Dynamic Labour Demand: Fixed and Quadratic Adjustment Costs

Paola Rota

University College London

July 1994

A thesis submitted for the degree of Ph.D. in the University of London
Acknowledgements

Chapter 1. is the result of joint work with Alessandra Del Boca, and benefits from the comments of Pietro Ichino. Chapter 2. contains material from joint work with Alessandra Del Boca.

I would like to thank my supervisor Costas Meghir, and also Richard Blundell, Alessandra Del Boca, Christopher Gilbert and Arthur van Soest. I would also like to acknowledge comments from participants in the Applied Microeconomics Workshop at UCL, and at SPES meetings at Brescia, Dublin and London [Microeconometrics of European Labour Markets project].

Financial support was provided by the University of Brescia through the EULO Association and by the European Community through a SPES scholarship.
Abstract

The thesis focuses on the way recent developments in labour market legislation have affected companies' hiring and firing practices. We consider Italy as a representative example of the effects of regulations on firms' employment decisions. We show that, in Italy, the costs imposed by hiring and firing rules are decreasing rather than increasing at the margin implying large and infrequent changes in employment. We therefore claim that the costs of varying employment are better approximated by a structure in which adjustment costs are fixed, i.e. they are independent of the size of the change in the labour input. This is in contrast with the standard theory of dynamic labour demand in which adjustment is based on strictly convex costs.

At a descriptive level, the salient feature of the data, on 3247 companies observed for the period 1982-89 in the Milan area, is a large peak corresponding to zero changes in employment which indicates infrequent changes in employment, as implied by the existence of fixed costs. In order to explain these facts we take two approaches. The first is to consider the presence of the peak at zero as a selection problem to be analyzed in the context of the standard model of dynamic labour demand based on strictly convex quadratic costs. Selection arises from either measurement errors or the presence of fixed costs. The second approach is to consider explicitly fixed costs. In this case, labour demand is governed by threshold rules which guide firms on whether or not to adjust in the face of exogenous shocks. In the presence of idiosyncratic companies, firms' labour demand is therefore characterized by alternate phases of adjustment and non-adjustment.

A final theoretical model encompasses both structures of adjustment costs.
INTRODUCTION

Chapter 1. HIRING AND FIRING REGULATION

Introduction

1.1. Italy as an extreme case of hiring and firing regulations
1.2. The effect of labour legislation on firms' firing decisions

   1.2.1. Unfair dismissal
   1.2.2. Collective firings

1.3. The effect of labour legislation on firms' hiring decisions

Appendix 1.1. Wage Supplementation Scheme
Appendix 1.2. Unemployed and first job seekers 1982-1991

Chapter 2. THE DATASET

Introduction

2.1. The dataset
2.2. Aggregate behaviour
2.3. Data description at the firm level
2.4. Interpretation of the "zeros"

   2.4.1. Wage Supplementation Scheme
   2.4.2. Turnover
   2.4.3. Measurement errors

2.5. Transition probabilities

Appendix 2.1. Variables specification
Appendix 2.2. Summary statistics at industry level
Appendix 2.3. Rate of change in employment, output and wages: summary statistics
Appendix 2.4. Percentage rate of change in employment by size
Appendix 2.5. Transition probabilities by industry
Chapter 3. QUADRATIC ADJUSTMENT COSTS: EULER EQUATION REPRESENTATION

Introduction

3.1. Theoretical model
3.2. Stochastic specification and choice of estimation method
3.3. The dataset
3.4. The selectivity model
3.5. Estimation of the selection term (step 1)
3.6. Estimation of the Euler equation (step 2)
3.7. Static model

Appendix 3.1. Euler equation
Appendix 3.2. Selection bias; distributional assumptions on the error terms
Appendix 3.3. Inverse Mills' ratio: summary statistics
Appendix 3.4. Plots of the probability of adjusting
Appendix 3.5. Program for GMM estimation
Appendix 3.6. Estimation of the Euler equation with time dummies

Chapter 4. FIXED ADJUSTMENT COSTS AND DYNAMICS

Introduction

4.1. Hamermesh's (1989) model
4.2. Discrete dynamic process: introduction

Chapter 5. THE PRINCIPLE OF OPTIMALITY AND THE DYNAMIC PROGRAMMING ALGORITHM: basic ingredients of the model

Introduction

5.1. Finite horizon problems
   5.1.1. Firm's problem: perfect state information
   5.1.2. The Principle of Optimality
   5.1.3. The DP algorithm
   5.1.4. Imperfect state information
   5.1.5. Reformulation of the problem of imperfect state information into a problem with perfect information
   5.1.6. Sufficient statistics
   5.1.7. Conditional probability measures
   5.1.8. Markov decision processes

5.2. Infinite horizon problems
   5.2.1. Formulation of the infinite-horizon problem
   5.2.2. Principle of Optimality
5.2.3. The DP algorithm
5.2.4. Contraction mappings
5.2.5. Convergence of the DP algorithm, necessary and sufficient conditions for the Bellman's equation

Appendix 5.1. Measurability problems

Chapter 6. A DYNAMIC MODEL OF LABOUR DEMAND WITH FIXED ADJUSTMENT COSTS

6.1. Structural model
6.2. Reduction of the model
6.3. S,s-type policy

Appendix 6.1. The logit model

Chapter 7. FIXED AND CONVEX ADJUSTMENT COSTS: can we recover the Euler equations?

7.1. First order conditions
7.2. Stochastic specification and the choice of estimation method
7.3. Estimation of the conditional choice probabilities
7.4. Estimation of the newly represented Euler equation

Appendix 7.1. Euler equation
Appendix 7.2. Logit estimates
Appendix 7.3. Calculation of short and long run elasticities

CONCLUSIONS AND TOPICS FOR FURTHER RESEARCH
### List of Tables

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Flexible accelerator</td>
</tr>
<tr>
<td>B</td>
<td>Euler equation and GMM</td>
</tr>
<tr>
<td>C</td>
<td>Linear absolute costs</td>
</tr>
<tr>
<td>D</td>
<td>Fixed costs</td>
</tr>
<tr>
<td>I</td>
<td>Intercountry comparisons</td>
</tr>
<tr>
<td>A.1.1</td>
<td>CIG hours</td>
</tr>
<tr>
<td>II</td>
<td>Number of firms per year</td>
</tr>
<tr>
<td>III</td>
<td>Number of firms by size, 1982</td>
</tr>
<tr>
<td>IV</td>
<td>Unemployment rate</td>
</tr>
<tr>
<td>V</td>
<td>Number of companies with zero employment change</td>
</tr>
<tr>
<td>VI</td>
<td>Number of companies with zero change in employment for two consecutive periods</td>
</tr>
<tr>
<td>VII</td>
<td>Employment change</td>
</tr>
<tr>
<td>VIII</td>
<td>Real output and wages (percentage)</td>
</tr>
<tr>
<td>IX</td>
<td>CIG, Extraordinary and Ordinary Scheme</td>
</tr>
<tr>
<td>X</td>
<td>Transition probabilities</td>
</tr>
<tr>
<td>XI</td>
<td>Standard Errors</td>
</tr>
<tr>
<td>XII</td>
<td>Frequencies</td>
</tr>
<tr>
<td>XIII</td>
<td>$\chi^2$</td>
</tr>
<tr>
<td>XIV</td>
<td>Equilibrium probabilities</td>
</tr>
<tr>
<td>A.2.1</td>
<td>Average employment across manufacturing industries</td>
</tr>
<tr>
<td>A.2.2</td>
<td>Average real wage across manufacturing industries</td>
</tr>
<tr>
<td>A.2.3</td>
<td>Average real output across manufacturing industries</td>
</tr>
<tr>
<td>A.2.4</td>
<td>Average gross profits across manufacturing industries</td>
</tr>
<tr>
<td>A.2.5</td>
<td>Average investments in new machinery across manufacturing industries</td>
</tr>
<tr>
<td>A.2.6</td>
<td>Rate of change in employment</td>
</tr>
<tr>
<td>A.2.7</td>
<td>Rate of change in real output</td>
</tr>
<tr>
<td>A.2.8</td>
<td>Rate of change in real wages</td>
</tr>
<tr>
<td>A.2.9</td>
<td>Transition probabilities: industry 1</td>
</tr>
<tr>
<td>A.2.10</td>
<td>Standard errors</td>
</tr>
<tr>
<td>A.2.11</td>
<td>Frequencies</td>
</tr>
<tr>
<td>A.2.12</td>
<td>Transition probabilities: industry 2</td>
</tr>
<tr>
<td>A.2.13</td>
<td>Standard errors</td>
</tr>
<tr>
<td>A.2.14</td>
<td>Frequencies</td>
</tr>
<tr>
<td>A.2.15</td>
<td>Transition probabilities: industry 3</td>
</tr>
<tr>
<td>A.2.16</td>
<td>Standard errors</td>
</tr>
<tr>
<td>A.2.17</td>
<td>Frequencies</td>
</tr>
<tr>
<td>A.2.18</td>
<td>$\chi^2$ test for state dependence</td>
</tr>
<tr>
<td>XV</td>
<td>Method of moments</td>
</tr>
<tr>
<td>XVI</td>
<td>Model &quot;(t-1)(t-2)(t-3)&quot;</td>
</tr>
<tr>
<td>XVII</td>
<td>Model &quot;(t-2)(t-3)&quot;</td>
</tr>
</tbody>
</table>
XVIII Model "(t-3)"
XIX Test of joint significance of groups of regressors with different timing:
  model "(t-1)(t-2)(t-3)"
XX (t-2)(t-3)
XXI Test of joint significance of groups of regressors with different timing:
  model "(t-2)(t-3)"
XXII Model "(t-1)(t-2)(t-3)"
XXIII Model "(t-2)(t-3)"
XXIV Model "(t-3)"
XXV Log-likelihood test
XXVI Pooled probits
XXVII Euler equation
XXVIII Characteristic roots
XIX Hausman test
XXX Euler equation: reparametrization
XXXI Static model
XXXII Wald test of joint significance of the coefficients of $L_{t+1}$, $L_{t}$, $L_{t+1}$
A.3.1 Inverse Mills' ratio calculated from Probit (t-1)(t-2)(t-3): $\lambda_{t}$ (levels)
A.3.2 $\lambda_{t-1}\lambda_{t}$ (differences)
A.3.3 Inverse Mills' ratio calculated from Probit (t-2)(t-3): $\lambda_{t}$ (levels)
A.3.4 $\lambda_{t-1}\lambda_{t}$ (differences)
A.3.5 Inverse Mills' ratio calculated from Probit (t-3): $\lambda_{t}$ (levels)
A.3.6 $\lambda_{t-1}\lambda_{t}$ (differences)
A.3.7 Probability of adjusting after having adjusted upwards at t-1
A.3.8 Probability of adjusting after not having adjusted at t-1
A.3.9 Probability of adjusting after having adjusted downwards at t-1
A.3.10 Euler equation with time dummies
XXXIII Euler equation with quadratic and fixed costs
XXXIV Characteristic roots
XXV Alternative model
XXVI Characteristic roots
XXVII Short and long run elasticities
A.7.1 Model "(t-1)(t-2)(t-3)"
List of Figures

1. Average employment
2. Average real wage
3. Average productivity
4. Average profitability
5. Average investments in new machinery
7. 1983-1984
8. 1984-1985
9. 1985-1986
10. 1986-1987
11. 1987-1988
12. 1988-1989
15. 1984-1985
16. 1985-1986
17. 1986-1987
18. 1987-1988
19. 1988-1989
20. Real wages 1982-1983
22. 1984-1985
23. 1985-1986
24. 1986-1987
25. 1987-1988
26. 1988-1989
27. Output
28. Real wage
29. Productivity
30. Profitability
31. Investments in new machinery
32. Adjustment regimes
33. Transitions 1982-1986
34. * Transitions 1985-1989
35. Equilibrium probabilities
A.2.1 Firms with 1-19 employees 1982-1983
A.2.2 1983-1984
A.2.3 1984-1985
A.2.4 1985-1986
A.2.5 1986-1987
A.2.6 1987-1988
A.2.7 1988-1989
A.2.8 Firms with 20-49 employees 1982-1983
A.2.9 1983-1984
A.2.10 1984-1985
A.2.11 1985-1986
A.2.12 1986-1987
A.2.13 1987-1988
A.2.14 1988-1989
A.2.15 Firms with 50-99 employees 1982-1983
A.2.16 1983-1984
A.2.17 1984-1985
A.2.18 1985-1986
A.2.19 1986-1987
A.2.20 1987-1988
A.2.21 1988-1989
A.2.22 Firms with 100-199 employees 1982-1983
A.2.23 1983-1984
A.2.24 1984-1985
A.2.25 1985-1986
A.2.26 1986-1987
A.2.27 1987-1988
A.2.28 1988-1989
A.2.29 Firms with 200-500 employees 1982-1983
A.2.30 1983-1984
A.2.31 1984-1985
A.2.32 1985-1986
A.2.33 1986-1987
A.2.34 1987-1988
A.2.35 1988-1989

36 Panel data estimation
37 Inverse Mills’ ratio
38 Probability of selection
A.3.1 Probability of adjusting after having adjusted upwards the previous period
A.3.2 Probability of adjusting after not having adjusted the previous period
A.3.3 Probability of adjusting after having adjusted downwards the previous period

39 Decision tree
40 Law of motion
41 Decision, decision rule and policy
42 Equation (180)
43 Valuation function and conditional valuation function
44 Decision tree: quadratic and fixed costs
45 Plot of pq
46 Plot of the probability of adjusting
INTRODUCTION

Firms face costs if they choose to adjust their staffing levels. During recessions firms hoard labour and during expansions they may operate to some extent understaffed in order to avoid the cost of adapting employment to the level they would attain in the absence of these costs. Both economists and employers acknowledge that the non-wage component of the costs of varying the labour input has played an increasingly important role in determining the temporal pattern of labour demand in response to exogenous shocks.

