0(1-\mu)$ i.e. the probability that an observation is observed as a job change is given by the probability that it truly is a job change (μ) and is correctly classified as such $(1-\alpha_1)$ plus the probability that it truly is not a job change $(1-\mu)$ but it has been misclassified as one (α_0). In what follows, let $\Pr(m_{it} = 1) = (1-\alpha_1)\mu + \alpha_0(1-\mu) = p$ for simplicity.

Consider, first of all, a model with a single binary regressor:

[81] This assumes that the misclassification rates are constant across individuals and time and that they only depend on the true value m_{it}^* and not on the other covariates in the model.

$$\Delta \log(w_{it}) = \beta m_{it}^{*} + \varepsilon_{it} \tag{7.4}$$

However, we cannot observe m_{it}^{*} only the mismeasured proxy m_{it}, given in (7.3), so:

$$\begin{aligned}\Delta \log(w_{it}) &= \beta m_{it} - \beta u_{it} + \varepsilon_{it} \\ &= \beta m_{it} + (\varepsilon_{it} - \beta u_{it})\end{aligned} \tag{7.5}$$

Using m_{it} as a proxy for m_{it}^{*} means the measurement error becomes part of the error term in (7.5) and therefore creates an endogeneity bias. Estimating the model given in (7.5) yields an OLS estimator for β with a probability limit:

$$\text{plim}\hat{\beta}_{OLS} = \frac{Cov(m_{it}, \Delta \log(w_{it}))}{Var(m_{it})} = \frac{Cov(m_{it}^{*} + u_{it}, \beta m_{it} + \varepsilon_{it} - \beta u_{it})}{Var(m_{it})}$$

$$= \frac{\beta Cov(m_{it}^{*}, m_{it}) + Cov(m_{it}^{*}, \varepsilon_{it}) - \beta Cov(m_{it}^{*}, u_{it}) + \beta Cov(u_{it}, m_{it}) + Cov(u_{it}, \varepsilon_{it}) - \beta Var(u_{it})}{Var(m_{it})}$$

$$= \frac{\beta Cov(m_{it}^{*}, m_{it}) - \beta Cov(m_{it}^{*}, u_{it}) + \beta Cov(u_{it}, m_{it}) - \beta Var(u_{it})}{Var(m_{it})}$$

(as $Cov(m_{it}^{*}, \varepsilon_{it}) = 0$ and $Cov(u_{it}, \varepsilon_{it}) = 0$ because u_{it} and ε_{it} are independent errors, caused by different things, so we do not expect them to be correlated with each other or with m_{it}^{*})

$$= \frac{\beta Cov(m_{it}^{*}, m_{it}) - \beta Cov(m_{it}^{*}, u_{it}) + \beta Cov(u_{it}, m_{it}^{*} + u_{it}) - \beta Var(u_{it})}{Var(m_{it})}$$

$$= \frac{\beta Cov(m_{it}^{*}, m_{it}) - \beta Cov(m_{it}^{*}, u_{it}) + \beta Cov(u_{it}, m_{it}^{*}) + \beta Var(u_{it}) - \beta Var(u_{it})}{Var(m_{it})}$$

$$= \beta\left(\frac{Cov\left(m_{it}^{*}, m_{it}\right)}{Var\left(m_{it}\right)}\right) = \beta\left(\frac{E\left[m_{it}^{*}m_{it}\right] - E\left[m_{it}^{*}\right]E\left[m_{it}\right]}{E\left[m_{it}^{2}\right] - E\left[m_{it}\right]^{2}}\right) = \beta\left(\frac{E\left[m_{it}^{*}E\left[m_{it}\middle|m_{it}^{*}\right]\right] - \mu p}{p(1-p)}\right)$$

$$= \beta\left(\frac{\left[\Pr\left(m_{it}^{*} = 1\right)\left[\Pr\left(m_{it} = 1\middle|m_{it}^{*} = 1\right)\right]\right] - \mu p}{p(1-p)}\right)$$

$$= \beta\left(\frac{\mu(1-\alpha_1) - \mu p}{p(1-p)}\right) = \beta\left(\frac{\mu/p\,(1-\alpha_1 - p)}{(1-p)}\right) = \beta\gamma^{0} \qquad (7.6)$$

where $\gamma^{o} = \dfrac{\mu/p\,(1-\alpha_1 - p)}{(1-p)}$

where γ^{0} is the attenuation coefficient in a model with a single misclassified regressor. As α_0, α_1, μ and p are all greater than zero but less than one and $\mu \approx p$, the attenuation coefficient γ^{0} given in (7.6) is less than one which implies that the OLS estimate of β is biased towards zero. Without knowledge about the misclassification rates, α_0 and α_1, and the probability that an observation is truly a job change, μ, we cannot identify the true β from our data. Furthermore, for very high levels of misclassification the expression for γ^{0} could be negative yielding an OLS estimate of the wrong sign (Kane, Rouse and Staiger (1999)).

Attenuation bias is typically exacerbated in multivariate regression (Angrist and Krueger (1999)). Card (1996) and others have shown that the attenuation factor in this case is given by (see Appendix 7A for details when there are two explanatory variables in the model):

$$\gamma = \frac{\gamma^{0} - \left(R^{2}/(1-\alpha_0 - \alpha_1)\right)}{1-R^{2}} \qquad (7.7)$$

where γ^0 is the attenuation factor from the model with no other covariates, given in (7.6), and R^2 is the theoretical R^2 from a regression of observed job changes on the other explanatory variables in the model.

Misclassification will cause both OLS and fixed effects estimates to be biased towards zero and inconsistent. However, measurement error bias is likely to be amplified in the fixed-effects estimates (Bound *et al.* (2001)). Correctly measured explanatory variables tend to be correlated across time so there is typically much less within-group variation in these variables than in the measurement error (as this will typically exhibit weak or no serial correlation). Therefore, measurement error in fixed effects models tends to reduce the variance in the signal relative to the variance in the noise so attenuation bias in this model can be more severe than both measurement error bias and heterogeneity bias in a pooled OLS model.

Implications for Instrumental Variable Estimation

With a misclassified binary regressor, instrumental variable estimation does not yield a consistent estimate of β. The intuition behind this result is straightforward. A valid instrument must be correlated with the true value, m_{it}^* , and uncorrelated with the error term which is made up of the random error ε_{it} and the measurement error u_{it}. As the measurement error, u_{it}, is correlated with m_{it}^*, any variable (potential instrument) which is correlated with m_{it}^* will also generally be correlated with the measurement error. If an instrument is available, IV estimation will remove the correlation between m_{it}^* and ε_{it} but not between m_{it}^* and u_{it} and so the IV estimate of β will be biased.

In the case of a model with a single binary (misclassified) explanatory variable the IV estimate of β is biased by a factor $1/(1-\alpha_0-\alpha_1)$ (Angrist and Krueger (1999), Kane, Rouse and Staiger (1999)). See Appendix 7B for details. As $0<\alpha_0+\alpha_1$ and generally $\alpha_0+\alpha_1<1$[82], the IV estimate will be biased upwards. The bias only depends on the

[82] If misclassification is so severe that $\alpha_0+\alpha_1>1$, then the estimate of β will have the wrong sign.

misclassification rates and not on the measurement error u_{it}. In a bivariate regression with a mismeasured binary explanatory variable the OLS estimate is biased downwards and the IV estimate is biased upwards so these estimates can be used to bound the true coefficient.

Approaches in the Literature

There are several approaches in the literature to dealing with measurement error in binary regressors. One approach is to exploit external estimates of misclassification rates. Validation surveys can be used to provide estimates of misclassification rates. For example, Freeman (1984) and Card (1996) examine the impact of union membership on wages. They use a validation survey that has both employer and worker reports of union status to estimate the misclassification rates in the reporting of union status.[83] Kane, Rouse and Staiger (1999) adopt a different approach and propose a generalised method of moments technique to obtain consistent estimates when a researcher has two noisy reports of the regressor.

Another approach is to try to bound the estimates. Bollinger (1996) establishes bounds for the true coefficients in a linear regression when a binary regressor is mismeasured.[84] In addition, Bollinger shows how these bounds can be made tighter if information is available on the misclassification rates. Frazis and Loewenstein (2003) extend the procedure proposed by Hausman *et al.* and compute bounds of the misclassification rates without making functional form assumptions. They combine these bounds with the OLS coefficient to bound the true effect of the mismeasured explanatory variable.

Card (1996) and Frazis and Loewenstein (2003) provide an expression for the inconsistency in OLS estimates due to misclassification, assuming the other explanatory variables in the model are perfectly measured, as follows:[85]

[83] Freeman (1984) assumes that the employer report of union status is correctly measured. Card (1996) allows for both the employee and employer reports to be measured with error but assumes that the rate of misreporting from both groups is equal and that the misclassification rates are equal.
[84] Bollinger (2001) extends the methodology to include fixed effects models.
[85] The expression inside the square brackets is equivalent to the reciprocal of the attenuation factor given in equation (7.7) when γ^0 (from equation (7.6)) and $\mu = \dfrac{p-\alpha_0}{1-\alpha_0-\alpha_1}$ have been substituted in for.

$$\beta = \left(\text{plim}\hat{\beta}_{OLS}\right)\left[\frac{p(1-p)(1-\alpha_0-\alpha_1)(1-R^2)}{(p-\alpha_0)(1-p-\alpha_1)-R^2 p(1-p)}\right] \tag{7.8}$$

In the sensitivity analysis in Section 7.4 this expression is used to provide "corrected" OLS estimates of the impact of job mobility on wage growth.

Another strand of the literature uses an instrumental variable approach to handle mismeasured binary regressors. For example, Mahajan (2006) assumes that additional information, in the form of a second variable, is available that is correlated with the unobserved true underlying variable but not related to the measurement error in the binary variable.

7.2.4 Empirical Strategy

This chapter adopts a two-step approach to controlling for misclassification in estimating the effect of job mobility on wage growth. The approach closely follows Brachet (2008) and is similar to Dustman and van Soest (2001).[86] The first step uses the modified probit estimator developed by Hausman *et al.* which generates consistent estimates of the coefficients as well as the misclassification probabilities and, most importantly for this chapter, the probability of truly being a job changer, $\left(\Pr\left(m_{it}^* = 1\right)\right)$. This yields a proxy for m_{it}^* that removes the impact of misclassification. In the second step, model (7.1) is estimated using pooled OLS substituting in for m_{it}^* using the fitted probabilities that an observation is truly a job change calculated in the first stage. The coefficient estimates will be consistent provided the functional form for F(.) in the first step has been correctly specified.[87]

The same approach can be used to control for both measurement error in job changes and unobserved heterogeneity; the wage growth equation in the second step is

[86] Brachet (2008) examines the effect of maternal smoking on infant health where smoking status may be misreported. Dustmann and van Soest (2001) investigate the effect of language fluency of immigrants on earnings where self-reported language proficiency may be misclassified. They allow for misclassification errors that are independent over time and errors that persist over time (a respondent who over (or under) reports once will always tend to over (or under) report). They jointly estimate the earnings and language fluency equations to allow for unobserved heterogeneity in both the language fluency and earnings equation to be correlated.

[87] See Appendix 7C for details.

estimated using a fixed effects or a random effects estimator. In addition, if we have an instrument for job mobility, it can be used in the first stage of the procedure to create a proxy for m_{it}^{*} that removes both the impact of misclassification and the correlation with the error term. It is hard to find good instruments for job mobility. This chapter attempts to control for reverse causality using non-wage elements of job satisfaction as instruments for job mobility.

7.3 Descriptive Statistics

The starting point for the empirical analysis is the sample of workers employed in successive two-year periods from the revolving balanced panel that uses the LIS data, as described in Chapter 3.[88] Two additional restrictions are placed on the sample. The first is that only income from paid employment is considered so self-employed workers and farmers are excluded from the analysis. These workers are excluded because of difficulties in measuring income from self-employment and farming. The second restriction is that workers are excluded in any year that they report they are working part-time (less than 30 hours).[89] These restrictions ensure some degree of homogeneity in the sample. The key dependent variable in the analysis is the change in log real gross hourly wages between period *t-1* and period *t*. [90] In each year, there are around 90 cases where either the wage in period *t-1* or the wage in period *t* or both are not available and these person-year observations are excluded from the analysis. The final sample consists of 1,206 workers and 5,346 person-year observations, observable for various durations over the period 1995 to 2001.

---

[88] The focus of this chapter is on job-to-job transitions. The sample restrictions and definition of job mobility used means that workers cannot be unemployed or leave the labour force for any considerable amount of time between jobs (specifically by more than the amount of time between interviews). Therefore, the sample is probably a length time biased sample of job changers; in the sense that it may over-represent those who experience a relatively short period of unemployment between jobs or who leave the labour force for a relatively short period between jobs and under-represent those who are unemployed or leave the labour force for longer durations between jobs.

[89] This means that part-time workers are included in the sample in other waves if they are working full-time; however the results presented in the next section are similar to those when part-time workers are deleted entirely from the sample.

[90] Another reason for focussing on full-time workers is the possibility of measurement error in reported usual hours worked. Baum-Snow and Neal (2009) show that there is substantial measurement error in hourly wages for part-time workers. Using US Census data, they perform a validation exercise with the Current Population Survey and find that a significant proportion of workers respond to a question about usual hours of work per week as if the question asked about usual hours of work per day.

Table 7.1 provides preliminary evidence of the relationship between wage growth and job mobility. The table shows that the annual average wage growth for all workers is 8.5 per cent. The next two rows of the table divides workers into job 'movers' and 'stayers', where job movers are those who change jobs *at some point* over the period 1995 to 2001 and stayers are those who are observed in the same job over the entire period. Job movers experience higher but more variable wage growth than those who stay in their jobs. An examination of real wage growth at different points in the distributions for job movers and stayers reveal that they are closest at the 25th percentile; however at the median and 75th percentiles wage growth of job changers is over 1.5 times that of job movers.

It may also be important to distinguish between different types of mobility when looking at wage effects associated with changing jobs. The table shows the number of workers that experience voluntary, involuntary and other types of mobility.[91] Voluntary movers experience higher wage growth than involuntary movers and stayers, as expected. However involuntary movers record higher wage growth than job stayers which is surprising. The previous empirical literature has shown that involuntary job movers can experience wage losses, not just at the time of job change but that these losses can be permanent, especially if there is a period of unemployment between jobs (e.g. Garcia Perez and Rebollo Sanz (2005)). This effect is not evident in Table 7.1. This may be due to the fact that the sample is one where workers have a very high attachment to the labour force; workers need to be employed in consecutive two-year periods to be included in the sample. This excludes the type of transition where a worker experiences a long spell of unemployment, such as where a worker moves from being employed to unemployed for more than a year to employed again. In addition, the time period under consideration is one with very strong economic and employment growth so it is possible that any reputation effects associated with involuntary mobility may be reduced and/or job search costs may be lower as jobs are more plentiful.

The table also shows the average wage growth for workers that move once and for workers that move more than once. Here, we do not distinguish between the types of

[91] Other movers are those who do not state a reason for their job separation or who experience different types of mobility (e.g. they experience both a voluntary and an involuntary quit).

move a worker may make, rather the number of moves. Workers who change jobs more than once experience higher wage growth than those who only move once. There do not appear to be any reputation effects associated with repeated mobiliy.[92]

Table 7.1: Average Within-Person Wage Growth*

	No. of People	*Mean*	*Standard Error*	*25th Percentile*	*50th Percentile*	*75th Percentile*
All Workers	1,206	0.085	0.005	0.01	0.06	0.13
Job Stayer	766	0.072	0.006	0.01	0.05	0.11
Job Mover	440	0.109	0.011	0.02	0.08	0.17
Voluntary Job Mover	223	0.118	0.014	0.03	0.10	0.17
Involuntary Job Mover	78	0.092	0.030	0.01	0.05	0.23
Other Movers	139	0.105	0.018	0.03	0.08	0.17
Move Once	256	0.101	0.015	0.01	0.07	0.16
Move more than Once	184	0.120	0.014	0.03	0.11	0.18

* Note: Wage growth is defined as $\log(w_t) - \log(w_{t-1})$ where w_t and w_{t-1} are real gross hourly wages in Euros reported at time *t* and *t-1*.

Table 7.1 examines average within-person wage growth. However, it does not control for the timing of job changes so the average wage growth reported for job movers refers to workers who move in any year over the period. Controlling for the timing of job changes helps to disentangle whether the higher wage growth of job movers described in Table 7.1 is attributable to a discrete jump in wages at the time of starting a new job or if changing jobs shifts a worker onto a higher wage growth profile. Table 7.2 shows the annual average wage growth for job 'moves' and job 'stays'. The unit of analysis has shifted from people in Table 7.1 to person-year observations so the mean wage growth reported in Table 7.2 is simply the average change in log wages between *t* and *t-1* across all observations. In the case of job moves this refers to the change in log wages associated with the previous job at time *t-1* and the new job at time *t*.

There are very large and variable wage gains related to job moves; a job move is associated with an average wage increase of around 17 per cent, compared to an average wage increase of around 6 per cent for a job stay. Comparing these figures with those from Table 7.1 implies that the bulk of the wage increase associated with job mobility happens at the time of changing jobs. The table also shows that wage

[92] Of course, we observe people at different stages in their working lives and the analysis cannot control for previous mobility history.

growth is greatest for voluntary moves and that involuntary moves are associated with wage gains higher than those of job stays. Wage growth does not differ much depending on the whether it is the first move that we observe a worker making over the period or their second or third move etc over the period.

Table 7.2: Average Wage Growth for Job Stays and Job Moves

	No. of Person-Year Observations	*Mean*	*Standard Error*	*25th Percentile*	*50th Percentile*	*75th Percentile*
All Observations	5,346	0.070	0.004	-0.051	0.045	0.190
Job Stays	4,897	0.061	0.004	-0.051	0.041	0.172
Job Moves	449	0.167	0.019	-0.048	0.134	0.383
Voluntary Job Moves	282	0.200	0.025	-0.022	0.190	0.422
Involuntary Job Moves	103	0.094	0.040	-0.113	0.076	0.328
Other Moves	64	0.140	0.046	-0.054	0.073	0.334
First Move	232	0.177	0.026	-0.042	0.155	0.395
2nd + Move	217	0.157	0.029	-0.050	0.110	0.355

The analysis so far has ignored differences in characteristics across workers and these differences may account for some of the variation in wage growth across job movers and stayers. Table 7.3 shows average wage growth for workers by different levels of labour market experience. The first line of the table indicates that wage growth declines with experience or that the wage-experience profile is concave. The next two lines of the table show that job changers with less experience record much higher wage growth than job stayers, however there is no significant difference in wage growth between job changers and stayers who have more than 5 years experience. Disaggregating the job movers by reasons for job separation shows that workers with less experience who undergo any type of mobility have greater wage growth than job stayers with similar levels of experience. This is consistent with there being a bigger return to match quality in the first few years of labour market experience.

Table 7.3: Average Within-Person Wage Growth by Experience*

	<=4 Years			*5-14 Years*			*15+ Years*		
	No. of People	*Mean*	*Std. Error*	*No. of People*	*Mean*	*Std. Error*	*No. of People*	*Mean*	*Std. Error*
All Workers	341	0.134	0.011	323	0.089	0.011	542	0.052	0.007
Job Stayer	141	0.094	0.015	191	0.096	0.014	434	0.053	0.006
Job Mover	200	0.163	0.015	132	0.078	0.017	108	0.048	0.023
Voluntary Job Mover	116	0.160	0.021	71	0.064	0.023	36	0.092	0.025
Involuntary Job Mover	23	0.194	0.044	20	0.061	0.032	35	0.042	0.057
Other Movers	61	0.156	0.024	41	0.112	0.034	37	0.011	0.034
Move Once	107	0.161	0.022	84	0.083	0.024	65	0.027	0.034
Move more than Once	93	0.165	0.021	48	0.071	0.020	43	0.080	0.029

* Experience refers to years of experience in the first year someone is observed in the sample

As before, Table 7.3 may be misleading as it classifies a worker as a job mover even if they only change jobs towards the end of the observation window. Table 7.4 shows wage growth by person-year observations, where each wage change between time *t-1* and time *t* is classified according to whether it is associated with a job move or a job stay. The table implies that for all levels of experience most of the wage increase associated with job mobility occurs at the time of changing jobs. Table 7.3 showed movers with more than 5 years experience have similar wage growth to stayers; however Table 7.4 indicates that at the time of job change moves have bigger impacts on wages than stays for these workers. In addition, as years of experience increase there is a sharp decline in the wage change associated with involuntary mobility as we go from one experience group to the next. Workers with more experience (15+ years) who undergo involuntary mobility do not record wage gains when moving to a new job.

Table 7.4: Average Wage Growth for Job Stays and Job Moves by Experience

	<=4 Years			*5-14 Years*			*15+ Years*		
	No. of Person-Year observations	*Mean*	*Std. Error*	*No. of Person-Year observations*	*Mean*	*Std. Error*	*No. of Person-Year observations*	*Mean*	*Std. Error*
All Observations	716	0.146	0.012	1,567	0.084	0.008	3,063	0.044	0.005
Job Stays	560	0.117	0.012	1,395	0.079	0.008	2,942	0.041	0.005
Job Moves	156	0.249	0.034	172	0.127	0.032	121	0.119	0.033
Voluntary Job Moves	112	0.245	0.043	107	0.147	0.037	63	0.210	0.048
Involuntary Job Moves	23	0.236	0.065	41	0.110	0.078	39	-0.007	0.051
Other Moves	21	0.285	0.077	24	0.065	0.085	19	0.074	0.066
First Move	77	0.283	0.051	82	0.120	0.042	73	0.128	0.037
2nd + Move	79	0.216	0.045	90	0.133	0.047	48	0.104	0.061

Table 7.5 shows average individual wage growth by gender and Table 7.6 shows wage growth by person-year observation and gender. The first row of Table 7.5 indicates that women experience higher average wage growth than men. This is not as expected, although it should be noted that the average wage level for women is around three quarters of the male wage.[93] One of the features of the labour market over this period is the dramatic rise in female labour force participation, driven by rising educational attainment and also improved labour market conditions which encouraged many married women to return to the labour market. In the sample, female employment is concentrated in the services sector – almost 80 per cent of female workers are in the services sector compared with around 45 per cent of male workers. It may be the case that, over this period, the labour market tightened more quickly in the sectors that women were more heavily concentrated in and, as a result, they experienced faster wage growth than men.

Female job movers have slightly stronger wage growth than their male counterparts. Male involuntary job movers do not appear to suffer wage losses, in fact their average wage growth is similar to male job stayers. However, wage growth for female involuntary job movers is actually higher than for female voluntary movers or stayers.[94] This is highlighted when we look at the between-job wage growth

[93] The average real hourly wage over the period is €12.98 for men and €10.24 for women.