It takes time and money for a firm to adjust the workforce to shocks. For instance, in the case of recruiting, a firm must advertise, recruit, interview, and train new workers. If the cost of hiring is high, a company, which perceives a shock as transitory, may decide to operate with a number of workers smaller than a comparable firm facing lower costs. Furthermore, if firing costs are high, understaffing may result also from the decision to avoid the possibility of bearing the expenses relating to labour reductions. It is clear, therefore, that hiring and firing (labour adjustment) costs affect the timing and the extent of the change of the labour-force in response to shocks. The type of constraint that adjustment costs impose on optimal employment decisions represents a crucial issue in understanding firms' behaviour over time. Any policy aimed at reducing employment fluctuations and lay-offs by changing the costs of hiring and firing should carefully evaluate the impact of these costs on the employment decisions of forward-looking firms.

It is not straightforward to define and categorize labour adjustment cost. We
broadly interpret these costs as the costs that relate to hiring and firing, and affect labour and/or job turnover, and the number of workers and/or hours. Not all costs have an explicit monetary dimension. An example is the cost of disruption generated by the expansion of total employment within a firm, which might imply rescheduling work across plants, training and the incorporation of new workers among the experienced. Another possible example is reduced morale among the remaining employees when some workers are laid-off. Even if we can find an acceptable definition of adjustment costs, therefore, measurement difficulties remain. Some costs affect companies' output decisions, and the mix of hours and workers, through disruption to production when the employment levels are changed. Other costs are simply out-of-pocket costs of hiring and firing such as advertising or, under certain institutional arrangements, payment by employers of unemployment benefits or severance payments to workers permanently laid-off. In Chapter 1, we identify an extra category of cost to add to what is an already complex classification, namely the cost implied by the attempts to circumvent legislation and legal decisions. This cost is likely to be directly proportional to the degree of restrictiveness of labour provisions in a country or in a market.

Some costs, affect the way the production process is organized and the optimal number of jobs within a company. Among these is the cost of terminating contract where strict unfair dismissal legislation applies. The disruption of the workforce provoked by expansion or contraction, like collective firings, provides another example of such costs, and there are also other expenses, like advertising, interviewing and the provision of unemployment subsidies which do not directly relate to the production process.
Another important distinction is between costs relating to net changes and costs relating to gross variations in employment. Net changes, i.e. the differences between positive and negative variations, are a relatively aggregative measure of employment adjustment. For instance net changes over a year do not take into account the complexity of the flows which have occurred over this time unit, such as labour turnover which includes hiring and firing of employees, voluntary quits, transfer of workers between plants, etc. Thus, we may refer to two substantially different sources of adjustment costs: net changes, which we may interpret as a measure of changing the optimal number of jobs within a firm; and gross changes, which occur whenever a worker joins or leave a firm and is replaced, and which measure the flows of people necessary to fill a given stock of jobs. Gross costs are important, because firms react to shocks by initiating different types of flow (collective firings, natural wastage, promotions, transfers, etc.) according to the constraints they face. Nevertheless, sufficiently detailed information on turnover is seldom available, and this forces most empirical work to consider only net changes.

Another difficulty concerns the direct measurement of adjustment costs. Early studies made some attempt to do this. Holt, Modigliani, Muth and Simon (1960) calculated that, over the period 1959-69, adding a new worker to staff cost about $360, while firing an employee implied an expense of about $170. Oi (1962), using data from the International Harvester Company, found that the amount of money invested per new employee in 1951, was on average about $380 (this included all the costs a company might expect to pay: recruiting, training, lay-off costs, unemployment compensation, etc.).

Generally, the evidence is only indirect. Some studies have attempted to
measure the stock of hoarded worker-hours. The studies by Fay and Medoff (1985) and Fair (1985), despite using very different methodologies, reach compatible results. Labour hoarding ranged between 4.5 and 8.5 percent of worker-hours during recessions in United Sates through the postwar period. Burgess and Dolado (1989) estimate that adjustment costs amount to 0.25 percent of the quarterly wage-bill for manufacture employment in the United Kingdom. But despite these examples, little progress has been made in measuring the magnitude of adjustment costs. Even if we manage to find a widely acceptable definition of these costs, the lack of information at a very disaggregate level, such as the firm or the plant, does not allow us to measure them directly.

An indirect approach to recognizing the existence and restrictiveness of adjustment costs is to conduct interviews with employers. A study by Emerson (1988) reports a survey of employers attitudes and reveals substantial concern over restrictive hiring and firing rules; Wells (1993) discusses the results of a survey of employers views across European countries on whether hiring and firing rules were an impediment to employing more people: in the UK, less than 10 percent of the employers considered inflexibility in hiring/shedding labour as a very important obstacle to increased employment whereas in Greece, France, the Netherlands, Italy and Spain the figures were much higher. Spain reported more than 80 per cent followed by Italy with nearly 80 percent. The other countries, Germany, Ireland, Portugal and Belgium, reported figures between 21 and 45 percent.

The structure of adjustment costs is very important in addition to the source of these costs to an understanding of firms' optimal employment decisions over time and the implications of policies for reduction of employment fluctuations over the
cycle.

Adjustment costs have traditionally taken in the literature to have a strictly convex structure, typically represented by a quadratic functional form. This specification was introduced by Holt, Modigliani, Muth and Simon (1960) who specifically studied the relationship between costs and optimal labour input in a case study of the Pittsburg Plate Glass Company. They assume a quadratic functional form for adjustment costs as a "tolerable approximation over a range" (p. 53). The hypothesis of adjustment costs increasing at the margin represented the theoretical justification of the distributed lag models and became the foundation of the subsequent empirical literature on dynamic labour demand.

Quadratic adjustment costs have been applied extensively to interrelated factor demands, and specifically to the demand for labour, but both these literatures derived from developments in investment theory. There is neither any a priori justification nor any general empirical argument in favour of strict convexity of adjustment costs. The hypothesis is based on the view that firms find it more costly to adjust inputs quickly rather than gradually. It is largely because of the

---

1 A detailed survey of models of labour demand with quadratic adjustment costs, may be found in the articles by Nickell (1986) and by Del Boca and Rota (1989b), and in the book by Hamermesh (1993). Alternative structures of adjustment costs are introduced by Nickell and more thoroughly developed in Hamermesh.

desirable properties of strict convexity in relation to mathematical tractability, that this assumption has become the centre of the dynamic model building.

The initial generation of works based on this assumption is represented by the flexible accelerator, partial adjustment and distributed lag models. Strictly convex costs represented the theoretical justification for a lag structure which had been superimposed on basically static models, as for instance in Jorgenson's theory of investment (Jorgenson, 1963). [see Nerlove, 1972, for a discussion]

A profit maximizing firm which faces quadratic costs in varying an input would determine the optimal adjustment path of that input by finding a convex combination of last period's level and a target, long term input level. If we indicate the input as \( L_t \), labour, we have

\[
\Delta L_t = \lambda (L_t - L^*)
\]  

(I)

or

\[
L_t = L_{t-1} - \lambda (L_{t-1} - L^*)
\]  

(II)

where \( L^* \) is the target, long term or equilibrium level of the input and \( \lambda \) is the accelerator coefficient.

Technically, equations (I) and (II) are the solution to a maximization problem where the first order conditions give rise to a second order difference equation. To solve this equation, a linearization is taken around the target level of the input. This allows us to calculate the characteristic roots of the polynomial form associated with the difference equation. The root which has a value less than one is the stable root implying a process which converges to the long run equilibrium level of the input. This root represents the flexible accelerator [see for instance
In Table A we report a brief description of the technical features of the flexible accelerator.

### Table A Flexible accelerator

Assume that a firm wishes to maximize profits $\Pi$, which, for simplicity, are only function of one input, labour. $L_{t-1}$ is the initial stock the firm starts with

$$
\max_{L_t} \sum_t \beta^t \left[ \Pi(L_{t-1}) - \frac{1}{2} \gamma (L_t - L_{t-1})^2 \right]
$$

where $\beta$ is the discount rate and $\frac{1}{2} \gamma \Delta L_t$ indicates quadratic adjustment costs, $\gamma > 0$. The first order conditions (i.e. the Euler equation) for this maximization are

$$
\Delta^2 L_t - \Delta L_t + \frac{1}{\gamma} \Pi'(L_{t-1}) = 0
$$

where $\Delta^2(\cdot)$ indicates second difference. Linearizing equation (IV) in the neighbourhood of the steady state we obtain

$$
\Delta^2 L_t - \Delta L_t + \Pi'(L^*) + \frac{1}{\gamma} \Pi''(L_{t-1} - L^*)
$$

The polynomial form associated to the difference equation (V) yields two characteristic roots. One of them, the stable root, is less than one. The solution to (V) has the form

$$
L = L^* + A_1 \lambda_1 + A_2 \lambda_2
$$

where $L^*$ is the particular integral and $A_1 \lambda_1 + A_2 \lambda_2$ is the general integral. Setting the coefficient of the unstable root to zero, we obtain the flexible accelerator formulation in equation I. Differentiation of (I) with respect to time gives equation (II).

The larger the gap between the actual and the desired level of the input the faster the adjustment. The process of adjusting the input slows down as the difference between the two levels becomes smaller. $L_t$ converges smoothly to $L^*$ and reaches it asymptotically. Firms react to exogenous shocks by distributing input.
variations over time and adjusting gradually to the desired level. 3

A great deal of attention has been devoted to the analysis and the estimation of the speed of adjustment, λ. The difficulty in this literature concerns the determination of the target level of the input, which is endogenous and is affected by the level of input attained at each stage of the adjustment process. This solution is only straightforward if expectations are static. In the more plausible case in which they are not static, we need to infer the target level of the input, i.e. where the company is aiming at, by observing the speed of adjustment. We may guess that if the speed of adjustment is high then it is likely that the firm has been strongly shocked away from the desired level of the input. By contrast, if the speed is low we may infer that the company is quite close to the equilibrium level. Nevertheless, as we will discuss later, if adjustment costs do not have a quadratic form we may reach incorrect conclusions. Adjustment may occur instantaneously, as in the case of linear costs, or by jumps and intermittently, as in the case of fixed costs.

The more recent generation of models has overcome these limitations by considering more complicated expectations formation processes, and through direct

---

3 This implies that, if the target level L* varies over time, the adjustment is less than it would be in the absence of adjustment costs, i.e. in a non dynamic context. In other words, the optimal employment path which results from the model discussed entails labour hoarding. This was a welcome result in the literature since it provided a possible interpretation of the phenomenon of the short run increasing returns to labour (SRIL): the classic finding that in the short run as output increases (decreases), manhours and, specially, employment increased (decreased) less than proportionally. This result was at first seen as contradicting the theory of the firm according to which, in the short run, with the stock of capital fixed and labour variable, the short run elasticity for labour with respect to output should be greater than unity and smaller than the long run elasticity, and both marginal and average costs should be rising with output. Among the possible explanations of this paradox, labour hoarding represented the most common rationale. [see Berndt (1981) for a survey]
exploitation of the first order conditions from the maximization, without needing a linearization in the neighbourhood of the steady state, and the explicit solution of the difference equation. The estimation method, known as "generalized method of moments" (GMM), was suggested by Hansen (1982) and Hansen and Singleton (1982) and allows us to directly exploit the Euler equation to provide orthogonality conditions with respect to an appropriate set of instruments. Under the GMM, we set the sample covariances which correspond to the theoretical moment restrictions as close to zero as possible. In Table B we summarize the main characteristics of the Euler equation approach.

---

Table B Euler equation and GMM

The firm starts at \( t \) with the stock of labour, \( L_{t-1} \). The objective is to maximize the expected present value of future stream of profits in the presence of adjustment costs:

\[
max_L E_t \left( \sum_{i} \beta^i \left[ \Pi_i (L_{t-i}) - C(\Delta L_i) \right] | I_t \right)
\]  

(\text{VII})

where \( E_t \) is the expectations operator conditional on information available at the start of period \( t \) indicated by \( I_t \) and \( C(\Delta L_i) = \frac{1}{2} \gamma \Delta L_i \) denotes quadratic adjustment costs, \( \gamma > 0 \). The optimization problem is:

\[
V_t(L_{t-1}) = \max_{L_t} \{ [\Pi_t(L_{t-1}) - C(\Delta L_t)] + \beta_t E_t[V_{t+1}(L_t)] | I_t \} 
\]  

(\text{VIII})

First order conditions are

\[
\frac{\partial [\Pi_t - C(\Delta L_t)]}{\partial L_t} + \beta_t E_t \left( \frac{\partial V_{t+1}}{\partial L_t} | I_t \right) = 0
\]  

(\text{IX})

Taking the second equation in (\text{IX}) one period ahead and substituting into the first equation we obtain a standard Euler equation representation:

\[
\frac{\partial \Pi_t}{\partial L_t} + \beta_t E_t \left( \frac{\partial \Pi_{t+1}(L_t)}{\partial L_t} | I_t \right) = 0
\]  

(\text{X})

Formulation (\text{X}) introduces an expectational error due to the fact that the company chooses employment conditionally on information at time \( t \)

\[
\eta_{t+1} - \beta_t \frac{\partial \Pi_{t+1} - C(\Delta L_{t+1})}{\partial L_t} - \beta_t E_t \left( \frac{\partial \Pi_{t+1} - C(\Delta L_{t+1})}{\partial L_t} | I_t \right)
\]  

(\text{XI})

Under rational expectations we have that \( E(\eta_{t+1} | I_t) = 0 \). Euler equation estimation methods exploits this orthogonality condition between the firm’s forecast error and elements within its information set.

This approach obviates the need to specify the target level of employment and allows us to infer important properties of the adjustment path simply by
looking at the first order conditions, without solving the difference equation to obtain a closed form solution, as in the flexible accelerator model. Moreover, we are able to consider more general expectations formation procedures than the myopic mechanisms previously supposed. Employment decisions are rational in the sense that they incorporate all the available information at the time that the decisions are made.

In this brief discussion of the evolution of the analysis of quadratic adjustment costs we have identified the three crucial components we need to consider in modelling dynamic labour demand: the initial stock of the input, expectations formation, and the structure of costs.

Recent studies have focused on the effects of unions, labour market legislation and unemployment, on the speed of adjustment. Burgess (1988) includes union resistance as a potential adjustment cost and finds that adjustment is slower when the fraction of unionized workers is greater; Smyth (1984) finds a positive correlation between the unemployment rate and speed of adjustment; and Burgess and Dolado (1989) find a significant effect of legislation and labour market tightness on the speed of adjustment.

There is almost a unanimous view that labour market legislation can slow adjustment. This has resulted in a widely shared view about the sluggish employment recovery in Europe over the past decade. Despite a period of real wage moderation in Europe, the response of employment has been modest. Job security legislation is claimed to be responsible for greater persistence in unemployment in Europe than in United States, or than in European countries in
the early 1960's: a phenomenon defined by some economists as "Eurosclerosis". In stark contrast, US employment grew, between 1973 and 1991, at an average rate of 1.7 percent, almost six times the corresponding EC rate. The fact that strict regulation of labour markets may help to explain the poor employment performance of European countries has shifted attention away from the conventional wisdom according to which labour costs, in particular excessive real wage demands, represent the primary obstacle to sustained job creation. [see for instance Sachs, 1979, 1983 and Artus, 1984].

A recent paper by Bentolila and Dolado (1994) focuses attention on the divergence of real wage growth between Europe [1.7 percent per annum] and the US [0.4 percent] during the 1970's and 1980's. They point out how, in spite of the introduction of flexibility enhancing measures (such as temporary and part-time contracts) job creation in EC countries has not been particularly remarkable in the

---

5 Some studies have focused on intercountry comparisons of labour market tightness and adjustment speed. Emerson (1988) provides a comparison of labour market restrictiveness across European countries [Italy appears to be the most regulated and the United Kingdom the least regulated country] and supports the idea that hiring and firing costs are substantially higher in Europe than in United States; Wells (1993) ranks European countries according to various indicators of job flexibility: part time and temporary contracts, statutory duration of probation of qualifying period for blue collar worker, time taken to carry out redundancy of a blue collar worker, maximum duration for fixed term contract and temporary work. He shows that UK and Denmark are the least restrictive countries. Lazear (1990) analyzes 22 countries for the period 1956-84 and suggests a role for mandated severance pay in explaining unemployment when changes are substantial. Abraham and Houseman (1993a) compare the adjustment of manufacturing employment and hours in West Germany, France and United States; and Abraham and Houseman (1993b) between Germany and the US. They conclude that although the adjustment of employment to changes in output is much slower in Germany, French and Belgian manufacturing sector than in US manufacturing, the adjustment of total hours worked is more similar. They also find little evidence that the weakening of job security regulations that occurred in Germany, France and Belgium during the 1980's affected employers' adjustment to changes in output.
1980's with unemployment remaining stubbornly high. In other words, labour market deregulation, apart from raising employers' willingness to hire, was also probably expected to yield higher real wage flexibility, which however has not resulted. They argue that legal changes encouraging the use of temporary contracts have increased the power of insiders in the countries characterized by high protection for permanent employees, and has only enhanced flexibility at the margin. Those changes have, in fact, separated the workforce into two groups: the insiders, and a new group of temporary workers with low firing costs, whose attachment to their companies is fragile, precisely because their role is to bear the brunt of employment adjustments. This fragility naturally results in wage bargaining being mostly in the hands of the core of permanent, full-time employees. Under these circumstances, the insiders may be able to obtain higher wages through the presence of a buffer formed by flexible employees which reduces the likelihood that insiders will lose their jobs. Analyzing the case of Spain, they suggest that policies aimed at enhancing flexibility in the labour market should aim at reducing rigidities which affect the core workers, e.g. reductions in notice periods, severance pay, administrative delays to dismissals, and so forth but avoiding increasing flexibility at the margin.

Recent studies have emphasized that non-convexities in the structure of

\footnote{Indeed, already in 1962, Oi provided a definition of labour as a quasi-fixed factor. He pointed out how, in addition to the wage-bill, the largest part of total labour costs, the firm incurred fixed employment costs in hiring a specific stock of workers. These fixed costs were considered as an investment by the firm in its labour force and introduced an element of capital in the use of labour. Rothschild (1971), writing on investment, claimed that while strictly convex adjustment costs provided a justification of the use of distributed lags in empirical works, concave and linear costs provided a rigorous theoretical justification for the distinction between fixed and variable factors of production.}
adjustment costs may result from technological and organizational aspects inherent in the productive process and that they may have been substantially enhanced by labour market regulations. In particular, some hiring and firing provisions add an element of fixity to the cost structure providing an incentive for firms to react intermittently to changes in the exogenous variables. In this case the assumption which has driven the models based on quadratic adjustment costs, i.e. that the firm finds it more costly to vary the input by large amounts and instantaneously is contradicted. When adjustment costs are independent of the size of the change, firms have the incentive to vary the labour input quickly and by large amount in order to outweigh these costs. The adjustment path is characterized by periods of inaction and discrete jumps. Continuous and gradual reactions to exogenous shock would penalize the firm since each small change would imply the payment of fixed costs.

This view has been emphasized in recent works among which number Nickell (1978,1986), Hamermesh (1989, 1990a, 1990b, 1992, 1993a, 1993b), Bentolila and Bertola (1990), Bertola and Caballero (1990), Burda (1991) and Bertola (1992). Some of these studies have considered the case of linear adjustment costs as alternative hypothesis to quadratic costs. Also in the case of linear costs, Nickell (1986) assumes linear costs of adjustment as an alternative to quadratic costs, in order to show the possibility of non-smoothness in labour adjustment over the cycle. Hamermesh (1989) demonstrates using plant data that the standard model of convex variable adjustment costs of labour is inferior to a specification based on fixed costs of adjustment. Bentolila and Bertola (1990) compare countries with different job protection and show for Italy, West-Germany, France and U.K. that institutional constraints give rise to higher firing costs that may also affect hiring decisions. Firms tend to hire less in good times because they expect thereby to fire less in bad times. Employment appears more stable and unemployment more persistent in high job security countries. Bertola and Caballero (1990) show how firms, in the presence of non-convex adjustment costs, often make
labour adjustment is very different from that implied by the conventional assumption of strict convexity. Linear costs imply that the firm will not find it optimal to adjust employment to shocks smoothly and gradually, as in the quadratic costs model. The firm will correct its labour input instantaneously, albeit to a lesser extent than in the absence of adjustment costs. There will be a zone of inaction around the desired level of employment due to the existence of linear costs.

Table C  Linear absolute costs

Consider the same objective function as in the previous tables:

$$\max_{L_t} E_t \left\{ \sum_{t} \beta^{t} [\Pi_t (L_{t-1}) - C(\Delta L_t)] \right\}$$

(XII)

Adjustment costs are linear and take the form

$$C(\Delta L_t) = a|\Delta L_t|$$

(XIII)

it is clear that first order conditions reduce to the following condition

$$\beta \Pi_t' = \pm a$$

(XIV)

which generates two target level of employment, $L^*$ and $L^-$ and a demand for labour of the following form

$$L_t = \begin{cases} L^* & \text{if } L_{t-1} \leq L^* \\ L_{t-1} & \text{if } L^- > L_{t-1} > L^* \\ L^- & \text{if } L_{t-1} \geq L^- \end{cases}$$

(XV)

The firm adjusts instantaneously and completely to the desired level of employment, $L^*$ if it reduces or $L^*$, if it increases the labour input. There will be a zone of non-adjustment when current employment lays within the two target levels.

infrequent, sharp corrections in preference to partial continuous adjustment. Burda (1991) estimates a model for manufacturing on annual data and finds limited evidence for quadratic adjustment costs while fixed costs are significant for all countries.
(Nickell, 1978, 1986; Leban and Lesourne, 1980; Gavin, 1986; Bentolila and Bertola, 1990; Bertola, 1992). In Table C we summarize the effect of linear costs. In the case of fixed costs, the demand for labour will be discontinuous, with changes triggered at thresholds values of employment determined by fixed costs. In the work of Hamermesh (1989), for instance, the firm will decide to change employment only if fixed costs are outweighed by the resulting increase in profitability. In other words, the desired level of labour must be sufficiently different from the current stock of employees in relation to fixed costs; otherwise, the firm will not vary the stock of workers. The result consists of two alternative adjustment regimes: an immediate, complete jump to the desired level of employment, in order to minimize the impact of fixed costs; or complete inaction. Indeed, adjusting continuously by following a smooth path, as in the case of quadratic costs, will be expensive, since each time the firm will be obliged to pay fixed adjustment costs. In Table D we summarize the main results from this hypothesis.

---

\(^8\) The type of policy which results is very similar to the so-called "one-bin policy" in the literature on inventories (see Blinder, 1990, where many of the fundamental articles are reproduced).

\(^9\) This behaviour is known in the literature as two-bin or \((S,s)\) policy.
Consider the same objective function as in the previous tables:

$$\max_L E_t \left\{ \sum_{t} \beta^{t} (\Pi_t(L_{t-1}) - C(\Delta L_t)) \right\}$$  \hspace{1cm} (XVI)

Adjustment costs are fixed and given by a positive constant $K$:

$$C(\Delta L_t) = K$$  \hspace{1cm} (XVII)

Fixed costs give rise to two reservation values, $l^-$ and $l^+$, which trigger the adjustment, i.e. they result in two levels of input at which the firm is indifferent over whether or not to adjust. We obtain the following decision rule:

$$L_t = \begin{cases} 
L^+ & \text{if} \quad L_{t-1} \leq l^+ \\
L_{t-1} & \text{if} \quad l^- > L_{t-1} > l^+ \\
L^- & \text{if} \quad L_{t-1} \geq l^-
\end{cases}$$  \hspace{1cm} (XVIII)

The firm adjusts instantaneously and completely to the desired level of employment, $L^+$ ($L^-$), if the stock of workers $L_{t-1}$ is less (greater) than the reservation level, $l^+$ ($l^-$), otherwise it maintains its stock unaltered.

Hamermesh (1989) analyzes monthly data on output and employment, for the period 1983-87, relating to seven manufacturing plants of a large US durable goods producer. He shows how, in the presence of large fluctuations in output, employment of production-workers is essentially constant, except for large changes associated with very large variations in output. This conflicts with the slow adjustment implied by the increasing variable costs hypothesis. His conclusions are that more attention needs to be paid to linking maximizing behaviour to the underlying structure of adjustment costs.

If that, by comparison with the literature on quadratic costs, relatively little has been published using the hypothesis of linear costs, this is even more so with
regard to fixed costs which have only been analyzed by Hamermesh. One important issue which arises from the study of the role of fixed costs in firms' dynamic employment decisions concerns aggregation: both spatial and temporal [Hamermesh, 1992]. He claims that aggregate data limit our ability to draw inferences on the structure of adjustment costs. They make it difficult to discriminate between smooth and discontinuous adjustment; they do not allow us to account for heterogeneity across firms, and for idiosyncratic behaviour; and they may not correctly identify the appropriate decision time unit. As a result, it is crucial to study labour demand at firm or plant level if we want to understand the effects of regulatory policies on optimal employment decisions.

Another issue related to aggregation and studied a somewhat more extensively in the literature concerns heterogeneity in the labour input. Symmetry in the costs of hiring and firing has always been an implicit assumption in the literature on quadratic adjustment costs, although in the early works by Holt, et. al. (1960) and Oi (1962) hiring and firing costs were considered and measured independently. Their results suggested much higher average adjustment costs for increasing than for decreasing employment. Pfann and Palm (1993) use the GMM technique to estimate the structural parameters of an Euler equation with asymmetric adjustment costs. They present a model where the traditional symmetric Euler model with rational expectations is nested. Their empirical results suggest that asymmetries in adjustment costs can explain part of the imbalance between changes of production and non-production employment during upward and

---

downward phases of the business cycle. They find a difference in the speed of adjustment of production and non-production workers, and that this gap becomes larger during recessions. Production workers are more easily fired during recessions than hired during upturns. Also, firms tend to hire non-production workers during expansions and do not fire them during downturns.