[94] This result for female involuntary mobility is not being driven by outliers in the data. As a robustness check, separate wage regressions by gender were estimated and a quarter of female involuntary movers who experience the highest wage growth were excluded from the analysis and it does not have a dramatic effect on the coefficient estimates.

associated with involuntary mobility; Table 7.6 shows that wage changes at the time of involuntary job moves for men are relatively flat while they are very high for involuntary moves made by women. As mentioned above, because female employment is concentrated in certain sectors, they may face different labour market conditions to male workers in other sectors.

Table 7.5: Average Within-Person Wage Growth by Gender

	Male			*Female*		
	No. of People	*Mean*	*Standard Error*	*No. of People*	*Mean*	*Standard Error*
All Workers	753	0.076	0.007	453	0.102	0.009
Job Stayer	477	0.059	0.006	289	0.092	0.011
Job Mover	276	0.104	0.014	164	0.119	0.015
Voluntary Job Mover	132	0.120	0.019	91	0.115	0.022
Involuntary Job Mover	52	0.060	0.041	26	0.155	0.035
Other Movers	92	0.104	0.024	47	0.106	0.023
Move Once	148	0.090	0.022	108	0.117	0.019
Move more than Once	128	0.120	0.017	56	0.122	0.024

Table 7.6: Average Wage Growth for Job Stays and Job Moves by Gender

	Male			*Female*		
	No. of Person-Year observations	*Mean*	*Standard Error*	*No. of Person-Year observations*	*Mean*	*Standard Error*
All Observations	3,564	0.063	0.005	1,782	0.084	0.006
Job Stays	3,272	0.054	0.005	1,625	0.075	0.006
Job Moves	292	0.159	0.024	157	0.182	0.031
Voluntary Job Moves	185	0.215	0.031	97	0.172	0.040
Involuntary Job Moves	71	0.022	0.044	32	0.255	0.079
Other Moves	36	0.147	0.068	28	0.132	0.061
First Move	144	0.152	0.035	88	0.217	0.037
2nd + Move	148	0.167	0.034	69	0.136	0.053

Average individual wage growth by education level is reported in Table 7.7. Workers with higher levels of education have the highest wage growth. Across each education category, job movers experience higher wage growth than stayers, particularly voluntary job movers. Involuntary job changers with a medium level of education do not record wage gains. For this group of workers, this may reflect shifting onto flatter wage profiles as wage gains are recorded at the time of involuntary job moves (see Table 7.8).

Table 7.7: Average Within-Person Wage Growth by Education

	Low Education			*Medium Education*			*High Education*		
	No. of People	*Mean*	*Std. Error*	*No. of People*	*Mean*	*Std. Error*	*No. of People*	*Mean*	*Std. Error*
All Workers	396	0.071	0.009	595	0.083	0.008	215	0.119	0.010
Job Stayer	247	0.050	0.008	370	0.074	0.009	149	0.102	0.012
Job Mover	149	0.105	0.018	225	0.098	0.016	66	0.157	0.019
Voluntary Job Mover	59	0.094	0.030	120	0.116	0.020	44	0.157	0.021
Involuntary Job Mover	45	0.135	0.031	28	-0.005	0.059	5	0.252	0.157
Other Movers	45	0.090	0.033	77	0.108	0.025	17	0.129	0.025
Move Once	91	0.116	0.026	130	0.073	0.022	35	0.169	0.027
Move more than Once	58	0.088	0.022	95	0.133	0.022	31	0.143	0.027

Table 7.8: Average Wage Growth for Job Stays and Job Moves by Education

	Low Education			*Medium Education*			*High Education*		
	No. of Person-Year observations	*Mean*	*Std. Error*	*No. of Person-Year observations*	*Mean*	*Std. Error*	*No. of Person-Year observations*	*Mean*	*Std. Error*
All Observations	2,232	0.058	0.006	2,389	0.075	0.006	725	0.089	0.010
Job Stays	2,047	0.051	0.006	2,171	0.064	0.006	679	0.079	0.009
Job Moves	185	0.132	0.029	218	0.181	0.028	46	0.241	0.064
Voluntary Job Moves	103	0.183	0.041	143	0.205	0.036	36	0.229	0.055
Involuntary Job Moves	57	0.050	0.044	38	0.147	0.069	8	0.156	0.260
Other Moves	25	0.113	0.080	37	0.123	0.051	2	0.793	0.452
First Move	97	0.154	0.038	110	0.181	0.040	25	0.247	0.072
2nd + Move	88	0.109	0.044	108	0.181	0.040	21	0.234	0.115

7.4 Results

This section presents formal econometric estimates of the impact of changing jobs on wage growth. It first presents pooled OLS results, which give an idea of the initial correlation between job mobility and wage growth. It then goes on to control for differences in unobservable characteristics and investigates whether there are differential wage impacts depending on the type of mobility. Then, crucially for this chapter, measurement error in the job change variable is controlled for. A sensitivity analysis illustrates the effect misclassification has on the estimated impact of job mobility on wage growth. Finally, an attempt is made to control for the bias due to the reverse causality between job mobility and wages.

Table 7.9 shows the pooled OLS estimates of the effect of changing jobs on wage growth. The dependent variable is the change in log wages between time *t-1* and time *t*. The first specification in the table (Model 1) contains no additional regressors (other than a constant term). The coefficient estimate on the job change dummy implies that the average increase in wage growth associated with changing jobs is around 10½ per cent and this effect is highly significant.

Model 2 in the table includes the standard set of control variables that determine wage growth.[95] Specifically, it includes traditional human capital variables such as age, experience and level of education. In addition, some job characteristics are controlled for, such as whether the job is in the public or private sector, the size of the firm etc. Year dummies are included to control for changes in the macroeconomic environment. These variables are included in an attempt to assess to what extent differences in observable characteristics across workers affect the premium associated with changing jobs. The estimate on the job change dummy variable is around 8 per cent indicating that some (around 2½ percentage points) of the higher wage growth associated with changing jobs is attributable to differences in observed characteristics.[96]

The results also indicate that wage growth declines with age and experience. This may reflect the fact that investment in human capital declines over the life-cycle or career-cycle or possibly that employers prefer younger workers. Wage growth is higher for those with third level degrees and above. The results also indicate that there is no significant difference in male and female wage growth or between public and private sector wage growth once differences in observable characteristics are controlled for. Workers in larger firms are expected to have higher wage growth, as larger firms are more likely to have internal labour markets etc, but the estimated effect is negative. Workers in sectors that are more exposed to market forces and where competitiveness is more important for growth, such as the manufacturing sector, have lower estimated

[95] All the explanatory variables (apart from the year dummies) are lagged by one year, so for job changers they refer to their characteristics prior to changing jobs and for job stayers they refer to their situation in the previous year.
[96] Including each of the regressors individually with the job change dummy variable indicates that most of them do not have a substantial impact on the estimated effect of changing jobs on wage growth. The inclusion of variables like age and experience reduce the estimate on the mobility variable by around 0.02.

wage changes. Working in the construction sector has a positive effect on wage growth, probably reflecting the fact that the sector was booming during the period under consideration. However, none of these sectoral wage effects are significant.

Finally, Model 3 in the table includes some job transition characteristics to capture the multifaceted nature of job mobility, such as controls for whether a job changer moves from a small to a big firm. The estimate on the job change dummy variable increases to around 9 per cent whereas the estimate from model 2 that does not control for any transition characteristics is around 8 per cent, indicating that the wage effect associated with changing jobs is not attributable to the nature of the transition. The results also indicate that moving to a big firm leads to higher wage growth and moving to a smaller firm has a negative effect on wage growth relative to those who change jobs but continue in a similar sized firm. The direction of these impacts is as expected. However, controlling for the nature of the transition does not affect the direction of the estimate on firm size, as it remains negative. Those who move from the private to the public sector experience higher wage growth and those who move from the public to the private sector experience wage losses relative to those who change jobs but stay in the same sector although the impacts are not significant. In addition, changing broad sectoral group leads to wage losses relative to those who stay in the same broad group but these effects are not significant.

Table 7.9: Pooled OLS Wage Growth Model^

	Model 1		*Model 2*		*Model 3*	
	Estimate	*Standard Error*	*Estimate*	*Standard Error*	*Estimate*	*Standard Error*
Job Change	**0.1064******	**0.0178**	**0.0794******	**0.0182**	**0.0891******	**0.0239**
Age	-	-	-0.0110**	0.0043	-0.0096**	0.0043
Age Squared	-	-	0.0001***	0.0001	0.0001**	0.0001
Experience	-	-	-0.0019	0.0022	-0.0027	0.0022
Experience Squared	-	-	0.0000	0.0000	0.0000	0.0000
Female	-	-	0.0014	0.0065	0.0016	0.0065
Education: (ref: Low Education)						
Education - Medium	-	-	-0.0087	0.0073	-0.0078	0.0072
Education - High	-	-	0.0078	0.0110	0.0095	0.0111
Public Sector	-	-	0.0070	0.0098	0.0104	0.0100
Number of Employees > 50	-	-	-0.0244***	0.0067	-0.0169***	0.0063
Occupation: (ref: Elementary Occupations)						
Manager	-	-	-0.0016	0.0111	-0.0047	0.0110
Professional	-	-	0.0210**	0.0099	0.0174*	0.0097
Clerk	-	-	0.0186**	0.0092	0.0157*	0.0089
Skilled	-	-	0.0125	0.0114	0.0110	0.0105
Sector of Origin: (ref: Non-Market Services)						
Agric., Mining & Utilities	-	-	0.0073	0.0184	-0.0035	0.0153
Manufacturing	-	-	-0.0108	0.0130	-0.0072	0.0131
Construction	-	-	0.0133	0.0181	0.0162	0.0188
Market Services	-	-	0.0037	0.0097	0.0073	0.0100
Year Dummies:						
1996	-	-	-0.0168	0.0167	-0.0162	0.0167
1997	-	-	0.0139	0.0155	0.0114	0.0154
1998	-	-	-0.0199	0.0146	-0.0186	0.0146
1999	-	-	-0.0069	0.0144	-0.0023	0.0143
2000	-	-	0.0120	0.0140	0.0123	0.0139
2001	-	-	-0.0072	0.0149	-0.0046	0.0149
Job Transition Characteristics:						
Private to Public	-	-	-	-	0.1145	0.0699
Public to Private	-	-	-	-	-0.0395	0.0683
Company Size: Small to Big	-	-	-	-	0.0956*	0.0501
Company Size: Big to Small	-	-	-	-	-0.1177**	0.0495
Sector: Services to Industry	-	-	-	-	-0.0677	0.0615
Sector: Industry to Services	-	-	-	-	-0.0103	0.0532
Constant	0.0608***	0.0028	0.2987***	0.0707	0.2709***	0.0697
Number of Observations	5,346		5,320		5,221	
R-squared	0.0112		0.0294		0.0372	
Prob > F	0.0000		0.0000		0.0000	

Notes: Standard errors are clustered by person. In the table * corresponds to 10%, ** to 5% and *** to 1% level of significance.

Next, we control for unmeasured individual characteristics that are constant over time. Table 7.10 presents results from the two standard panel data models that deal with unobserved heterogeneity, namely a fixed effects and a random effects model.[97] The dummy variables capturing job mobility and the other regressors are similar to what was included in pooled OLS Model 2 (reported in Table 7.9), the results for which are repeated in Table 7.10. The key point to note about the results is that mobility has a strong, positive and significant effect on wage growth even after controlling for unobserved heterogeneity.