Hiring and firing entail costs which have different origins and may result in differing patterns of adjustment. Nevertheless, as Hamermesh (1993) points out, in studying asymmetries there may be a difficulty in identifying variations in labour demand, and hence adjustment cost effects, separately from changes in labour supply. He argues that, for example, we may find that firms adjust employment more slowly in response to shocks to product demand when the short-run supply curve of labour is less elastic. Unless supply is specified very carefully, we would infer that adjustment costs are greater at such times (probably times of low unemployment) even though the phenomenon has nothing to do with the underlying asymmetry in the costs of adjusting demand. Asymmetries may not always stem from structural differences in adjustment costs.

In this thesis we analyze dynamic labour demand in the presence of quadratic and fixed adjustment costs. We analyze three theoretical models: the first, based on the traditional assumption of quadratic adjustment costs, is an Euler equation model; the second, characterized by fixed costs, is a discrete decision process where a firm decides whether or not to adjust and incur fixed costs; and in the third model we assume both quadratic and fixed costs and recover the parameters of the Euler equation from a model of a discrete decision process. We
view firms’ decisions as taken in an uncertain environment characterized by shocks to wages and output prices.

We pay attention to both theoretical and empirical aspects of dynamic labour demand. As mentioned in the previous pages, moving away from the hypothesis of quadratic costs to the assumption of fixed costs requires that we take into account discontinuities in labour demand generated by periods in which firms’ optimally choose not to adjust. This makes the Euler equation-GMM approach no longer appropriate and requires a framework in which a forward-looking firm may choose among alternatives and form an optimal policy which comprises a sequence of decisions of whether or not to adjust. The structure which allows us to consider dynamic labour demand in the presence of fixed costs is that of a "discrete decision process". We obtain a structural model of dynamic labour demand where employment decisions are modelled as a sequential decision process.

We extend the model to the case of both fixed and quadratic costs and, by conditioning on the company optimally deciding to adjust employment and bear fixed costs, we recover an Euler equation which is driven by quadratic costs but is also characterized by an extra term which summarizes the intertemporal choice of whether or not to adjust the firm continues to face given the existence of fixed costs in the model.

The theoretical implications are not the only interesting aspect of the analysis of the effects of adjustment costs on employment decisions. Many studies claim that labour adjustment costs have been substantially enhanced by labour market regulations, such as job security provisions. Institutional factors have not only increase the magnitude of these costs but have added elements of fixity in their
structure providing incentive for firms to adjust less frequently and by large amounts. It is crucial, therefore to carefully analyze the institutional frameworks which characterize labour markets in order to understand the type of constraints imposed on companies by regulations. We consider the Italian labour market which is regarded by many as providing a very clear example of a highly regulated market. Knowledge about the practices followed by companies in order to circumvent regulations, the living law versus the written law, appears to be very important in understanding the effects of hiring and firing costs on employment decision.

The empirical part of the thesis also focuses on Italy. We estimate the models with strictly convex costs, and with both quadratic and fixed costs using a panel data of 3247 Italian firms. The dataset is provided by Centrale dei Bilanci and covers period 1982-1989. The data show, each year, a recurrent peak of zero net changes in employment amounting to about 20 percent of the sample. One of the possible explanations for this phenomenon supports the idea that fixed costs inhibit adjustment for some companies in each period. [We also consider other explanations such as the possibility of measurement errors.] The existence of this peak poses for the quadratic cost model, where firms in the face of shocks always adjust in a smooth and continuous way, a problem of selectivity. We discuss a general, consistent method of analyzing selectivity bias within panel data GMM estimation.

Only a few studies of dynamic labour demand have considered microdata. As Hamermesh (1989 and subsequent works) shows it is not possible to analyze fixed costs at an aggregate level, such as the industry or the economy. By aggregating, we smooth over the discontinuities and miss out the effects of fixed
costs resulting in possibly biased conclusions on the effects of labour market policies on employment. In our work, microdata play a crucial role by adding measurement errors, which are characteristic of datasets of this type, to exogenous shocks as an extra source of uncertainty. This second type of uncertainty allows us to apply discrete choice methods and obtain estimates of the probability that a firm optimally adjusts or not in the presence of fixed costs.

The thesis is organized as follows. Chapter 1 considers the relationship between institutions and adjustment costs through the analysis of hiring and firing legal provisions in Italy. Chapter 2 reports the descriptive evidence on adjustment costs from a dataset of Italian companies. Chapter 3 analyzes the case of quadratic adjustment costs and considers an Euler equation-GMM approach in the presence of endogenous selection. Chapter 4 introduces fixed costs and the theoretical problems which arise when considering them. Chapter 5 outlines the basic technical tools for a discrete dynamic process. Chapter 6 develops a model of dynamic labour demand in the presence of fixed costs. Chapter 7 extends the model of fixed costs to the case of both quadratic and fixed costs recovering the Euler first order conditions and estimating them. Concluding remarks and hints for further work terminate the dissertation.
Chapter 1. HIRING AND FIRING REGULATIONS

Introduction

1.1. Italy as an extreme case of hiring and firing regulations

1.2. The effect of labour legislation on firms' firing and hiring decisions
   1.2.1. Unfair dismissal
   1.2.2. Collective firings

1.3. The effect of labour legislation on firms' hiring decisions

Appendix 1.1. Wage Supplementation Scheme

Appendix 1.2. Unemployed and first job seekers 1982-1991
Introduction

The purpose of this chapter is to draw attention to the way recent developments in labour market legislation have affected companies’ hiring and firing practices. The current literature on the economics of optimal employment decisions over time emphasizes the differing implications of alternative adjustment cost structures in the determination of the way companies choose the sequence of profit-maximizing labour inputs over the cycle. Traditionally, the economics literature has assumed strictly convex adjustment costs. In other words, the costs associated with varying the number of employees is always increasing at the margin. This hypothesis is neither entirely justified theoretically nor empirically but it generates a smooth adjustment process towards the desired level of employment and a directly estimable empirical specification for the labour demand equation. By contrast, many technologically and legally imposed labour costs are fixed in the sense of being independent of the size of the change in the firm’s employment level. If fixed costs really matter, then firms’ employment decisions result in a process very different from that described by the traditional economic literature on dynamic labour demand: the firm will vary employment in a discontinuous way, either instantaneously jumping to the desired level of employment or, alternatively, keeping an unchanged number of employees. There will not be the gradual adjustment, distributed over time generated if costs are entirely convex. For this reason, it is essential to look at the actual form of the constraints imposed by regulations on hiring and firing decisions if we are to understand optimal decision making by firms in the labour market.
The institutions regulating the Italian labour market are a very instructive example of how such rules may influence firms' employment strategies. Our analysis suggests that fixed components in the structure of adjustment costs are likely to substantially affect the way employers react to shocks. While we do not propose an extended deep examination of the intricate web of Italian job security legislation, we focus instead on the interactions between rules and practices: the so-called "diritto scritto", the written law, versus "diritto vivente", the living law which appears to have played an important role in Italian industrial relations. The written law in Italy is the outcome of a lengthy bargaining process: it appears a complicated, often contradictory body of regulations, sometimes difficult to enforce. The gap between written and living law has widened over the years: the way the practice departs from the written law has become essential to an understanding of how things function.

It is not straightforward to define and categorize labour adjustment costs. Not all costs have an explicitly monetary dimension and for this reason even after reaching an acceptable definition the difficulty of measurement remains. For example, the cost of disruption generated by the expansion of the total work force in a company may imply rescheduling the work across plants, training and including new workers among the experienced ones and so on. Another instance might relate to reduced morale among the remaining employees when some workers are laid-off. Some costs affect companies' output decisions and the mix of hours and workers

---

11 We use the word "institution" in a broad sense as described by Matthews (1986): a set of rights and obligations affecting people in their economic lives. This encompasses four main features: property rights, conventions and norms, types of contract, and authority.
through disruption to production when the level of employment is changed. Other costs are simply out-of-pocket costs of hiring and firing such as advertising, or unemployment benefits paid by the employer, or severance payments to workers permanently laid-off. 12

In what follows we identify an extra element of cost to add to what is an already complex categorization, namely the cost implied by the attempts to circumvent the legislation and legal decisions. This cost is likely to be directly proportional to the degree of restrictiveness of labour provisions in a country or in a market. These costs are similar to fixed costs in that the employer has an incentive to bear them whenever a large change in employment is involved; or alternatively can leave employment unaltered and thereby avoid costs. As we will see in the case of firings, some of these costs have an explicit monetary dimension.

1.1. Italy as an extreme case of hiring and firing regulations

Italy has often been considered as the representative case in which the regulatory burden of hiring and firing has imposed a major constraint on employers' decisions with regard to their desired level of employment. Among OECD countries, Italy records the lowest dismissal figures, the highest percentage of first job seekers and the most strictly regulated recruitment environment. A broad inter-country comparison shows Italy with an extremely low degree of labour turnover, great difficulty associated with enforcing dismissals, and the highest percentage of

12 Hamermesh (1993a) calls the first category "internal costs" and the second "external costs" and provides a very detailed description.
companies claiming insufficient flexibility in hiring and shedding labour as a reason for not employing more staff (Emerson, 1988). As a result, Italy emerges as a country characterized by a very stable employment. In Table I we show some indicators relating to 1989. The share of lay-offs as a percentage of total unemployment, in 1989, was lowest in Italy at 28.7%, against 64.7% in France, 62.7% in Germany and 47.5% in Great Britain. (OECD, 1989) This illustrates our contention that shedding labour has been more difficult in Italy than in other countries.

Table I Intercountry comparisons

<table>
<thead>
<tr>
<th>% of unemployment: 1989</th>
</tr>
</thead>
<tbody>
<tr>
<td>country: Germany France Great Britain Italy</td>
</tr>
<tr>
<td>lay-offs</td>
</tr>
<tr>
<td>job losses as a reason for seeking alternative occupation</td>
</tr>
<tr>
<td>people seeking work after job loss</td>
</tr>
<tr>
<td>first job seekers</td>
</tr>
</tbody>
</table>

Source: OECD and Eurostat

From the "Labour Force Survey" (Eurostat 1989) it also emerges that the risk of job loss as a reason for seeking alternative occupation is again lowest in Italy
at 8.5% against 11.4% in France, 25.0% in Germany and 11.7% in Great Britain. The same pattern is followed by the percentage of people seeking work after job losses: 10.2% in Italy, 61.6% in France, 64.7% in Germany and 53.0% in Great Britain. Conversely but consistently with this outlook, the share of first job seekers is the highest in Italy, reaching 62.9%, while it is dramatically lower in all the other countries: 10.3% in France, 7.0% in Germany and 8.4% in Great Britain.  \(^{13}\)

On the hiring side, Italy is the only European country which has attempted to regulate precisely who is to be recruited (Emerson, 1988). The constraints on recruitment were drastically reduced as late as 1991. Italy has also the highest percentage of mandatory hiring in total employment (number of disabled over number of employees) in Western Europe: 15%, against 10% in France, 6% in Germany, 3% in Great Britain and zero in Sweden, in 1987 (Ichino and Violi, 1988).

Obviously, these large differences across countries reflect at least in part differences in labour market regulations, policies and practices: job security legislation, regulation of lay-offs and regulation on recruitment and industrial relations. For instance, the fact that lay-offs as a percentage of unemployment in Italy is half the EEC average suggests that shedding labour is difficult; nevertheless, employers have often succeeded in negotiating other means to reduce labour force

\(^{13}\) A recent study by Houseman (1991) shows that in Italy dismissals and lay-offs, for manual workers account for the lowest proportion of job leavers compared to France, Germany, Great Britain, Belgium and Luxembourg. Dismissals and lay-offs as a percentage of work force were dramatically higher in Britain, over the period 1975-1983. Italy had the lowest percentage among the Continental countries [West Germany, France, Belgium, Luxembourg]. Over the same period, dismissal and lay-offs accounted for a negligible portion of leavers in France, Italy, and Luxembourg. Again Britain has a much higher percentage, followed by Belgium, Germany and France.
levels in order to acquire some additional flexibility (see Del Boca and Rota, 1989a). In attempting to understand these figures we should bear in mind that, in Italy, most of the restrictions imposed by labour regulations have coexisted with forms of contracts and practices designed to circumvent the rules. This modifies the way companies are constrained and the associated costs as well as the extent and nature of workers' protection in relation to that envisaged in the legislation. It is for this reason that we regard the analysis of the institutional features and of the practices followed by firms as crucial to an understanding of the actual form of adjustment costs and labour adjustment in Italy.

1.2. The effect of labour legislation on firms' firing decisions

The seventies and eighties were characterized by a very mixed and uneven body of regulation relating to the firing decision. Very strict regulations coexisted with exemptions, relaxations of the rules for small firms, and areas lacking any discipline. The existence of exceptions in the regulation in the case of small

---

14 Until the beginning of the eighties, union action typically aimed at stabilizing employment, even when companies were going out of business. Between 1968 and 1979 the unions succeeded in obtaining legislation that "froze" potential collective firings and transformed, in practice, temporary lay-offs subsidies into benefits of indeterminate duration. This practice started with Law no. 1115 (1968) and was confirmed by Law no. 464 (1972). Unions did not worry about the consequences of imposing so many constraints on the labour market and on firms: until the first half of the seventies they did not perceive limits to their market power and believed that economic growth would take care of the negative consequences of the rigidities imposed on labour demand (Reyneri, 1987, 1992). Their leadership was weakened by the massive industrial reorganization throughout the 1980s. As companies reorganized their plants, invested in new technologies and shed labour, national industrial unions found themselves crowded out. With each company embarking on a different adjustment strategy, it became increasingly difficult for national unions to negotiate in relation to the process of reorganization (Locke, 1992).
has often provided incentives to companies not to expand their personnel beyond the thresholds imposed by legislation, but instead to organize themselves on the basis of single plant companies (Piore and Sabel, 1984).