Overall, the fixed effects estimates are broadly comparable to pooled OLS estimates.[98] The results indicate that the impact of changing jobs on wage growth is around 11 per cent when we control for unobserved heterogeneity and the effect is significant at the 1 per cent level. This compares with the 8 per cent pooled OLS estimate so controlling for unobserved heterogeneity leads to an increase in the estimated impact of changing jobs on wage growth. This is consistent with the unobservable characteristic being negatively correlated with job mobility (and so the OLS estimate may be biased downwards). However, the F-test for the individual effects does not reject the null hypothesis that the individual effects are not jointly significantly different from zero.[99] This goes against the prediction of the mover-stayer model. The effects of the other variables included in the model are broadly comparable to the estimates from the pooled OLS model.[100]

---

[97] The fixed effects models for wage growth presented in the tables all exclude time dummies. The reason for excluding them is that variables like age, and to some extent experience, change within individuals in the same way over time so the effect of a variable like age in a fixed effects model is interpretable more as a linear time trend. As a result, there is little reason to include these types of variables and also time dummies in a fixed effects model. In addition, the estimate of the effect of job change on wage growth when age and experience are excluded and time dummies are included is practically identical. The regressions also exclude the education variables as they have little within-person variation and reported changes in education level may reflect measurement error (in particular where the reported education level decreases).

[98] As with the pooled OLS models the standard errors in the fixed effects models are clustered at the individual level. Fixed effects account for the time-constant part of the unobservable differences across people. However, it may be the case that unobserved random shocks that influence an individual at time t may also affect their behaviour at time $t+1$ therefore leading to correlated errors within people.

[99] The F-test for the individual effects is calculated from a regression that does not use clustered standard errors because the test is based on the assumption of serially uncorrelated errors.

[100] There are some differences in estimated effects across the pooled OLS and fixed effects models. For example, the estimated effect on the experience variable changes sign. However this variable has low within person variation. Consequently, the coefficient estimate may vary significantly and even in the other direction from the pooled OLS estimate. Generally, in fixed effects models, it is hard to obtain reliable estimates for variables that only change slowly over time.

The estimates from the random effects model are very similar to the pooled OLS estimates. In this model the variation across individuals is assumed to be random and uncorrelated with the other regressors in the model. A Breusch-Pagan Lagrange Multiplier test helps to discriminate between a random effects and OLS regression. The null hypothesis in this test is that the variances across individuals are zero. The test indicates that we reject the null hypothesis and so we conclude that there is a significant difference across individuals so a random effects model is appropriate.

This indicates that the random effects model is the preferred model. This is somewhat surprising as we would expect unobserved effects to be correlated with the explanatory variables i.e. that a fixed effects model is appropriate. A Hausman test can help decide between a fixed effects and random effects model. The null hypothesis in this test is that there is no correlation between the individual effects and regressors. If this is true then both estimators are consistent but the fixed effects estimator is inefficient. If the individual effects and regressors are correlated then the random effect estimator is inconsistent. The Hausman test follows a chi-squared distribution and is equal to 22.47 with a corresponding p-value of 0.0962. This indicates we cannot reject the random effects model at the 5 per cent level of significance, but it can be rejected at the 10 per cent level of significance.[101]

[101] The Hausman test is essentially testing whether the coefficient estimates from the fixed effects model are equal to those from the random effects model. As the fixed effects estimator only uses a small part of the information in the sample it usually has a large standard error. In practice, the Hausman test can very often accept the null hypothesis. Accepting the null hypothesis implies that either the two sets of coefficient estimates are reasonably close or it could indicate that the fixed effect estimates have very large standard errors and so we fail to reject the null hypothesis or we cannot conclude that the two estimators are significantly different (Wooldridge (2002)).

Table 7.10: Pooled OLS, Fixed Effects and Random Effects Wage Growth Models^

	Pooled OLS		***Fixed Effects***		***Random Effects***	
	Estimate	*Standard Error*	*Estimate*	*Standard Error*	*Estimate*	*Standard Error*
Job Change	**0.0794*** **	**0.0182**	**0.1122*** **	**0.0237**	**0.0805*** **	**0.0183**
Age	-0.0110**	0.0043	-0.0211	0.0267	-0.0113**	0.0044
Age Squared	0.0001***	0.0001	0.0003	0.0002	0.0002***	0.0001
Experience	-0.0019	0.0022	0.0062	0.0210	-0.0017	0.0022
Experience Squared	0.0000	0.0000	-0.0002	0.0002	0.0000	0.0000
Female	0.0014	0.0065	-	-	0.0016	0.0066
Education: (ref: Low Education)						
Education - Medium	-0.0087	0.0073	-	-	-0.0086	0.0074
Education - High	0.0078	0.0110	-	-	0.0080	0.0111
Public Sector	0.0070	0.0098	-0.0028	0.0282	0.0073	0.0100
Number of Employees > 50	-0.0244***	0.0067	-0.0497***	0.0147	-0.0251***	0.0068
Occupation: (ref: Elementary Occupations)						
Manager	-0.0016	0.0111	-0.0076	0.0278	-0.0013	0.0113
Professional	0.0210**	0.0099	-0.0015	0.0268	0.0212**	0.0101
Clerk	0.0186**	0.0092	0.0106	0.0240	0.0190**	0.0093
Skilled	0.0125	0.0114	0.0385	0.0329	0.0128	0.0117
Sector of Origin: (ref: Non-Market Services)						
Agric., Mining & Utilities	0.0073	0.0184	0.1026*	0.0606	0.0080	0.0189
Manufacturing	-0.0108	0.0130	0.0269	0.0383	-0.0107	0.0133
Construction	0.0133	0.0181	0.0799**	0.0389	0.0137	0.0184
Market Services	0.0037	0.0097	0.0501*	0.0298	0.0039	0.0099
Year Dummies:						
1996	-0.0168	0.0167	-	-	-0.0166	0.0167
1997	0.0139	0.0155	-	-	0.0140	0.0155
1998	-0.0199	0.0146	-	-	-0.0198	0.0145
1999	-0.0069	0.0144	-	-	-0.0070	0.0144
2000	0.0120	0.0140	-	-	0.0120	0.0140
2001	-0.0072	0.0149	-	-	-0.0072	0.0149
Constant	0.2987***	0.0707	0.3859	0.4649	0.3021***	0.0716
Number of Observations	5,320		5,320		5,320	
Number of People	1,205		1,205		1,205	
R-squared within			0.0157		0.0144	
R-squared between			0.0281		0.0715	
R-squared overall			0.0156		0.0294	
R-squared	0.0294					
Prob > F	0.0000		0.0000			
Prob > chi squared					0.0000	
F test that all $\eta_i = 0$			F(1,204,4100)=0.68			
$H_0 : \sigma^2_{\eta_i} = 0$						chi-squared(1)=162.76

Notes: Standard errors are clustered by person. In the table * corresponds to 10%, ** to 5% and *** to 1% level of significance.

We expect to see different wage impacts associated with job mobility depending on the reason for job separation. Table 7.11 reports the random effects estimates of different types of mobility on wage growth.[102,103] The first model in the table does not distinguish between different types of mobility. Model 2 distinguishes between voluntary, involuntary and other types of job changes.[104] Voluntary moves have the highest effect on wage growth, as expected. The table indicates that voluntary changes are associated with a 14 per cent increase in short-term wage growth and this effect is significant at the 1 per cent level. The table also shows that involuntary moves do not have a negative impact on wage growth; in fact the estimated effect is positive, although it is insignificant and much smaller than for voluntary moves. Although the sign of estimate is not as expected it is not significant and, as discussed before, may be attributable to the construction of the sample. In addition, it could reflect the tightness in the labour market over the period under consideration where workers had many alternative employment opportunities and also employers may have been more willing to disregard any reputation effects associated with involuntary mobility. The estimated effect of 'other' types of mobility on wage growth is in-between the effects of voluntary and involuntary mobility. Model 3 excludes cases from the 'other' category and just separates out voluntary and involuntary mobility. The results are practically identical to those of Model 2.

Model 4 distinguishes between whether this is the first move a worker makes or whether they are observed changing jobs more than once during the observation window. The estimate associated with the job change being the first move observed is above that of a second or higher move but there is no evidence of wage penalties associated with repeated mobility. However, as mentioned before, it is important to note that in many cases we do not observe a workers' entire prior mobility history.

[102] The same tests were conducted to help choose between the pooled OLS, fixed effects and random effects specification. The random effects model is the preferred specification although the coefficient estimates form all three models are broadly comparable.
[103] These models only include the relevant job change variable(s) and a constant term.
[104] Other types of job changes are those where the reason for changing jobs is not reported or the respondent chooses the 'other' category from a list of possible reasons for changing jobs.

Table 7.11: Random Effects Wage Growth Models, Controlling for Type of Job Mobility^

	Model 1		*Model 2*		*Model 3*		*Model 4*	
	Estimate	*Std. Error*	*Estimate*	*Std. Error*	*Estimate*	*Std. Error*	*Estimate*	*Std. Error*
Job Change	0.1068***	0.0182						
Voluntary Job Change			0.1397***	0.0242	0.1394***	0.0242		
Involuntary Job Change			0.0335	0.0380	0.0337	0.0381		
Other type of Job Change			0.0791*	0.0479				
First Job Change							0.1153***	0.0264
Second plus Job Change							0.0973***	0.0241
Number of Observations	5,346		5,346		5,282		5,346	
Number of People	1,206		1,206		1,203		1,206	
R-squared within	0.0095		0.0115		0.0105		0.0093	
R-squared between	0.0142		0.0154		0.0176		0.0156	
R-squared overall	0.0112		0.0134		0.0129		0.0113	
Prob > chi squared	0.0000		0.0000		0.0000		0.0000	

^ Notes: * corresponds to 10%, ** to 5% and *** to 1% level of significance. Standard errors are clustered by person. Only constant terms and the job change variables reported in the table are included in the regressions. Model 1 does not distinguish between different types of mobility. Model 2 distinguishes between voluntary, involuntary and other types of mobility. Model 3 excludes other types of job changes. Model 4 distinguishes between whether a job change is the first one a worker is observed making or whether they change jobs more than once.

Next we formally examine the impact that misclassification in job changes has on the estimated effect of job change on wage growth. We use the procedure outlined in Section 7.2.4 to control for misclassification in a binary regressor. The first step involves using the Hausman *et al.* modified probit estimator to control for misclassification in a model of job change.[105] Table 7.12 shows the estimates of the misclassification probabilities. The estimated probability of misclassification for job stays, α_0, is very small at a ¼ of one per cent and the estimated probability of misclassification for job changes, α_1, is high at 51 per cent. Significance tests on α_0 and α_1 can be used as tests of misclassification. Workers who have truly changed jobs are more likely to be misclassified, as α_1 exceeds α_0. This means that the measure of job change is likely to undercount the true number of job changes. To put this estimate α_1 in context, the average mobility rate in the sample used in the wage

[105] The analysis only examines controlling for misclassification in the overall job change dummy variable.

growth regressions is around 8 per cent and this estimate for α_1 implies that the true mobility rate is around 12 per cent.

These first stage estimates are then used to construct the predicted probabilities that an observation is truly a job change. In the second step, this generated regressor is included instead of the job change dummy variable in the wage growth regression.[106,107]

Table 7.12: Effect of Job Mobility on Wage Growth Controlling for Misclassification^

	Estimate	*Standard Error*
First Stage Estimates		
$\hat{\alpha}_0$	0.0025	0.0086
$\hat{\alpha}_1$	0.5113**	0.2385
Second Stage Estimates		
Job Change	0.1372**	0.0532
Number of Observations	5,217	
R-squared	0.0238	
Prob > F	0.0000	

^ Notes: * corresponds to 10%, ** to 5% and *** to 1% level of significance. Standard errors are clustered by person. The standard errors in the second stage are also adjusted to take account of the fact that a generated regressor is included in the model. The first stage model also includes controls for experience, experience squared, gender, whether the person has children, whether they are married or living in a couple, education level, whether the person has undergone recent training, whether they report if they are overeducated, working in the public sector, firm size, occupation, sector, year dummies and the national unemployment rate. The second stage regression includes the predicted probabilities from the first stage and also the other variables in Model 2 in Table 7.9.

The results from the second step indicate that the impact of changing jobs on wage growth is closer to 14 per cent when we control for misclassification.[108] The comparable result from the model that ignores misclassification is around 8 per cent

[106] The identification of the model comes from the fact that certain variables, such as whether a person reports if they are overeducated, have children, have undergone recent training and the national unemployment rate, are included in the model in the first stage of the estimation procedure but not in the second stage and also that the predicted probabilities are non-linear functions of the explanatory variables.

[107] In addition, the standard errors are adjusted in the second stage to take account of the additional variance due to the inclusion of generated regressor as described by Newey and McFadden (1994) and Murphy and Topel (1985).

[108] The results in Table 7.12 use a pooled OLS model in the second stage. The comparable estimates using a random effects model and a fixed effects model in the second stage are 13.77 per cent and 14.84 per cent respectively.

(see Model 2 in Table 7.9). Therefore failing to control for misclassification leads us to seriously underestimate the wage effects of job mobility.

Sensitivity Analysis: Effect of Different Rates of Measurement Error on Estimates

This section illustrates the effect different rates of misclassification have on the estimates of job mobility in the wage growth regressions. It applies the formula for attenuation bias described in Section 7.2.3 and uses a range of misclassification rates to generate corrected OLS estimates. These 'corrected' estimates can be compared to the pooled estimate of 0.0794 from Model 2 in Table 7.9. Table 7.13 reports adjusted OLS estimates for different rates of misclassification. Using the first stage estimates of α_0 (i.e. the probability that a job stay is misclassified) and α_1 (the probability that a job change is misclassified) from the previous section generates an adjusted OLS estimate of around 0.10, around 30 per cent above the estimate from Model 2 in Table 7.9.

The table also shows comparable corrected OLS estimates when α_0 is assumed to be equal to zero and α_1 varies between 1 per cent and 80 per cent. The corrected estimates indicate that when α_1 is low that the adjusted estimates are quite close to the pooled OLS one. However, as α_1 increases the adjusted estimate moves increasingly further away from pooled OLS estimate. In addition, the table reports corrected OLS estimates when α_0 is 1 per cent and 5 per cent and α_1 is assumed to be equal to zero. Even for these relatively low rates of misclassification, the adjusted OLS estimates are quite far away from the pooled OLS estimate. This stronger impact from misclassifying job stays arises from the fact that the proportion of observed job changes in the sample is around 8 per cent, so the proportion of job stays is 92 per cent and therefore the misclassification rate applies to a much higher number of cases.

Table 7.13: Adjusted Pooled OLS Estimates for a Range of Misclassification Rates

	Estimate	***Standard Error***
Reference:		
Pooled OLS estimate (from Model 2, Table 7.9)	0.0794***	0.0182
Corrected OLS Estimates:		
Using α_0=0.0025 and α_1=0.5113 (from Table 7.12)	0.1008	
Varying α_1 (assume α_0=0):		
α_1=0.01	0.0795	
α_1=0.05	0.0802	
α_1=0.10	0.0811	
α_1=0.20	0.0832	
α_1=0.30	0.0860	
α_1=0.40	0.0902	
α_1=0.50	0.0968	
α_1=0.60	0.1086	
α_1=0.70	0.1364	
α_1=0.80	0.2798	
Varying α_0 (assume α_1=0):		
α_0=0.01	0.0903	
α_0=0.05	0.2131	

Besides measurement error and unobserved heterogeneity, an additional difficulty with investigating the effect of job mobility on wage growth is the possible endogeneity of job mobility. In particular, there is a potential issue with reverse causality; wage growth is affected by job changes but also job changes may occur in expectation of higher wage growth. This can be addressed using an instrumental variable approach. We need instruments that are highly correlated with job mobility and that are uncorrelated with wage growth, so they have no independent effect on wage growth other than through job mobility. The approach taken is to use non-wage aspects of job satisfaction as instruments for job mobility.

Kristensen and Westergård-Nielsen (2004) argue that job satisfaction may be a proxy for the worker's assessment of the quality of the match. Job satisfaction may capture unobserved aspects of work, such as the organisation of work, harsh working conditions etc. As such, job satisfaction should be a strong predictor of job mobility. However, the difficulty with using a measure of overall job satisfaction as an instrument for job mobility is that we expect a worker's satisfaction with earnings to dominate such as measure. Therefore it is likely that overall job satisfaction is correlated with wage growth. However, the LIS asks workers how satisfied they are with different aspects of their job, where satisfaction with earnings is only one component. Nonetheless, it may still be the case that satisfaction with earnings influences a worker's assessment of their satisfaction with other aspects of the job and this should be borne in mind when interpreting the results.

Table 7.14 shows the percentage of workers who are satisfied with various aspects of their jobs.[109] From the table, dissatisfaction with earnings is the most common source of dissatisfaction with the job.

Table 7.14: Satisfaction with Various Aspects of Job

	% Satisfied	*% Not Satisfied*
Satisfied with:		
Earnings	67%	33%
Job Security	82%	18%
Type of Work	90%	10%
Number of Hours	83%	17%
Distance to Job/Commuting	88%	12%
Working Times (i.e. daytime, night-time, shifts etc.)	88%	12%
Working conditions/environment in place of work	88%	12%

To assess whether satisfaction with wages affects satisfaction with other aspects of the job, Table 7.15 reports the percentage of workers satisfied with other areas of their jobs of those who are not satisfied with their earnings. The table shows that high proportions of workers are satisfied with other areas of their jobs even though they are unhappy with their earnings. This indicates that (dis)satisfaction with earnings may not influence satisfaction with other areas of the job. Therefore, the non-wage aspects of job satisfaction may be appropriate instruments for job mobility.

[109] Workers are asked to indicate their degree of satisfaction with each area of their work on a scale of 1 to 6, where 1 indicates that they are not satisfied at all and 6 indicates that they are fully satisfied. In the table, satisfied corresponds to workers who report a level or 4 or above and not satisfied refers to those who report a satisfaction level of 3 or below.

Table 7.15: Satisfaction with Various Aspects of Job for those not Satisfied with Earnings

	% Satisfied	*% Not Satisfied*
Satisfied with other aspects of job, if not satisfied with earnings:		
Job Security	66%	34%
Type of Work	80%	20%
Number of Hours	70%	30%
Distance to Job/Commuting	83%	17%
Working Times (i.e. daytime, night-time, shifts etc.)	81%	19%
Working conditions/environment in place of work	78%	22%

The quality of the instruments can be checked by testing their significance in the first stage of the two-step approach. The results from the Hausman *et al.* modified probit estimator that includes all the exogenous variables and all the instruments show that satisfaction with distance to job and working conditions are not significant and the coefficient on satisfaction with working hours has the incorrect sign and is only significant at the 10 per cent level. Consequently, these three variables are dropped from the analysis. Satisfaction with job security, type of work and working times are used as instruments for job mobility in the first stage. Using the Hausman *et al.* estimator in the first step, we can calculate the probability of truly being a job changer, $\left(\Pr\left(m_{it}^{*}=1\right)\right)$. In the second step, the wage growth equation is estimated using pooled OLS substituting in for m_{it}^{*} using the fitted probabilities from the first stage. This controls for both misclassification and endogeneity.

As discussed in Section 7.2.3, we expect the IV estimates that don't control for misclassification to be biased upwards. Table 7.16 shows the results from the two-step procedure using IV but where misclassification is ignored i.e. the predicted probabilities in the first stage come from a standard probit model. The estimate on the job change dummy variable indicates that the increase in wage growth associated with changing jobs is around 26 per cent. This compares to the pooled OLS estimate of 8 per cent (see Model 2, Table 7.9). As expected, the IV estimate is above the OLS one, but it is dramatically higher and arguably implausibly large.[110]

[110] This type of estimate is consistent with what Davia (2005) finds when she controls for endogeneity in job mobility using ECHP data. For most of the countries in her analysis, the estimates that control for endogeneity are multiples of the pooled OLS estimates (see Davia (2005), Table 2, page 24).

Table 7.16: Second Stage IV Estimates of Job Mobility on Wage Growth^

	Estimate	*Standard Error*
Job Change	0.2590***	0.0706
Number of Observations	4,428	
R-squared	0.0275	
Prob > F	0.0000	

^ Notes: * corresponds to 10%, ** to 5% and *** to 1% level of significance. Standard errors are clustered by person. The standard errors in the second stage are also adjusted to take account of the fact that a generated regressor is included in the model. The second stage model includes the same controls as Model 2 in Table 7.9. The first stage model includes the same controls as the first stage model in Table 7.12 as well as the three instruments

Table 7.17 shows the results from the two-step approach controlling for both endogeneity and misclassification. The estimates of the misclassification probabilities from the first stage are practically identical to the estimates in Table 7.12. The second stage IV estimate implies that the impact of changing jobs on wage growth is around 13 per cent, when we control for misclassification. This is around half the IV estimate that ignores misclassification, implying that ignoring misclassification leads to a significant upwards bias in the IV estimate. In addition, the estimate is around 1.6 times the size of the pooled OLS estimate but quite similar to the estimate that controls for misclassification but ignores the possible reverse causality of job mobility.

Table 7.17: IV Estimates of Job Mobility on Wage Growth, Controlling for Misclassification^

	Estimate	*Standard Error*
<u>*First Stage Estimates*</u>		
$\hat{\alpha}_0$	0.0072**	0.0030
$\hat{\alpha}_1$	0.5107***	0.1337
<u>*Second Stage Estimates*</u>		
Job Change	0.1242***	0.0348
Number of Observations	4,428	
R-squared	0.0274	
Prob > F	0.0000	

^ Notes: * corresponds to 10%, ** to 5% and *** to 1% level of significance. Standard errors are clustered by person. The standard errors in the second stage are also adjusted to take account of the fact that a generated regressor is included in the model. The second stage model includes the same controls as Model 2 in Table 7.9. The first stage model includes the same controls as the first stage model in Table 7.12 as well as the three instruments.

Finally, Table 7.18 provides a summary of the various estimates of job mobility on wage growth.

Table 7.18: Summary of Estimates of Job Mobility on Wage Growth

	Estimate	*Standard Error*
Pooled OLS (See Model 2, Table 7.9)	0.0794***	0.0182
Random Effects (See Table 7.10)	0.0805***	0.0183
Controlling for Misclassification (see Table 7.12)	0.1372**	0.0532
IV (see Table 7.16)	0.2590***	0.0706
IV & Controlling for Misclassification (see Table 7.17)	0.1242***	0.0348

7.5 Conclusions

This chapter adds to the literature on the effect of job mobility on wage growth. The chapter finds OLS estimates of the effect of job mobility on wage growth of around 8 per cent. The chapter also finds that wage effects differ depending on the reason for job separation, as expected. Voluntary job changes are associated with a 14 per cent increase in wage growth. However, there is no evidence of wage penalties associated with involuntary mobility. This may be attributable to the fact that the sample considered is one where workers have a very high attachment to the labour force or it may be due to the very high growth rates and tightness in the labour market over the time period under consideration.