While the legislation on firing greatly penalizes individual dismissals, it tends rather to facilitate large scale redundancies. In the industrial sector, the long lasting wage supplementation scheme, *Cassa Integrazione Guadagni (CIG)* (see Appendix 1.1. for a description), has, over two decades, provided the most common route to make employees redundant: an easily extendable wage subsidization programme for reduction in working time and for temporary lay-offs, applicable to a wide range of situations, such as bankruptcy, reorganization and restructuring, and sectoral problems. The programme is essentially financed by the government and provides 80 percent of the wage relative to shortfall of hours. Two features make CIG different from most other European countries unemployment benefit programs: an institutionally fixed replacement ratio and a rotation system which, at least in principle, imposes a labour sharing regime.

In spite of the important changes of 1990s, the firing legislation is still dominated by high procedural costs and restrictive attitudes that deters firms from reducing payrolls. The introduction of Laws no. 108 (1990) and no. 223 (1991) which regulate collective firings, has clarified the legal position so that the process of employment reductions has become more certain but in most cases no less expensive.

In what follows we examine the main features of the legislation that has so greatly influenced Italian labour market outcomes.
1.2.1. Unfair dismissal

According to the International Organization of Employers, in 1985, Italy ranked among those countries in which the obstacles to the termination of the employment contract are fundamental (Emerson, 1988). Unfair dismissal regulation is a significant example. The legislation exclusively concerns individual firings and is applicable to firms with more than 15 employees. As in most EC countries, the proof of the existence of a justified reason for firing a worker rests on the employer, but what counts as a reason for dismissing a worker is not incontrovertible. The judge has substantial power in determining whether a lay-off is justifiable. If his response is negative, the decision to lay-off will turn out to be very expensive for the company. Whenever an employee is fired without a justifiable and fair reason he/she has to be re-employed within 30 days and compensated with an amount equal to at least five months and up to a maximum of 12 months wages. The indemnity for the worker who accepts a job loss judged to have been unjustified is up to 20 months wages including 5 months wages as a fine on the company.

There is an asymmetry in the contractual position of the two parties. If the dismissal is judged unfair, the worker has the option to be either re-hired or reimbursed with a generous indemnity, whereas the company is left with no

---

15 A company may fire a worker on the basis of extreme lack of respect for his/her duties but this must be proved. The worker may also ask for union representation during the dispute.

alternative but either to accept the worker's decision to come back or to pay the fine. In practice, except for rare cases of very evident unsatisfactory behaviour in the work place, a dismissed employee will invariably initiate a claim for unfair dismissal. Bargaining between the firm and the worker will typically end up either in a job buy-out, or in rehiring, or in a large payment by the company.

We interviewed a representative officer of the employers' association (Assolombarda) and, also, some employers of medium sized companies to get a grasp of how, in practice, the bargaining process functions. They indicated amounts of upward of 6 months pay as a compensation for the buy-out (from 6 million lire (£2,500) upwards). One of the employers we talked to reported that his company had to pay 30 million lire (£12,500) as the result of a judgement of unfair dismissal. However, the amount may go far beyond this level, up to 150 million lire (£62,500), depending on the pay of the person involved. They also mentioned cases in which employers had attempted to create a very unfavourable environment around particular workers in order to provoke resignation, but they noted that this strategy can be quite time consuming and stressful. They explained how, in situations in which they needed to decrease the number of employees and firing was impossible, they attempted to transfer some employees, especially if "recommendable," to jobs with the same characteristics in other companies, often customers or suppliers. Another device to circumvent the burden imposed by the legislation is to remain beneath the critical size in order to take advantages of the exemptions to the rules. Small firms (up to 15 employees) and craft-workshops can
hire and fire much more freely than larger firms. This can induce firms to remain just beneath the size limit and hence restrain hiring. This incentive creates a discontinuity in employment behaviour implying hiring up to critical threshold and then inaction.

The fact that the introduction of Laws no. 108 (1990) and no. 223 (1991) implies that, except for large firings or plant closures, whenever a reduction is proposed, it is entirely up to the judge to approve the company’s decision is perceived as the greater limitation to companies’ freedom (Giugni, 1992). This view was also shared by the employers we interviewed. As a result, companies have an incentive to operate large reductions since smaller ones have a more uncertain and potentially expensive outcome. There are other incentives to operate firings in large aggregates such as the Wage Supplementation Scheme. Since the reduction of employment on an individual basis has been difficult and virtually unsubsidized, collective firings have been the most popular way to operate redundancies over the past two decades. We deal with collective firings in the next section.

1.2.2. Collective firings

Collective firings were not officially regulated in Italy until 1991. Instead they were internally regulated by an agreement between unions and the employers’ association (Confindustria), but these arrangements were effective only for the

17 Companies with less than 15 employees are subject to smaller firing costs: up to 2-6 months pay [Law no. 198 (1990)]. They are, in general, subject to lighter sanctions [Law no. 604 (1966)].

18 Law no. 604 (1966).
members of these two parties. The EC Court of Justice condemned Italy twice for
not having complied with the EC directive 19, according to which collective firings
had to be regulated by universally applicable laws. For fifteen years, until 1991, the
law did not provide any guidelines for collective firings which were considered
simply as instances of unfair dismissal. However, it was unclear that collective
firings should be subject to the same regulation as individual firings (Magnani,
1987). The practice, if not the legislation, seemed to suggest that reorganization of
productive activity leading to a reduction in the number of employees would
represent a fair reason for firings. But it took two decades before this question was
finally settled.

To fill the vacuum relating to collective firings and to solve the urgent
problems of industrial reorganization, Law no. 675 (1977) proposed "mobility" 20
to facilitate the transfer of labour from one firm to the other in cases of
redundancies in the industrial sector. The law failed in this objective for a large
number of reasons: in particular, the lack of adequate financial provision, of clear
objectives and of an adequate set of incentives for both employers and employees.
Firms were supposed to make a listing of excess employees, but they had no power
to terminate the employment relation unless the listed employees failed to find
another job. However, expanding firms had no incentive to hire from a list of the

19 EEC Directive no. 129 (17 Dec. 1975) concerning collective firings. Italy was
required to conform to the EEC directions twice: 8 June, 1982, Foro Italiano, 1982,

20 Unions' opposition to lay-offs was expressed in their claim that mobility could
only take place from one job to another, not from a job into unemployment.
"undesirable" employees made redundant by other companies. 21

Subsequent extensions and modifications 22 did not substantially increase the freedom of the companies to plan employment levels: for instance, Law no. 215, (1978) states that any worker who fails a trial period in the new company where he/she has been sent through the mobility list may return to his previous job. Only after the enactment of the Law no. 223, in 1991, were rules specifying the criteria for identifying collective firings introduced 23. Job mobility becomes the temporary last stage of social security protection: in principle, the company may start the process of reorganization with temporary lay-offs, financed by Wage Supplementation Scheme, for a duration of a maximum of 12 month in the North and 24 months in the Southern part of the country, more heavily hit by unemployment. If the firm realizes that the lay-offs are not temporary, it transfers the workers onto the job mobility programme which implies collective firing. The laid-off workers receive a subsidy of the same amount as the Wage Supplementation Scheme (about 80% of the wage) for a period of up to 12 months.

If the new law has brought more certainty to the process of employment reductions it has also added costs. On the one hand, the company has to pay 6 months of "mobility subsidy" (indennita' di mobilita') if it does not reach an agreement with the union. On the other hand, its ability to reduce employment is

21 At the end of 1982, only 29 cases of restructuring had been presented. In 1984, 164 applications were accepted. (Ferrera, 1987)


23 The company must have more than 15 employees and must be willing to fire at least 5 workers as a consequence of reorganization or of plant closure, and this must be within 120 days from the first dismissal.
again, as in the case of unfair dismissal, greatly limited by the judge’s discretion. Employees made redundant are ranked by the company according to i) family burden, ii) seniority and iii) skill. If the judge rejects this ordering the company incurs costs: penalties and legal expenses as in the case of unfair dismissal. Nevertheless, in case of collective firings, the larger the reduction the more certain and less expensive is the outcome: on the one hand the judge is less likely to interfere with the firm’s ranking; on the other, large reductions are more likely to entitle workers to CIG and mobility subsidy.

At the time of the discussion and the approval of the Law no. 223 (1991), union attitudes were favourable. However, with the worsening of the recession in 1992 and 1993, their attitudes have changed, as the aspects of liberalization brought in by the new law have been increasingly exploited by companies. In fact, when the mobility process implied by the law comes to an end nothing can prevent the employer from terminating the labour contract. Hence, the mobility programme is little more than a firing device. Union pressure on the government for more protection prompted Decree no. 398 (1992). Its aim is to protect employees from lay-offs implied by the mobility procedure. The text of the law recommends that all possible alternatives to permanent lay-off be tried before dismissal: transfer of employees across plants or to less qualified jobs, part-time contracts for older workers and job-sharing contracts (contratti di solidarieta’) implying a sharing of the reduction in working time between employees.\footnote{Unions have been in favour of solidarity contracts since early eighties, but it is only recently, with the new legislation, that the government has provided large incentives (reductions in payroll taxes) to companies who signs contracts involving a reduction of 30% in working time.} The decree also partly extends
the applicability of mobility procedure to firms with less than 15 employees. Small companies are not entitled to the mobility subsidy, but they may exploit the financial incentives for firms which hire workers from the mobility lists such reduction in payroll taxes. In practice, the mobility list has become the main way of organizing lists of unemployed workers, since, except in cases of firing for justifiable reasons, employees may register in order to be reallocated to other companies. However, the original function of the mobility regulation, the allocation of the excess labour generated by other companies, is becoming less effective since all unemployed workers with the exception of first job seekers may join the lists.

Despite the deep concern for employment protection that permeates the whole regulation of collective firings, including the very recent developments (the old philosophy that the mobility process should only end when another job is found), this legislation favours large scale reductions in employment through CIG and mobility lists. Marginal, small decreases in labour would encounter the restrictive and expensive unfair dismissal legislation made more costly by the uncertainty of the outcome due to disputes. A study by Del Boca and Rota (1989a) shows that companies which accessed the CIG and mobility programmes managed to eliminate large amounts of excess labour. Using a selected sample of firms under CIG Extraordinary Scheme relating to Milan province, they found that those companies managed to decrease employment by 60% on average over the period 1985-1989. Collective firings are typically characterized by non convex costs: applications for CIG and mobility, agreement with the union, waiting time for the application to be processed (2 months for the more temporary scheme, CIGO and 1 year for the long-lasting scheme, CIGS), and lower morale in the company all
represent costs which provide an incentive for companies to reduce workers by large amounts rather than by small and frequent decreases.

1.3. The effect of labour legislation on firms' hiring decisions

Regulation of recruitment is a significant example of the costs generated by the circumvention of institutional constraints. In what follows we briefly analyze the evolution of the legislation relating to labour exchange offices (Uffici di Collocamento), a public institution which forms the main instrument for regulating recruitment in Italy, and illustrate some of the practices companies have followed in order to ease the restrictions imposed on them by the legislation.

Until 1948, labour chambers (Camere del Lavoro) run by the union CGIL (Confederazione Generale Italiana del Lavoro) were in charge of recruitment and training of manpower. In 1949, the law proposed by labour minister Fanfani, and enacted by the Christian Democrat government, established a monopoly of public labour exchange offices in hiring, thereby taking the control over recruitment away from union. This change was enabled by the weakening of the unions, by the electoral defeat of the left, and by weak market conditions (high unemployment augmented by the massive outflow from agriculture and the south of Italy (Reyneri, 1987). A few years later, the other union, CISL (Confederazione Italiana Sindacati Lavoratori), and in particular its catholic wing, proposed as a focus for action employment guarantees and the improvement of job conditions in the work place, thus attempting to exert its power through plant level bargaining. (CISL General

---

25 Law no. 264 (1949).
Committee, Ladispoli, 1953). CGIL, the largest union, continued to bargain at national level until the middle of the fifties, when loss of control of FIAT's union bargaining committee (*Commissione Interna*), a symbol of the union's bargaining power, induced it to shift the attention to a two level bargaining strategy at national and plant levels (Horowitz, 1966; Turone, 1976). Plant level bargaining reached its peak during the "hot autumn" (1969) when it became completely unregulated. It was not until 1973 that the unions managed to recentralize bargaining. Centralization reached its peak with the 1975 agreement on the indexation clause (*Unificazione del Punto di Contingenza*), a strongly egalitarian wage escalator system aimed at narrowing wage differentials. This lasted until the very recent (July 1993) trilateral agreement on labour costs.  