The chapter argues that the OLS estimate of the effect of changing jobs on wage growth may be biased due to unobserved heterogeneity, reverse causality and also because of measurement error which is the main concern of the chapter. The chapter finds that the effect of job mobility on wage growth persists even after controlling for unobserved individual heterogeneity. The magnitude of the estimates obtained from OLS regressions and regressions that control for unobserved heterogeneity are broadly in line with the existing empirical literature (see Chapter 2). For example, OECD (2010) finds a wage premium associated with changing jobs of around 9 per cent for Ireland which is very similar to what is found in this chapter. However, these estimates ignore measurement error in job changes.

This chapter adopts a two-step approach to controlling for misclassification in a binary explanatory variable. It finds that controlling for misclassification has a substantial effect on the estimated impact changing jobs has on wage growth. The effect of job mobility on wage growth is estimated to be closer to 14 per cent when measurement error is controlled for. Finally, controlling for reverse causality using an instrumental variables approach and ignoring misclassification produces an estimate that seems questionably high; however a more plausible estimate is obtained when the IV strategy is combined with the measurement error approach.

Appendix 7A: Misclassification Bias in Multivariate Regression

Recall from (7.1) the true model is:

$$\Delta \log(w_{it}) = \beta m_{it}^{*} + \delta x_{it} + \varepsilon_{it} \tag{7A.1}$$

However, we do not observe the binary variable m_{it}^{*}, only the mismeasured proxy m_{it}, such that:

$$m_{it} = m_{it}^{*} + u_{it} \tag{7A.2}$$

Therefore:

$$\begin{aligned}\Delta \log(w_{it}) &= \beta m_{it} - \beta u_{it} + \delta x_{it} + \varepsilon_{it} \\ &= \beta m_{it} + \delta x_{it} + (\varepsilon_{it} - \beta u_{it})\end{aligned} \tag{7A.3}$$

The measurement error becomes part of the error term in the regression equation and creates an endogeneity bias. To assess the size of the bias consider the probability limit of OLS estimator for β in the two variable case:

$$\text{plim}\hat{\beta}_{OLS} = \frac{Var(x_{it})Cov(\Delta \log(w_{it}), m_{it}) - Cov(x_{it}, m_{it})Cov(\Delta \log(w_{it}), x_{it})}{Var(m_{it})Var(x_{it}) - Cov(x_{it}, m_{it})^2} \tag{7A.4}$$

$$= \frac{Var(x_{it})\left[Cov(\beta m_{it}^{*} + \delta x_{it} + \varepsilon_{it}, m_{it})\right] - Cov(x_{it}, m_{it})\left[Cov(\beta m_{it}^{*} + \delta x_{it} + \varepsilon_{it}, x_{it})\right]}{Var(m_{it})Var(x_{it}) - Cov(x_{it}, m_{it})^2}$$

$$= \frac{Var(x_{it})\left[\beta Cov(m_{it}^{*}, m_{it}) + \delta Cov(x_{it}, m_{it})\right] - Cov(x_{it}, m_{it})\left[\beta Cov(m_{it}^{*}, x_{it}) + \delta Var(x_{it})\right]}{Var(m_{it})Var(x_{it}) - Cov(x_{it}, m_{it})^2}$$

$$= \frac{\left[\dfrac{\beta Cov(m_{it}^{*}, m_{it})Var(m_{it})Var(x_{it})}{Var(m_{it})} + \delta Cov(x_{it}, m_{it})Var(x_{it})\right] - Cov(x_{it}, m_{it})\left[\dfrac{\beta Cov(m_{it} x_{it})}{(1 - \alpha_0 - \alpha_1)} + \delta Var(x_{it})\right]}{Var(m_{it})Var(x_{it}) - Cov(x_{it}, m_{it})^2}$$

(A linear projection of m_{it} on m_{it}^* is: $m_{it} = \alpha_0 + (1-\alpha_0-\alpha_1)m_{it}^* + \eta_{it}$, where η_{it} is uncorrelated with m_{it}^*, so $Cov(m_{it}, x_{it}) = (1-\alpha_0-\alpha_1)Cov(m_{it}^*, x_{it})$)

$$= \frac{\beta\left[\gamma^0 Var(m_{it})Var(x_{it})\right] - \left[\frac{\beta Cov(x_{it}, m_{it})^2}{(1-\alpha_0-\alpha_1)}\right]}{Var(m_{it})Var(x_{it}) - Cov(x_{it}, m_{it})^2}$$

(where $\frac{Cov(m_{it}^*, m_{it})}{Var(m_{it})} = \gamma^0 = \frac{\mu/p\,(1-\alpha_1-p)}{(1-p)}$ from (7.6))

$$= \frac{\beta\left[\gamma^0 - R^2/(1-\alpha_0-\alpha_1)\right]}{1-R^2}$$

(dividing above and below by $Var(x_{it})Var(m_{it})$ and where $R^2 = \frac{Cov(x_{it}, m_{it})^2}{Var(x_{it})Var(m_{it})}$)

$$= \beta\gamma \qquad \text{where } \gamma = \frac{\gamma^0 - \left(R^2/(1-\alpha_0-\alpha_1)\right)}{1-R^2}$$

Appendix 7B: Bias in Instrumental Variable Estimates in a Model with a Single Binary Regressor

Suppose a binary variable, z_{it}, that is highly correlated with m_{it}^* is available as an instrument. As discussed in Section 7.2.3 the measurement error, u_{it}, is correlated with m_{it}^*, so any instrument which is correlated with m_{it}^* will also generally be correlated with the measurement error. When there are no other covariates in the model the IV estimator is given by the Wald estimator so that:

$$\text{plim}\hat{\beta}_{IV} = \frac{E(\Delta \log(w_{it})|z_{it}=1) - E(\Delta \log(w_{it})|z_{it}=0)}{E(m_{it}|z_{it}=1) - E(m_{it}|z_{it}=0)} \tag{7B.1}$$

$$= \frac{\beta\left[\Pr(m_{it}^*=1|z_{it}=1) - \Pr(m_{it}^*=1|z_{it}=0)\right]}{E(m_{it}|z_{it}=1) - E(m_{it}|z_{it}=0)} \tag{7B.2}$$

Expanding the individual terms in (7B.2):

$$\Pr(m_{it}^*=1|z_{it}=1) = \frac{\Pr(z_{it}=1|m_{it}^*=1)\Pr(m_{it}^*=1)}{\Pr(z_{it}=1)} \qquad \text{(by Bayes' Rule)}$$

$$= \frac{\Pr(z_{it}=1|m_{it}^*=1)\Pr(m_{it}^*=1)}{\Pr(z_{it}=1|m_{it}^*=1)\Pr(m_{it}^*=1) + \Pr(z_{it}=1|m_{it}^*=0)\Pr(m_{it}^*=0)}$$

(using the law of iterated expectations)

$$= \frac{a\mu}{a\mu + b(1-\mu)} \tag{7B.3}$$

(where $a = \Pr(z_{it}=1|m_{it}^*=1)$ and $b = \Pr(z_{it}=1|m_{it}^*=0)$)

Similarly:

$$\Pr\left(m_{it}^* = 1 \middle| z_{it} = 0\right) = \frac{(1-a)\mu}{(1-a)\mu + (1-b)(1-\mu)} \tag{7B.4}$$

And:

$$E\left(m_{it} \middle| z_{it} = 1\right) = \Pr\left(m_{it} = 1 \middle| z_{it} = 1\right) = \frac{\Pr\left(m_{it} = 1, z_{it} = 1\right)}{\Pr\left(z_{it} = 1\right)}$$

$$= \frac{\Pr\left(m_{it} = 1, z_{it} = 1 \middle| m_{it}^* = 1\right)\Pr\left(m_{it}^* = 1\right) + \Pr\left(m_{it} = 1, z_{it} = 1 \middle| m_{it}^* = 0\right)\Pr\left(m_{it}^* = 0\right)}{\Pr\left(z_{it} = 1 \middle| m_{it}^* = 1\right)\Pr\left(m_{it}^* = 1\right) + \Pr\left(z_{it} = 1 \middle| m_{it}^* = 0\right)\Pr\left(m_{it}^* = 0\right)}$$

$$= \frac{(1-\alpha_1)a\mu + \alpha_0 b(1-\mu)}{a\mu + b(1-\mu)} \tag{7B.5}$$

Similarly:

$$E\left(m_{it} \middle| z_{it} = 0\right) = \frac{\mu(1-\alpha_1)(1-a) + \alpha_0(1-b)(1-\mu)}{(1-a)\mu + (1-b)(1-\mu)} \tag{7B.6}$$

Substituting (7B.3), (7B.4), (7B.5) and (7B.6) into (7B.2) yields:

$$\operatorname{plim}\hat{\beta}_{IV} = \frac{\beta\left[\frac{a\mu}{c} - \frac{(1-a)\mu}{(1-c)}\right]}{\frac{(1-\alpha_1)a\mu + \alpha_0 b(1-\mu)}{c} - \frac{\mu(1-\alpha_1)(1-a) + \alpha_0(1-b)(1-\mu)}{(1-c)}} \tag{7B.7}$$

(where $c = a\mu + b(1-\mu)$)

$$= \frac{\beta\left[\frac{a\mu(1-c) - (1-a)\mu c}{c(1-c)}\right]}{\frac{(1-\alpha_1)a\mu(1-c) + \alpha_0 b(1-\mu)(1-c) - (1-\alpha_1)(1-a)\mu c - \alpha_0(1-b)(1-\mu)c}{c(1-c)}}$$

Simplifying:

$$= \frac{\beta\left[a\mu - \mu c\right]}{(1-\alpha_1)a\mu + \alpha_0 b - \alpha_0 b\mu - (1-\alpha_1)\mu c - \alpha_0 c + \alpha_0 \mu c} \tag{7B.8}$$

Substituting for c in (7B.8) yields:

$$= \frac{\beta\left[a\mu - a\mu^2 - b\mu(1-\mu)\right]}{(1-\alpha_1)a\mu + \alpha_0 b - \alpha_0 b\mu - (1-\alpha_1)a\mu^2 - (1-\alpha_1)b\mu(1-\mu) - \alpha_0 a\mu - \alpha_0 b(1-\mu) + \alpha_0 a\mu^2 + \alpha_0 b\mu(1-\mu)}$$

$$= \frac{\beta\left[a\mu - a\mu^2 - b\mu(1-\mu)\right]}{(1-\alpha_0-\alpha_1)a\mu - (1-\alpha_0-\alpha_1)a\mu^2 - (1-\alpha_0-\alpha_1)b\mu(1-\mu)}$$

Therefore:

$$\text{plim}\hat{\beta}_{IV} = \frac{\beta}{(1-\alpha_0-\alpha_1)}$$

Appendix 7C: Consistency of Estimator in Two-Step Approach

The two-step approach outlined in Section 7.2.4 can be viewed as a two-step GMM procedure. Brachet (2008) argues that the estimator in the two-step approach is consistent provided the conditional distribution of truly being a job changer in the first stage has been correctly specified.

Brachet (2008) demonstrates the consistency of the estimator by using a model with a single binary regressor m_{it}^* which is assumed to be correctly measured but correlated with the error term, as follows:

$$\Delta \log(w_{it}) = \beta m_{it}^* + \varepsilon_{it} \tag{7C.1}$$

The first step of the two-step approach comes from the Hausman *et al.* procedure and involves estimating a nonlinear model which has $E(m_{it}^* | x_{it}) = \Phi(\delta x_{it})$ where $\Phi(.)$ denotes the cdf of the standard normal distribution. As the chapter assumes that the error terms are normally distributed, a modified probit model is estimated in the first step. The second step substitutes the estimate of $F(\delta x_{it})$ for the variable m_{it}^* for in (7C.1) above and the resulting regression model uses the following moment condition:

$$E\{\Phi(\delta x_{it}) \times [\Delta \log(w_{it}) - \beta \Phi(\delta x_{it})]\} = 0 \tag{7C.2}$$

(7C.2) only holds if $E(m_{it}^* | x_{it}) = \Phi(\delta x_{it})$ i.e. that the standard normal distribution is the correct choice in the first stage. This can be shown as follows:

$$\begin{aligned} E\{\Phi(\delta x_{it}) \times [\Delta \log(w_{it}) - \beta \Phi(\delta x_{it})]\} &= E\{\Phi(\delta x_{it}) \times [E(\Delta \log(w_{it}) | x_{it}) - \beta \Phi(\delta x_{it})]\} \\ &= E\{\Phi(\delta x_{it}) \times [\beta E(m_{it}^* | x_{it}) - \beta \Phi(\delta x_{it})]\} \\ &\qquad \text{(assuming } E(\varepsilon_{it} | x_{it}) = 0\text{)} \\ &= 0 \text{ if } E(m_{it}^* | x_{it}) = \Phi(\delta x_{it}) \end{aligned}$$

References

Abbott, M. and C. Beach (1994), “Wage Changes and Job Changes of Canadian Women”, *Journal of Human Resources*, Vol. 29, No. 2, pp. 429-460.

Aigner, D.J. (1973), “Regression with a Binary Independent Variable Subject to Errors of Observation”, *Journal of Econometrics,* Vol. 1, pp. 49-60.

Altonji, J. and R. Shakotko (1987), “Do Wages Rise with Job Seniority”, *The Review of Economic Studies*, Vol. 54, No. 3, (Jul., 1987), pp. 437-459.

Angrist, J.D. and A. B. Krueger (1999), “Empirical Strategies in Labor Economics”, in *Handbook of Labor Economics*, Volume 3A, O. Ashenfelter and D. Card (eds), Amsterdam and New York: Elsevier Science.

Artís, M., Ayuso, M. and M. Guillén (2002), “Detection of Automobile Insurance Fraud with Discrete Choice Models and Misclassified Claims”, *The Journal of Risk and Insurance,* Vol. 69, No. 3, pp. 325-340.

Bachmann, R., Bauer, T. and P. David (2010), “Labour Market Entry Conditions, Wages and Job Mobility”, IZA Discussion Paper No. 4965.

Baker, M. and G. Solon (1999), *“Earnings Dynamics and Inequality among Canadian Men, 1976-1992: Evidence from Longitudinal Income Tax Records”,* NBER Working Paper No. 7370.

Barrett, A., FitzGerald, J. and B. Nolan (2002), “Earnings Inequality, Returns to Education and Immigration into Ireland", *Labour Economics*, Vol. 9, Issue 5, (Nov., 2002) pp. 665-680.

Barron, J. M., Black, D.A. and M.A. Loewenstein (1993), “Gender Differences in Training, Capital and Wages”, *The Journal of Human Resources,* Vol. 28, No. 2. (Spring, 1993), pp. 343-364.

Bartel, A. and G. Borjas (1981), "Wage Growth and Job Turnover: An Empirical Analysis", in *Studies in Labor Markets*, Vol. 3, Rosen S. (ed.), pp. 65-84. University of Chicago Press.

Bartel, A. and F. Lichtenberg (1987), "The Comparative Advantage of Educated Workers in Implementing New Technology", *The Review of Economics and Statistics*, Vol. 69, No. 1. (Feb., 1987), pp. 1-11.

Bassanini, A., Garnero, A., Marianna, P. and S. Martin (2010), "Institutional Determinants of Worker Flows: A Cross-Country/Cross-Industry Approach", OECD Social, Employment, and Migration Papers, No. 107.

Baum-Snow, N., and D. Neal (2009), "Mismeasurement of usual hours worked in the Census and ACS", *Economic Letters,* Vol. 102, Issue 1, (Jan., 2009), pp. 39-41.

Becker, G. (1962), "Investment in Human Capital: A Theoretical Analysis", *Journal of Political Economy*, Vol. 70 (Oct. 1962), pp. 9-49.

Blau, F. and L Kahn (1981), "Race and Sex Differences in Quits by Young Workers", *Industrial and Labor Relations Review,* Vol. 34, No. 4, pp. 563-577.

Blumen, I. Kogen, M. and P. McCarthy (1955), "*The Industrial Mobility of Labor as a Probability Process"*, Ithaca, New York: Cornell University Press.

Bollinger, C.R. (1996), "Bounding Mean Regressions When a Binary Regressor is Mismeasured" *Journal of Econometrics*, Vol. 73, pp. 387-399.

Bollinger, C.R. (2001), "Response Error and the Union Wage Differential", *Southern Economic Journal*, Vol. 68, No. 1, pp. 60-76.

Booth, A. and M. Franscesconi (2000), "Job Mobility in 1990s Britain: Does Gender Matter?", *Research in Labor Economics*, Vol. 19, pp. 173-189.

Booth, A., Francesconi, M. and C. Garcia-Serrano (1999), "Job Tenure and Job Mobility in Britain", *Industrial and Labor Relations Review*, Vol. 53, No. 1, (Oct., 1999), pp. 43-70.

Borghans, L. and B. Golsteyn (2011), "Job Mobility in Europe, Japan and the United States", *British Journal of Industrial Relations,* Article in press.

Borjas, G. and S. Rosen (1980), "Income Prospects and Job Mobility of Younger Men" in *Research in Labor Economics*, Vol. 3, Ehrenberg R.G. (ed), pp. 159-181.

Borsch-Supan, A. (1987), "*The Role of Education: Mobility Increasing or Mobility Impeding*?", NBER Working Paper No. 2329.

Bound, J., Brown, C. and N. Mathiowetz (2001), "Measurement Error in Survey Data", in J. Heckman and E. Leamer, eds, *Handbook of Econometrics*, Vol. 5. Elsevier, North Holland.

Brachet, T. (2008), "Maternal Smoking, Misclassification and Infant Health", MPRA Working Paper No. 21466. Available at: http://mpra.ub.uni-muenchen.de/21466/

Brown, J. and A. Light (1992), "Interpreting Panel Data on Job Tenure", *Journal of Labor Economics*, Vol. 10, No. 3 (Jul., 1992), pp. 219-257.

Burdett, K. (1978), "A Theory of Employee Job Search and Quit Rates", *American Economic Review,* Vol. 68, No. 1 (Mar., 1978), pp. 212-220.

Burdett, K. and M. Coles (2003), "Equilibrium Wage-Tenure Contracts", *Econometrica*, Vol. 71, No. 5 (Sep., 2003), pp. 1377-1404.

Burdett, K. and D. Mortensen (1998), "Wage Differentials, Employer Size and Unemployment", *International Economic Review*, Vol. 39, No. 2, (May, 1998), pp. 257-273.

Burgess, S., Lane, J., and D. Stevens (2000), "Job Flows, Worker Flows, and Churning", *Journal of Labor Economics,* Vol. 18, No. 3, pp. 473-502

Campbell, C.M. (1997), "The Determinants of Dismissals, Quits and Layoffs: A Multinomial Logit Approach", *Southern Economic Journal,* Vol. 63, No.4, pp. 1066-1073.

Campbell, D. (2001), "Estimating the Wage Effects of Job Mobility in Britain" Studies in Economics, No. 0117, Department of Economics, University of Kent.

Caparros Ruiz, A., Navarro Gomez, M.L., and M.F. Rueda Narvaez (2004), "Gender wage gaps and job mobility in Spain", *International Journal of Manpower*, Vol. 25, No. 3/4, pp. 264-278.

Card, D. (1996), "The Effect of Unions on the Structure of Wages: A Longitudinal Analysis", *Econometrica*, Vol. 64, No. 4, pp. 957-979.

Caudill, A. and F. Mixon (2005), "Analysing Misleading Discrete Responses: A Logit Model Based on Misclassified Data", *Oxford Bulletin of Economics and Statistics,* Vol. 67, No. 1, pp. 105-113.

Connolly, H. and P. Gottschalk (2006) "Differences in Wage Growth by Education Level: Do Less-Educated Workers Gain Less from Work Experience?", IZA Discussion Paper No. 2331.

Cornelissen, T., Hubler, O., and S. Schneck (2007), "Cyclical Effects on Job-to-Job Mobility: An Aggregated Analysis on Microeconomic Data", University of Hannover, Discussion Paper 371.

Davia, M. (2005), "Job Mobility and wage mobility at the beginning of the working career: a comparative view across Europe", ISER Working Paper No. 2005-03, Institute for Social and Economic Research.

Doris, A. (2001), "Labour Supply Elasticity Changes during the 1990s", *Quarterly Economic Commentary,* December.

Duncan, G.J. and D.H. Hill (1985), “An Investigation of the Extent and Consequences of Measurement Error in Labor-Economic Survey Data”, *Journal of Labor Economics*, Vol. 3, No. 4, (Oct., 1985), pp. 508-532.

Dustmann, C. and C. Meghir (2005), “Wages, Experience and Seniority”, *The Review of Economic Studies*, Vol. 72, Issue 1, (Jan., 2005), pp. 77-108.

Dustmann, C. and S. Pereira (2008), “Wage Growth and Job Mobility in the United Kingdom and Germany”, *Industrial and Labor Relations Review,* Vol. 61, No. 3, (Apr., 2008), pp. 374-393.

Dustmann, C. and A. van Soest (2004), “An Analysis of Speaking Fluency of Immigrants Using Ordered Response Models with Classification Errors”, *Journal of Business and Economic Statistics,* Vol. 22, No. 3, pp. 312-321.

Dustmann, C. and A. van Soest (2001), “Language Fluency and Earnings: Estimation with Misclassified Language Indicators”, *Review of Economics and Statistics*, Vol. 83, No. 4, pp. 663-674.

Dye, R. and D. McMillen (2007), “Teardowns and land values in the Chicago metropolitan area”, *Journal of Urban Economics,* Vol. 61, pp. 45-63.

Fairlie, R. (2005), “An extension of the Blinder-Oaxaca decomposition technique to logit and probit models”, *Journal of Economic and Social Measurement*, Vol. 30, pp. 305-316.

Farber, H. (1994), “The Analysis of Interfirm Worker Mobility”, *Journal of Labor Economics*, Vol. 12, No. 4, (Oct., 1994), pp. 554-593.

Farber, H. (1999), “Mobility and Stability: The Dynamics of Job Change in Labor Markets”, in O. Ashenfelter and D. Card, eds, *Handbook of Labor Economics, Vol.3.* Elsevier, North Holland.

Flathmann, J. and S. Sheffrin (2003), *“Are Surveys of Taxpayers’ Honesty Honest?”*, IIPF Congress Proceedings, Prague, August 2003.