The conflict between government and unions ended in what may be interpreted as a political exchange: unions traded their traditional primary role in labour placement and training for the introduction of regulation in recruitment designed to provide protection against discriminatory hiring: the so-called "numerical" system (*chiamata numerica*). Recruitment thus came to be regulated in the following way: employers could only hire new workers through the labour exchange office. One half of any labour increase had to be accounted for by hiring

---

26 The agreement signed by employers, unions and the government is of historical importance for the industrial relations system. Two major novelties characterize the agreement: 1) the escalator indexation clause has been abolished; 2) collective bargaining must take place every two years. Italian collective bargaining contracts have now a fixed and a shorter duration and they are agreed at national and sectoral level. This brings Italy into line with the bargaining process in other European countries.

27 Law no. 264 (1949); Law no. 300 (1970); Law no. 877 (1973); Law no. 863 (1984).
in line with the rank ordering of a list of unemployed. This list was formed by the public employment offices on the basis of social criteria, in particular length of unemployment spell and family size. The other half of the new hiring could be chosen directly by the employer, the so-called "nominative" system (chiamata nominativa). However, this rule did not apply in the following cases: i) companies with less than 10 employees ii) hiring of managerial staff iii) direct transfer from one firm to another. Employees made redundant had priority in re-hiring if the company which laid them off started to expand employment within a year of a previous labour reduction. For this reason, before employing new workers, firms were obliged to declare whether they had made employees with the same skills redundant in the previous year. (See Ichino (1987, 1988) and Napoli (1989, 1992) for a detailed account of legislation.)

Although, on the one hand, the loss of control on labour market induced unions to bargain a recruitment policy based on guarantees and universalism, employers did not oppose these developments in the context of the post-war excess of supply over demand, although they were subsequently affected by the lack of an efficient, market-oriented employment service. The obligation to hire through labour exchange offices made the recruitment of new employees a potentially very expensive process for firms. For instance, the process of trying out a worker for a provisional period of 15 days and then, in case of rejection, going back to the labour exchange office for a replacement, was costly in terms of time and lost output. On the other hand, avoidance of the constraints imposed by the legislation
or reliance on the public exchange officers’ "benign neglect" was also costly. In principle, the fact that the labour exchange offices constrained companies to hire an employee from the official list, in order to hire a second employee freely, might have induced the firm either to hire more workers than would otherwise have been necessary, or to postpone the decision of increasing its employment until it was prepared to take on a larger increase than initially anticipated. Not surprisingly, these strict regulations induced firms to contrive devices to ease the constraints. In conversations we had with some employers we were informed of a number of ways of avoiding some of the restrictions on hiring. For example, in order to recruit a particular worker to a vacant position, a medium sized company could arrange with a firm with less than fifteen employees, often a supplier, to hire the desired worker and then transfer him/her to the vacant job in the original company, since in this case the transfer would not be regulated. However, this 'device' was costly even if the cost was outweighed by the additional productivity of the desired worker, in that it required coordination between the two firms. Nevertheless, this recruitment practice was efficient for medium sized companies in cases where hiring a specific worker (or ensuring that the worker hired had a minimum level of skill) was deemed crucial to the firm's production plans.

The legislation on hiring was gradually changed over the course of 1980s. A first hint of liberalization came in 1984, when fixed term apprenticeship contracts were introduced to ease employment recruitment regulation (see Ichino, 1988). Workers between 15 and 29 years old could be hired for a period of up to 24

28 The Minister of Labour created in 1949 a public employment bureaucracy (i collecatori). The attitude of these public employment officers was inherently bureaucratic rather than service-oriented.
months on apprenticeship contracts without reference to any ranking. This legislation was introduced in order to remedy the high unemployment rate among first job seekers and to allow firms to obtain more flexibility. Indeed, a substantial increase in hiring through this type of contract followed its introduction. For example, in Milan province hiring through apprenticeship contracts accounted for 47% of total hiring in 1988 (estimates of the Lombardia Employers Association, Assolombarda, (1989)).

In 1987, after eight years of consultation, a new law instituted Regional Employment Committees (Commissioni Regionali dell'Impiego) and Regional Employment Agencies (Agenzie per l'Impiego) for the industrial sector. The Regional Employment Committees, actually, already existed. Introduced in 1970, they never worked effectively.³⁹ The aim of the law was the formation of more active, less bureaucratic public employment services. The Regional Employment Committee is an institution in which union and firms' representatives plan the manpower needs in each region. The Employment Agencies are in charge of the processing of all the information relevant to the matching between jobs and unemployed. They remain governmental institutes (under the Ministry of Labour) but the head of each agency is hired through the market and he remains separate from the civil service hierarchy. However, despite the potential for innovation, the recruitment was still governed by the labour exchange offices through the old numerical system.

The complete elimination of the rank ordering happened only in 1991 ³⁰

³⁹ Law no. 56 (1987).
while the public employment offices’ monopoly on recruitment still remains unaltered. Law no. 223 aims at a comprehensive reconstruction of the labour market legislation, but its institutional structure remains centralized and bureaucratic. Labour exchange offices have been left merely in charge of the function of certification of employment status, and the administration of mandatory hiring. They collect information on those registered, which, in principle, they should convey to the Regional Employment Agencies. Regional Employment Agencies convey to firms information on workers seeking jobs and are in charge of hiring of workers made redundant by other firms through a programme called "Mobility Lists" (Liste di Mobilita’). Regional Employment Committees have the overall planning power but in practice they have very little power.

Despite the 1991 legislation, the considerable institutional dispersion of the various functions of employment policy therefore remains. There is no coordination between mandatory hiring, administered by the powerless labour exchange offices, job mobility subsidies and active job placement, organized by the regional agencies. This dispersion and the delay in the start-up of the regional employment agencies (only a few agencies are fully operative) implies that in effect there has been little change with respect to the previous situation. As a result, on the one hand, the Italian unemployed are essentially left without a guide or services to face the difficulties of job search; while, on the other hand, companies are unable to avail themselves of a set of unified and coordinated institutions oriented to increase labour market efficiency. The continuing monopoly status of public employment offices makes private employment agencies illegal 31 (Ichino, 1992). An

31 Law no. 369 (1960).
employment contract is still subject to annulment if it is not ratified by a public employment office (except in the case of hiring managerial staff). 32

Our overview of the developments of the regulations on entry shows the labour market dominated by complex and conflicting arrangements subject to strong inertia. New regulations are introduced by patching up existing ones rather than by starting from scratch. Each new step in the process of institutional change is still determined by the prior distribution of power and authority. (Matthews, 1986)

The interpretation of the effects of hiring costs on optimal labour decisions is complicated by the fact that hiring decisions may be largely influenced by restrictions on firing. If employers anticipate that hiring new workers today implies that they will not be able to dismiss them during recessions, they will not increase their labour to the same extent they would in the absence of restrictions on firing.

Until recently, in the case of hiring the main source of rigidity was the barrier created by recruitment regulation on firms' labour demands and their ability to select workers. In order to give effect to their employment plans, companies were obliged to devise methods for circumventing the rules or to come to terms with the official bureaucracy. Again, this represented a costly activity: time spent in trying to get round the law, uncertainty of the result, deferred hiring as in the cases described above, and revision of plan in the case of the decision of not hiring an undesired worker. These costs together with the possible effects of the firing

32 This regulation is in the process of being amended by the recent trilateral agreement. One of the important issues on which employers, government and unions have agreed is the establishment of private employment agencies with the function of allocating temporary work and increasing the flexibility of the market. But the text of the law is still to be defined in concrete terms and it will be sometime before it becomes effective. 

54
restrictions on the decision of hiring new workers may have resulted in fewer hirings of young and unskilled people. This could explain the very high percentage of first job seekers in Italy despite some success obtained through the introduction of apprenticeship contracts. Table A.1.2. in Appendix 1.2. shows the number of unemployed, first job seekers and other people searching for a job in Italy over the period 1982-1991.  

In this chapter we have investigated the legal constraints imposed on the companies hiring and firing decisions and have analyzed how they affect the structure of adjustment costs and firms' employment strategies in Italy. Italy has provided a very clear example of the effects of regulations on companies employment plans. Indicators of job security among European countries show how, in Italy, the legislation on lay-offs and recruitment is more restrictive. In this study of firing and hiring regulations and of practices followed by employers, in order to circumvent the obstacles introduced by legal provisions, we show that in Italy the costs of adjustment are high and probably non convex. These costs appear to be decreasing rather than increasing at the margin implying large and infrequent changes in employment as we have described in the previous pages. Therefore, they are likely to be better approximated by a fixed rather than quadratic costs structure. Our discussion suggests that, in the case of firing, legally imposed costs of adjustment may be approximated by a fixed structure, namely costs independent of

33 The source is Bank of Italy, Relazione Generale del Governatore, 1992. The percentage of first job seekers over the total of unemployed, about 50%, results form this table lower than the one reported by Eurostat in Table I where the figure may have a different basis.
the size of the employment variation. Indeed, in certain cases clear fixed and out-of-pocket amounts of money are imposed through the legal system. The legal constraints on employment reductions appear to provide an incentive for companies to vary employment by large amounts, since it is more profitable to incur once and for all the fixed costs and dismiss a large number of workers rather than to do it gradually and have to pay those costs each period. Gradual adjustment would, in fact, imply the uncertain and expensive outcome imposed by the unfair dismissal procedures.

The case of hiring legislation is not equally straightforward, in part because hiring decisions are likely to be affected by the difficulty of making workers permanently laid-off. Forward-looking firms know that reducing the number of workers in bad times is costly and may adopt a policy of hiring less than they would do in the absence of restrictions on firings. Also, the legislation on hiring has developed in a very patchy way. Nevertheless, the monopoly status of the public employment offices in recruitment still remains. In the case of hiring, fixity in costs emerges from the attempts by employers to circumvent the legal restrictions in order to acquire flexibility. This type of cost, which also recurs in the case of firing, represents an important source of fixity which should be added to the fixed costs documented above in relation to firing. The scale of the devices employed to circumvent regulations are in proportion to the restrictiveness of the legal provisions and to the way the written and the living legislation differ. We have emphasized the gap between practices followed by companies and the text of the law. These practices are crucial to understanding how Italian firms react to shocks and make employment plans over time.
Appendix 1.1. Wage Supplementation Scheme

Wage supplementation (Cassa Integrazione Guadagni (CIG), Ordinaria and Straordinaria) in Italy is a programme which subsidizes reduction in working time and temporary lay-offs without breaking the employment relation. Cassa Integrazione Guadagni Ordinaria (CIGO) provides 80% of the wage relating to shortfall of hours caused by unfavourable transitory market conditions not attributable to either the employers or the employees. The firm is allowed to use the scheme for three months with the possibility of extension up to a maximum of 52 weeks (out of 104). CIGO is assigned at the request of the firm after consultation with trade unions and local authorities (Commissione Provinciale). The fund is financed by INPS (Istituto Nazionale della Previdenza Sociale), and by contributions of 2.2% paid by every firm (1.9% for firms with less than 50 employees), and additional contributions of 8% (4% for firms with less than 50 employees) of the wage supplementation by firms which require the intervention.

The Cassa Integrazione Guadagni Straordinaria (CIGS) provides the same coverage as CIGO in relation to wages (with a maximum of a monthly 1.250.000 lire) but it is more widely applicable: it subsidizes restructuring and reorganization. Its duration is linked to the reorganization process: initially up to 24 months with the possibility of subsequent extension to a maximum of 36. (The duration of the scheme has been changed by the Law no. 223 (1991): 12 months in the North and 24 in the South.) Although the application procedure is more demanding than that for CIGO (involving the authorization of a governmental committee, CIPI - Comitato per la Programmazione Industriale) the firms contributions to the scheme
are 4.5%, and 3% for companies with less than 50 employees. These contributions double after the 25th week of use. A complete account of the CIG legislation may be found in Napoletano (1992).

In Table A.1.1, we report the number of hours covered by Wage Supplementation Scheme for the period 1982-91 for industry only.

Table A.1.1. CIG hours

<table>
<thead>
<tr>
<th>year</th>
<th>Ordinary Scheme</th>
<th>Extraordinary Scheme</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1982</td>
<td>189,410</td>
<td>357,110</td>
<td>546,520</td>
</tr>
<tr>
<td>1983</td>
<td>226,037</td>
<td>443,441</td>
<td>669,478</td>
</tr>
<tr>
<td>1984</td>
<td>194,363</td>
<td>536,956</td>
<td>731,319</td>
</tr>
<tr>
<td>1985</td>
<td>117,066</td>
<td>496,363</td>
<td>613,429</td>
</tr>
<tr>
<td>1986</td>
<td>97,787</td>
<td>458,800</td>
<td>556,587</td>
</tr>
<tr>
<td>1987</td>
<td>85,189</td>
<td>358,758</td>
<td>444,947</td>
</tr>
<tr>
<td>1988</td>
<td>59,681</td>
<td>290,496</td>
<td>350,177</td>
</tr>
<tr>
<td>1989</td>
<td>48,249</td>
<td>232,581</td>
<td>280,830</td>
</tr>
<tr>
<td>1990</td>
<td>74,040</td>
<td>193,667</td>
<td>267,716</td>
</tr>
<tr>
<td>1991</td>
<td>139,098</td>
<td>188,744</td>
<td>327,842</td>
</tr>
</tbody>
</table>

in thousands

Source: Bank of Italy and INPS
Appendix 1.2. Unemployed and first job seekers: 1982-1991

Table A.1.2.

<table>
<thead>
<tr>
<th>Year</th>
<th>Unemployed</th>
<th>First job seekers</th>
<th>Other people in search of a job</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1982</td>
<td>281</td>
<td>1156</td>
<td>615</td>
<td>2,052,000</td>
</tr>
<tr>
<td>1983</td>
<td>352</td>
<td>1291</td>
<td>621</td>
<td>2,263,000</td>
</tr>
<tr>
<td>1984</td>
<td>466</td>
<td>1136</td>
<td>703</td>
<td>2,304,000</td>
</tr>
<tr>
<td>1985</td>
<td>468</td>
<td>1215</td>
<td>699</td>
<td>2,382,000</td>
</tr>
<tr>
<td>1986</td>
<td>501</td>
<td>1296</td>
<td>814</td>
<td>2,611,000</td>
</tr>
<tr>
<td>1987</td>
<td>547</td>
<td>1354</td>
<td>932</td>
<td>2,832,000</td>
</tr>
<tr>
<td>1988</td>
<td>535</td>
<td>1398</td>
<td>936</td>
<td>2,868,000</td>
</tr>
<tr>
<td>1989</td>
<td>507</td>
<td>1404</td>
<td>954</td>
<td>2,866,000</td>
</tr>
<tr>
<td>1990</td>
<td>468</td>
<td>1266</td>
<td>888</td>
<td>2,621,000</td>
</tr>
<tr>
<td>1991</td>
<td>469</td>
<td>1285</td>
<td>899</td>
<td>2,653,000</td>
</tr>
</tbody>
</table>

Source: Bank of Italy, Istat: rilevazione campionaria delle forze di lavoro.
Chapter 2. DATA DESCRIPTION

Introduction

2.1. The dataset

2.2. Aggregate behaviour

2.3. Data description at the firm level

2.4. Interpretation of the zeros
    2.4.1. Wage Supplementation Scheme
    2.4.2. Turnover
    2.4.3. Measurement errors

2.5. Transition probabilities

Appendix 2.1. Variables specification

Appendix 2.2. Summary statistics at industry level

Appendix 2.3. Rate of change in employment, output and wages: summary statistics

Appendix 2.4. Rate of change in employment by industry

Appendix 2.5. Transition probabilities by industry
Introduction

In the previous chapter we discussed the legal constraints imposed on the companies hiring and firing decisions and have analyzes how they may affect the structure of adjustment costs. Italy provided a representative example of the effects of regulations on companies employment plans. Our analysis of the legislation suggests that non-convex costs may provide a better description of the structure of adjustment costs.

Before concentrating on the modelling of the effects of different forms of adjustment costs on intertemporal employers' optimal decisions, we focus on the descriptive evidence from the dataset which we will use in the empirical part of our investigation. We focus on the behaviour through the sample period of the variables which are likely to affect the decision of whether or not to vary the labour input. We study these variables both at an aggregate and a disaggregate (industry and individual) level, relating to Lombardia: an industrialized region, representative of Northern Italy. Does our sample reveal, at a descriptive level, any information on the existence of non-convex adjustment costs to add to the indications we have obtained by analyzing the legislation? This question will be addressed in the following pages where we provide the descriptive statistics.
2.1. The dataset

The dataset is provided by Mediocredito Lombardo and is drawn up from the Centrale dei Bilanci databank of company accounts. It contains information on companies located in the highly industrialized district of Lombardia and which in 1982 had less than 500 employees. This data source covers more than 3000 manufacturing companies for the period 1982-1989. It provides, on an annual base, information on employment, sales, value added, the wage-bill and industry. Employment includes both productive and non productive workers. Relatively few of these companies are quoted (from 20 to 25 throughout the period) reflecting the predominant small-size. The number of companies in the panel is given in Table II, for each year in the sample.

Table II  Number of firms per year

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>no of firms</td>
<td>3253</td>
<td>3258</td>
<td>3258</td>
<td>3260</td>
<td>3263</td>
<td>3263</td>
<td>3263</td>
<td>3266</td>
</tr>
</tbody>
</table>

The configuration of firms by size (number of employees) in the starting year is reported in Table III. Small-sized firms (1-49 employees) account for nearly 50 per cent of the sample.

---

34 The companies included in the dataset are strictly "Societa' di capitale" even when they have a very small size. This implies that craft-workshops are not included in the dataset.
Table III Number of firms by size, 1982

<table>
<thead>
<tr>
<th>no. of employees</th>
<th>1-19</th>
<th>20-49</th>
<th>50-99</th>
<th>100-199</th>
<th>&gt;200</th>
</tr>
</thead>
<tbody>
<tr>
<td>frequency</td>
<td>494</td>
<td>1125</td>
<td>827</td>
<td>519</td>
<td>288</td>
</tr>
</tbody>
</table>

2.2. Aggregate behaviour

The behaviour of employment, real wages, productivity, over the entire period for the full sample is illustrated below in Figures 1-3 while Tables A.2.1.-A.2.3. in Appendix 2.2. report the actual figures. The behaviour of these variables is consistent with aggregate data. The period we analyze, 1982-89, saw transformations that, although not as dramatic as those experienced in the seventies, were nevertheless substantial. The decade 1981-90 was marked by two well-defined cycles which are paralleled by two phases of industrial development: the 1980-85 recession which resulted in a large reorganization in industrial production and the 1986-89 expansion in which the effects of the transformation of the eighties became apparent.

Figure 1 shows falling average employment across both industries and in the aggregate during the first half of the sample period and a recovery in the second
half.  

In 1989, indeed, the unemployment rate in Lombardia was 4.9 per cent, considerably lower than the national rate.

---

35 Industry 1 = ferrous, non-ferrous metals, chemicals and artificial fibres; Industry 2 = metal manufacturing, machinery, electrical equipment, transport equipment and instrument manufacturing; Industry 3 = food, drink and tobacco, textiles, leather goods, timber and furniture, paper and printing, rubber manufacturing and others.
Table IV shows the unemployment rate in Lombardia, Northern Italy and Lombardia North I. Italy for the whole Italy over the period 1982-90. From the table we see that Lombardia showed the same pattern of unemployment as Northern Italy as a whole, even if the figures for the latter were slightly higher. There is, by contrast, a large difference between the unemployment rates in Lombardia (and Northern Italy) and Italy as a whole [The remainder of Italy is Centre and South]. This difference is emphasized in the last three years in the table where the unemployment rate for the whole of Italy was more than twice the rate for Lombardia.

Real wages increased over this period as shown in Figure 2 and productivity, measured as output per worker accelerated after 1986-87 (Figure 3). It may be already seen from the disaggregation by broad industry groups that there can be remarkable differences in firms behaviour.
Figure 2 Average real wage
We also present the plots of gross profits and investment in new machinery aggregated over companies. We will use these variables later as indicators of profitability and investment activity respectively. They are plotted in real terms. As explained in Appendix 2.1., they are divided by the value added price for manufactures, the same deflator applied to real wages and sales. We have also scaled these two variables dividing them by employment in the first year of the sample period, 1982. The two variables are illustrated in Figures 4 and 5 while
Tables A.2.4.-A.2.5. in Appendix 2.2. report the actual figures.

Figure 4 Average profitability
In both cases the aggregate figures show little variability across years. In particular, within industry variations across years appear to be very smooth. There are, as for the previous graphs, differences across industries. While gross profits seem to slightly recover only in the final years of the sample period, the series on investment in new machinery follow more closely the performance of the economy as a whole: decreasing during the 1982-85 recession and increasing during the
period of expansion in 1986-89. It is interesting to notice that the aggregate series shows this very smooth behaviour despite strongly heterogenous behaviour at the company level.

According to census data on employment in Lombardia for the years 1980, 1985 and 1990 (collected and elaborated by ASPO), very small firms have made the largest contribution to the employment growth of employment, confirming a trend apparent in the seventies towards smaller, more specialized one plant companies.\textsuperscript{36} Companies with less than 50 employees have generated a 30-40 per cent increases in employment, whereas firms employing more than 50 workers have reduced the number of workers by 20-30 per cent. The size class from 1 to 19 employees accounts for 45-50 per cent of total employment in 1989. \textsuperscript{37}

2.3. Data description at the firm level

To obtain a first approximation to the process of employment adjustment we analyze the distribution of the rate of change in employment each year, starting from 1982, in relation to the rate of change in sales and wages (see Figures 6-26). A number of features suggest the existence of some degree of fixity in the cost of adjusting labour.

\textsuperscript{36} ASPO census data consider local units while the account data also include plants outside the district in case of multiplant companies.

\textsuperscript{37} Company turnover has been very intense: 50 per cent of the firms existing today were born after 1981. Between 30 and 40 per cent of productive units existing in 1981 are no longer operating. The high natality rate is recorded not only in services and trade sector but also in mechanical engineering (49.7 per cent) and in more traditional industries like textiles and clothing. The mortality rate appears strongly inversely correlated with size. (Rapporto ASPO Lombardia, 1991)
First, changes in employment show a recurring pattern throughout the sample period (Fig. 6-12). The very high spike at zero shown, in the graphs and recorded in Table V reveals a considerable stickiness in employment. Throughout the period, the rate of change in employment was zero for more than 20 per cent of the firms, on average. By contrast the rate of change in sales and in wages is characterized by a more standard bell-shaped distribution (Fig. 13-26). [Appendix 2.3. illustrates the main descriptive statistics for these variables.] This contrasts sharply with the appearance of the data aggregated over the 3000 companies.

<table>
<thead>
<tr>
<th>Year</th>
<th>( E_t-E_{t-1} = 0 )</th>
<th>Total no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1982-83</td>
<td>913</td>
<td>3253</td>
</tr>
<tr>
<td>1983-84</td>
<td>640</td>
<td>3258</td>
</tr>
<tr>
<td>1985-85</td>
<td>717</td>
<td>3258</td>
</tr>
<tr>
<td>1985-86</td>
<td>735</td>
<td>3260</td>
</tr>
<tr>
<td>1986-87</td>
<td>662</td>
<td>3263</td>
</tr>
<tr>
<td>1987-88</td>
<td>648</td>
<td>3263</td>
</tr>
<tr>
<td>1988-89</td>
<td>587</td>
<td>3266</td>
</tr>
</tbody>
</table>

\( E = \text{employment} \)

---

38 One issue often raised is about the accuracy of the data on the number of employees reported by companies since these figure do not explicitly appear in the balance sheet. In other words, this variable could be subject to some degree of approximation and misreporting. However, since the dataset records companies which have applied for loans at Mediocredito Lombardo, we know that the bank requires a detailed documentation and checks, in particular, that the all the social security contributions have been paid up to the application date. This provides at least a partial a check on the accuracy of the figures reported by the companies.
Figure 12 1988-1989

Figure 13 1982-1983

Figure 14 1983-1984

Figure 15 1984-1985

Figure 16 1985-1986
Figures 6-12 also show a discontinuity in the rate of change in employment occurring each year above the zero mode on the right hand side of the distribution. Firms appear to increase employment beyond a threshold around 2 per cent rather than adjusting continuously by small amounts, as implied by the convex adjustment costs model; however, no jumps are apparent on the left hand side of the distribution. Note that negative changes include dismissals as well as voluntary quits and retirements. However, firms faced with legal restrictions on employment
decisions may rely on natural wastage as a policy of labour reduction.

This discontinuity may also be generated by a size effect: if very small firms have created the majority of jobs, then the change we observe only beyond a certain threshold may be the result of indivisibilities. By contrast, on the left side of the distribution we may not be able to see the same discontinuities if the destruction of jobs is mainly generated by large companies to which the problem of indivisibilities does not apply. In Appendix 2.4. we present the plots of the rate of change in employment by size. Size is divided into five categories: 1-19, 20-49, 50-99, 100-199 and 200-500 employees. By looking at those figures, we notice first of all the "indivisibilities" which occur when companies have a very small size. For five employee companies, for example, hiring or firing a person means an increase/decrease in its workforce by 20 percent. This explains why the percentage variations in employment for the category containing very small firms are more scattered around the horizontal axis. However, even in the case of small firms, the gap between the peak at zero adjustment and positive or negative net changes in employment tends to be wider on the right hand side. As mentioned before, we believe that the left hand side may hide a number of movements, like quits, which, at least in principle, involve voluntary decisions by the workers and not by the firm. The gap becomes more filled in the case of companies with more than 100 employees, where it is clear that the possibilities of smoothing are larger. Nevertheless, the gap does not entirely disappear.

All these findings prompted us to investigate whether the companies which do not adjust show any recurrent pattern in their outputs. We did not find any
behaviour of this sort. Rather, these firms exhibit a wide range of output variations.

As a further control we have divided the sample into three groups: companies with respectively zero, positive or negative changes in employment. We then, calculated the average rate of change in employment one period ahead, and the average change in sales and real wages one period lagged. Tables VI-VIII show the results. Table VI shows, for each pair of years, the number of companies which did not change employment at all. In Table VII we also distinguish between average positive and average negative changes in employment. If adjustment costs are not simply quadratic, we should expect larger employment variations at \( t+1 \) for firms which did not adjust employment at time \( t \) and smaller changes for firms which adjusted in both periods. This happens quite clearly in column (2) for positive changes at time \( t+1 \). The contrast between change at \( t+1 \) after no change at \( t \) and change over both periods is less sharp in column (3) which considers negative variations in employment. As we have already noticed, these numbers also include voluntary quits and retirements making employment adjustment on the negative side more complex to measure.