Frazis, H. and M.A. Loewenstein (2003), “Estimating Linear Regressions with Mismeasured, Possibly Endogenous, Binary Explanatory Variables”, *Journal of Econometrics*, Vol. 117, No. 1, pp. 151-178.

Freeman, R.B. (1984), “Longitudinal Analysis of the Effects of Trade Unions”, *Journal of Labor Economics*, Vol. 2, No. 1, pp. 1-26.

Garcia Perez, J.I. and Y. Rebollo Sanz (2005), “Wage changes through job mobility in Europe: A multinomial endogenous switching approach”, *Labour Economics,* Vol. 12, pp. 531-555.

Gibbons, R. and L. Katz (1991), “Layoffs and Lemons”, *Journal of Labor Economics,* Vol. 9, No. 4, (Oct., 1991), pp. 351-380.

Greenwald, B. (1986) “Adverse Selection in the Labour Market”, *The Review of Economic Studies,* Vol. 53, No. 3, (Jul., 1986), pp. 325-347.

Greenwood, M. (1975), “Research on Internal Migration in the United States: A Survey”, *Journal of Economic Literature,* Vol. 13, No. 2. (Jun., 1975), pp. 397-433.

Groot, W. and M. Verberne (1997), “Aging, Job Mobility and Compensation”, *Oxford Economic Papers*, Vol. 49, No.3, (Jul., 1997), pp. 380-403.

Hausman, J., Abrevaya, J. and F.M. Scott-Morton (1998), “Misclassification of the dependent variable in a discrete-response setting”, *Journal of Econometrics*, Vol. 87, pp. 239-269.

Hausman, J. (2001), "Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left", *The Journal of Economic Perspectives,* Vol. 15, No. 4 (Aut., 2001), pp. 57-67.

Hersch, J., and L. Stratton (1997), "Housework, Fixed Effects, and Wages of Married Workers", *Journal of Human Resources,* Vol. 32, No. 2, pp. 285-307.

Holmlund, B. (1984), "Job Mobility and Wage Growth: A Study of Selection Rules and Rewards, in *Studies of Labor Market Dynamics*, Neemann G. and N. Westgard-Neilson (eds), pp. 238-260, Berlin: Springer-Verlag.

Ibsen, R., Trevisan, E. and N. Westergaard-Nielsen (2008), "Job Mobility and Skill Transferability: Some Evidence from Denmark and a Large Italian Region", Working Paper No. 40 (2008), Department of Economics, Ca'Foscari University of Venice.

Jann, B. (2006), "Fairlie: Stata module to generate nonlinear decomposition of binary outcome differentials", available at http://ideas.repec.org/c/boc/bocode/s456727.html.

Jensen, P., Palangkaraya, A. and E. Webster (2008), "*Misclassification in Patent Offices"*, Intellectual Property Research Institute of Australia, Melbourne Institute of Applied Economic and Social Research, University of Melbourne, Working Paper No. 02/08, May 2008.

Johnson, W. (1978), "A Theory of Job Shopping"*, The Quarterly Journal of Economics,* Vol. 92, No. 2, (May, 1978), pp. 261-278.

Johnson, W. (1979) "The Demand for General and Specific Education with Occupational Mobility" *The Review for Economic Studies*, Vol. 46, No. 4, (Oct.1979), pp. 695-705.

Jovanovic, B. (1979a), "Firm-specific capital and Turnover", *The Journal of Political Economy*, Vol. 87, No. 6, (Dec., 1979), pp. 1246-1260.

Jovanovic, B. (1979b), "Job Matching and the Theory of Turnover", *The Journal of Political Economy*, Vol. 87, No. 5, Part 1, (Oct., 1979), pp. 972-990.

Kahn, S. and H. Griesinger (1989), "Female Mobility and the Returns to Seniority: Should EEO Policy Be Concerned with Promotion?" *The American Economic Review,* Vol. 79, No. 2, (May, 1989), pp. 300-304.

Kane, T., C.E. Rouse and D.O. Staiger (1999), "Estimating Returns to Schooling when Schooling is Misreported", National Bureau of Economic Research Working Paper No. 7235.

Keith, K. (1993), "Reputation, Voluntary Mobility and Wages", *The Review of Economics and Statistics,* Vol. 75, No. 3, (Aug. 1993), pp. 559-663.

Keith, K. and A. McWilliams (1999), "The Returns to Mobility and Job Search by Gender", *Industrial and Labor Relations Review*, Vol. 52, No. 3, (Apr., 1999), pp. 460-477.

Keith, K. and A. McWilliams (1997), "Job Mobility and Gender-Based Wage Growth Differentials", *Economic Inquiry,* Vol. 35, pp. 320-333.

Kenkel, D., Lillard, D. and A. Mathios (2004), "Accounting for misclassification error in retrospective smoking data", *Health Economics*, Vol. 13, pp. 1031-1044.

Kidd, M. (1991), "An Econometric Analysis of Interfirm Labour Mobility", *Canadian Journal of Economics*, Vol. 24, No. 3, pp. 517-535.

Kristensen, N. and N. Westergård-Nielsen (2004), "Does Low Job Satisfaction Lead to Job Mobility?, IZA Discussion Paper No. 1026.

Lazear, E. (1986), "Raids and Offer Matching", *Research in Labor Economics*, Vol. 8, part A, Ehrenberg, R.G. (ed), pp. 141-165.

Leece, D. (2000), "Household Choice of Fixed versus Floating Rate Debt: A Binomial Probit Model with Correction for Classification Error", *Oxford Bulletin of Economics and Statistics,* Vol. 62, No. 1, pp. 61-82.

Le Grand, C. and M Tahlin (2002), "Job mobility and earnings growth", *European Sociological Review*, Vol.18, No. 4, pp. 381-400.

Light, A. and K. McGarry (1998), "Job Change Patterns and the Wages of Young Men", *The Review of Economics and Statistics,* Vol. 80, No. 2, pp. 276-286.

Loprest, P. (1992), "Gender Differences in Wage Growth and Job Mobility", *The American Economic Review*, Vol. 82, No. 2, (May, 1992), pp. 526-532.

Mahajan, A. (2006), "Identification and Estimation of Regression Models with Misclassification", *Econometrica*, Vol. 74, No. 3, (May, 2006), pp. 631-665.

Marshall, R. and G. Zarkin (1987), "The Effect of Job Tenure on Wage Offers", *Journal of Labor Economics*, Vol. 5, No. 3, pp. 301-324.

Mincer, J. (1986), "Wage Changes in Job Changes", in *Research in Labor Markets,* Rosen S. (ed.), NBER and University of Chicago Press, pp. 21-63.

Mincer, J. and B. Jovanovic (1981), "Labor Mobility and Wages", in *Studies in Labor Markets*, Vol. 3, Rosen S. (ed.), pp. 21-63. University of Chicago Press.

Mortensen, D. (1978), "Specific Capital and Labor Turnover", *The Bell Journal of Economics*, Vol. 9, No. 2, (Autumn, 1978), pp. 572-586.

Mortensen, D. (1988), "Wages, Separations, and Job Tenure: On-the-Job Specific Training or Matching?", *Journal of Labor Economics,* Vol. 6, No. 4. (Oct., 1998), pp. 445-471.

Munasinghe, L. and K. Sigman (2004), "A Hobo Syndrome? Mobility, Wages and Job Turnover", *Labour Economics,* Vol. 11, pp. 191-218.

Murphy, K. and R. Topel (1985), “Estimation and Inference in Two-Step Econometric Models”, *Journal of Business and Economic Statistics,* Vol. 3, No. 4, (Oct., 1985), pp. 370-379.

Naticchioni, P. and D. Panigo (2004), “Employment Protection, Job-Tenure and Short Term Mobility Wage Gains: A New Explanation for the Italian Case” CEPREMAP Working Paper, no. 2004-02, CEPREMAP, Paris.

Neal, D. (1999), “The Complexity of Job Mobility among Young Men”, *Journal of Labor Economics,* Vol. 17, No. 2 (Apr., 1999), pp. 237-261.

Newey, W. K., and D. McFadden (1994), “Large Sample Estimation and Hypothesis Testing”, in R. F. Engle, and D. L. McFadden, eds, *Handbook of Econometrics*, Vol. 4. Elsevier, North Holland.

OECD (2010), “Institutional and Policy Determinants of Labour Market Flows”, Chapter 3 in 2010 OECD Employment Outlook: Moving Beyond the Jobs Crisis.

Oreopoulos, P., von Wachter, T., and A. Heisz (2008), “The Short- and Long-Term Career Effects of Graduating in a Recession: Hysteresis and Heterogeneity in the Market for College Graduates”, IZA Discussion Paper No. 3578.

Parsons, D. (1972), “Specific Human Capital: An Application to Quit Rates and Layoff Rates”, *The Journal of Political Economy,* Vol. 80, No. 6, (Nov. – Dec., 1972), pp. 1120-1143.

Pavlopoulos, D., Fouarge, D., Muffels., R. and J. Vermunt (2007), “Who benefits from a job change: The dwarfs or the giants?”, IRISS Working Papers, No. 2007-16, December 2007.

Postel-Vinay, F. and J.M. Robin (2002), “Equilibrium Wage Dispersion with Worker and Employer Heterogeneity”, *Econometrica,* Vol. 70, No. 6 (Nov., 2002), pp. 2295-2350.

Shimer, R. (2003), “The Cyclical Behavior of Equilibrium Unemployment and Vacancies: Evidence and Theory”, NBER Working Paper No. 9536.

Simpson, W. (1990), “Starting Even? Job Mobility and the Wage Gap Between Young Single Males and Females”, *Applied Economics*, Vol. 22, pp. 723-737.

Stevens, M. (2004), “Wage-Tenure Contracts in a Frictional Labour Market: Firms’ Strategies for Recruitment and Retention”, *The Review of Economic Studies*, Vol. 71, No. 2 (Apr., 2004), pp. 535-551.

Stigler, G. (1962), “Information in the Labor Market”, *Journal of Political Economy*, Vol. 70 (Oct., 1962), pp. 94-104.

Topel, R. (1991), “Specific Capital, Mobility and Wages: Wages Rise with Job Seniority”, *Journal of Political Economy*, Vol. 99, No. 1 (Feb., 1991), pp. 145-176.

Topel, R. and M. Ward (1992), “Job Mobility and the Careers of Young Men”, *The Quarterly Journal of Economics*, Vol. 107 (May, 1992), pp. 430-479.

Ureta, M. (1992), “The Importance of Lifetime Jobs in the U.S. Economy, Revisited”, *American Economic Review*, Vol. 82, No. 1, pp. 322-335.

Veum, J. (1997), “Training and Job Mobility among Young Workers in the United States”, *Journal of Population Economics,* Vol. 10, No. 2, (Jun., 1997), pp. 219-233.

Viscusi, W. (1979), “Job Hazards and Worker Quit Rates: An Analysis of Adaptive Worker Behaviour”, *International Economic Review,* Vol. 20, No. 1 (Feb., 1979), pp. 29-58.

Weiss, A. (1984), “Determinants of Quit Behavior”, *Journal of Labor Economics*, Vol. 2, No.3. (Jul., 1984), pp. 371-387.

Weiss, D., Dawis, R., England, G. and L. Lofquist (1961), "Validity of Work Histories Obtained by Interview", Industrial Relations Center, University of Minnesota.

Wooldridge, J. (2002), Econometric Analysis of Cross Section and Panel Data, MIT Press, Cambridge, Massachusetts.