<table>
<thead>
<tr>
<th>Year</th>
<th>( \Delta E_t=0 )</th>
<th>( \Delta E_{t+1}=0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>82-83</td>
<td>913</td>
<td>270</td>
</tr>
<tr>
<td>83-84</td>
<td>640</td>
<td>209</td>
</tr>
<tr>
<td>84-85</td>
<td>717</td>
<td>225</td>
</tr>
<tr>
<td>85-86</td>
<td>735</td>
<td>207</td>
</tr>
<tr>
<td>86-87</td>
<td>648</td>
<td>186</td>
</tr>
<tr>
<td>87-88</td>
<td>587</td>
<td>171</td>
</tr>
</tbody>
</table>

\( E = \text{employment} \)
Table VII Employment change

<table>
<thead>
<tr>
<th></th>
<th>ΔE,</th>
<th>no.</th>
<th>ΔE,&gt;0</th>
<th>no.</th>
<th>ΔE,&lt;0</th>
<th>no.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(mean)</td>
<td></td>
<td>(mean)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>82-83</td>
<td>=0</td>
<td>913</td>
<td>23.44</td>
<td>300</td>
<td>-11.52</td>
<td>347</td>
</tr>
<tr>
<td></td>
<td>&gt;0</td>
<td>978</td>
<td>15.84</td>
<td>410</td>
<td>-10.25</td>
<td>387</td>
</tr>
<tr>
<td></td>
<td>&lt;0</td>
<td>1354</td>
<td>17.58</td>
<td>455</td>
<td>-10.01</td>
<td>708</td>
</tr>
<tr>
<td>83-84</td>
<td>=0</td>
<td>640</td>
<td>19.66</td>
<td>241</td>
<td>-10.82</td>
<td>191</td>
</tr>
<tr>
<td></td>
<td>&gt;0</td>
<td>1171</td>
<td>13.21</td>
<td>556</td>
<td>-9.90</td>
<td>503</td>
</tr>
<tr>
<td></td>
<td>&lt;0</td>
<td>1443</td>
<td>14.40</td>
<td>503</td>
<td>-9.21</td>
<td>708</td>
</tr>
<tr>
<td>84-85</td>
<td>=0</td>
<td>717</td>
<td>13.89</td>
<td>255</td>
<td>-10.80</td>
<td>239</td>
</tr>
<tr>
<td></td>
<td>&gt;0</td>
<td>1300</td>
<td>11.50</td>
<td>580</td>
<td>-9.14</td>
<td>434</td>
</tr>
<tr>
<td></td>
<td>&lt;0</td>
<td>1239</td>
<td>13.74</td>
<td>439</td>
<td>-9.80</td>
<td>573</td>
</tr>
<tr>
<td>85-86</td>
<td>=0</td>
<td>735</td>
<td>16.37</td>
<td>314</td>
<td>-10.45</td>
<td>214</td>
</tr>
<tr>
<td></td>
<td>&gt;0</td>
<td>1274</td>
<td>11.99</td>
<td>678</td>
<td>-9.32</td>
<td>378</td>
</tr>
<tr>
<td></td>
<td>&lt;0</td>
<td>1248</td>
<td>13.10</td>
<td>459</td>
<td>-9.26</td>
<td>1248</td>
</tr>
<tr>
<td>86-87</td>
<td>=0</td>
<td>662</td>
<td>18.73</td>
<td>279</td>
<td>-13.03</td>
<td>196</td>
</tr>
<tr>
<td></td>
<td>&gt;0</td>
<td>1452</td>
<td>14.04</td>
<td>750</td>
<td>-9.97</td>
<td>436</td>
</tr>
<tr>
<td></td>
<td>&lt;0</td>
<td>1144</td>
<td>14.22</td>
<td>462</td>
<td>-9.75</td>
<td>489</td>
</tr>
<tr>
<td>87-88</td>
<td>=0</td>
<td>648</td>
<td>16.25</td>
<td>289</td>
<td>-11.95</td>
<td>186</td>
</tr>
<tr>
<td></td>
<td>&gt;0</td>
<td>1491</td>
<td>12.01</td>
<td>776</td>
<td>-10.44</td>
<td>471</td>
</tr>
<tr>
<td></td>
<td>&lt;0</td>
<td>1120</td>
<td>17.21</td>
<td>516</td>
<td>-9.78</td>
<td>432</td>
</tr>
</tbody>
</table>

E = employment

Table VIII illustrates the average rate of change of real output and wages in the period before the employment decision. As we expect, firms which on average decreased employment exhibit a lower rate of increase in sales. By contrast, firms which expand employment are characterized throughout the period by the highest rate of increase in output. It is interesting to notice, instead, how companies which do not vary their number of workers are on average characterized by differing patterns of sales in different years: they are nearly in the exact middle between positive and negative variation in 1982-84, very close to firms with positive
Table VIII

<table>
<thead>
<tr>
<th></th>
<th>ΔE_t</th>
<th>ΔQ_t</th>
<th>no.</th>
<th>ΔW_t</th>
</tr>
</thead>
<tbody>
<tr>
<td>83-84</td>
<td>=0</td>
<td>17.17</td>
<td>638</td>
<td>6.18</td>
</tr>
<tr>
<td></td>
<td>&gt;0</td>
<td>22.07</td>
<td>1165</td>
<td>11.52</td>
</tr>
<tr>
<td></td>
<td>&lt;0</td>
<td>12.37</td>
<td>1442</td>
<td>4.40</td>
</tr>
<tr>
<td>84-85</td>
<td>=0</td>
<td>21.89</td>
<td>715</td>
<td>5.75</td>
</tr>
<tr>
<td></td>
<td>&gt;0</td>
<td>26.92</td>
<td>1300</td>
<td>6.39</td>
</tr>
<tr>
<td></td>
<td>&lt;0</td>
<td>16.36</td>
<td>1238</td>
<td>2.06</td>
</tr>
<tr>
<td>85-86</td>
<td>=0</td>
<td>19.72</td>
<td>734</td>
<td>5.14</td>
</tr>
<tr>
<td></td>
<td>&gt;0</td>
<td>21.66</td>
<td>1274</td>
<td>8.29</td>
</tr>
<tr>
<td></td>
<td>&lt;0</td>
<td>12.65</td>
<td>1246</td>
<td>3.80</td>
</tr>
<tr>
<td>86-87</td>
<td>=0</td>
<td>9.47</td>
<td>662</td>
<td>3.36</td>
</tr>
<tr>
<td></td>
<td>&gt;0</td>
<td>14.06</td>
<td>1451</td>
<td>3.29</td>
</tr>
<tr>
<td></td>
<td>&lt;0</td>
<td>8.27</td>
<td>1144</td>
<td>1.58</td>
</tr>
<tr>
<td>87-88</td>
<td>=0</td>
<td>10.41</td>
<td>646</td>
<td>6.42</td>
</tr>
<tr>
<td></td>
<td>&gt;0</td>
<td>11.98</td>
<td>1491</td>
<td>8.50</td>
</tr>
<tr>
<td></td>
<td>&lt;0</td>
<td>6.53</td>
<td>1121</td>
<td>4.56</td>
</tr>
<tr>
<td>88-89</td>
<td>=0</td>
<td>15.67</td>
<td>587</td>
<td>4.28</td>
</tr>
<tr>
<td></td>
<td>&gt;0</td>
<td>17.70</td>
<td>1581</td>
<td>7.12</td>
</tr>
<tr>
<td></td>
<td>&lt;0</td>
<td>11.50</td>
<td>1089</td>
<td>1.98</td>
</tr>
</tbody>
</table>

Q = real output
W = real wage

change in 1984-85 and 1986-89, while the opposite happens in 1985-86. It seems, therefore, that there is no consistent behaviour in the rate of change in sales across companies which do not vary employment.

In the same way, when looking at the rate of change in real wages in the previous period we should expect that non-adjusting firms experience a lower increase in wages than other companies. This only partially appears in Table VIII, but it is possible that the figures in the table are the most strongly affected by fixed effects. If employment and wages are jointly determined, as in the case of bargaining between unions and employers, then firms which increase employment will experience larger increases in wages; and firms which reduce their number of workers will show a reduction in wage costs. This appears to be the case in this table.
In the following figures we maintain the same division into companies which show zero, positive or negative net change in employment and plot the averages of the variables we are going to use to determine firms' labour decisions. These are real output, real wages, productivity, profitability and new investment.

Figure 27 Output

As far as output, measured by real sales, is concerned it is interesting to notice how the firms which experience negative net variations in employment show a level of output higher than the ones which do not vary their manpower. This could imply
that companies with zero changes may be experiencing labour hoarding while the companies with negative variations are undertaking substantial restructuring. In the case of companies with positive net changes we may be picking up a fixed effect: firms which perform well experience both increases in output and employment.

Figure 28 reveals how the process of adjustment is also affected by the level of wages. Firms which exhibit positive adjustment show less variable dynamics. The opposite happens to the companies which reduce employment in the same year,
Figure 29 Productivity

while real wages for the no change firms lie in the middle. Productivity in Figure 29, measured as output per employee, is clearly higher for companies with negative changes in employment, and this is reinforced by the fact that, as we saw from Figure 27, these firms exhibit the highest level of average output. The pattern of productivity growth is more similar for the companies with zero and positive changes. The interesting feature of Figure 30 relates to these firms with zero net changes in employment which show the lowest average level of profitability. Again,
as pointed out in the previous pages, this may imply some level of labour hoarding.

Figure 30 Profitability

The last figure on new investment, Figure 31, confirms the same point of view. Firms with zero net changes in the level of employment exhibit the lowest level of new investment, consistent with stationery situation likely to be characterized by labour hoarding.
2.4. Interpretation of the "zeros"

In this section we look for possible explanations of the existence of the zeros. These explanations need not necessarily refer to a fixed costs framework. In particular, we consider the effects of the wage subsidies, the Cassa Integrazione Guadagni (CIG) and its two principal components, the Ordinary Scheme or "Cassa
Integrazione Ordinaria" (CIGO) and the Extraordinary Scheme or "Cassa Integrazione Straordinaria" (CIGS); turnover; and measurement errors. The Wage Supplementation Scheme or Cassa Integrazione Guadagni was described in more detail in Chapter 1, in particular in Appendix 1.1.

2.4.1. Wage Supplementation Scheme

Could zero adjustment hide subsidized excess labour rather than a decision not to adjust? The practice of counting employees working shorter time or temporarily laid-off and covered by the Wage Supplementation Scheme as employed, until they are either fired or reabsorbed, makes it more difficult to interpret firm’s redundancies policies. As mentioned in the previous chapter, while some lay-offs represent real labour hoarding, some others may be interpreted as hidden firings as the results of limits imposed by very restrictive regulation.

To disentangle labour hoarding from subsidized excess labour, we have matched our company data with the data on wage supplementation (the source is INPS, the institute administering the scheme) to find out in which cases a zero change in employment is in fact hiding subsidized excess labour. If we look at CIGS, the special component of the programme, we find that, each year, only about

<table>
<thead>
<tr>
<th>Year</th>
<th>Extr.</th>
<th>Ord.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1983</td>
<td>2.2</td>
<td>4.7</td>
</tr>
<tr>
<td>1984</td>
<td>2.0</td>
<td>6.4</td>
</tr>
<tr>
<td>1985</td>
<td>1.4</td>
<td>7.8</td>
</tr>
<tr>
<td>1986</td>
<td>1.6</td>
<td>6.8</td>
</tr>
<tr>
<td>1987</td>
<td>1.5</td>
<td>4.5</td>
</tr>
<tr>
<td>1988</td>
<td>1.8</td>
<td>4.5</td>
</tr>
<tr>
<td>1989</td>
<td>1.9</td>
<td>1.8</td>
</tr>
</tbody>
</table>
2 percent of the companies which do not change employment are in fact using the scheme. Table IX reports the percentages for both schemes. The zero variations in employment do not appear to be a consequence of the Ordinary Scheme either, being, on average, around 5 per cent. These calculations are based on matching rules that we have adopted to establish a correct correspondence between a zero variation and a subsidy. In fact, to assess whether a zero change in employment is the effect of a subsidy is complicated because the legislation induces a number of time lags and spillover effects across years. It is therefore incorrect to accept as matchings only the exact correspondences between zero changes and CIG authorization in particular years. For these reasons we have adopted a broader matching rule that allows us to capture such lags. The year in which companies appear on the INPS dataset corresponds to the authorization of the scheme. Suppose a company is granted CIG at the end of 1984: we may find no variation in employment in 1985 which is certainly the effect of CIG in 1984, but this would not be captured by the matching and could result in underestimation of the cases in 1985. On the other hand we may also be induced to overestimate: if the company shows no variation in 1984, the zero will be attributed to the authorization at the end of 1984. There are also a few cases in which the request for renewal is disputed and the authorization delayed. When the final approval comes, it is attributed for bureaucratic reasons to the year of the initial application rather than to the year of actual authorization. We have dealt with this problem by considering the average delay time to be one year. For instance, if a company applies for CIG in 1984 and the authorization is delayed to 1985, CIG will still be registered and appear in our statistics in 1984. But the zero adjustment may take place in 1985.
along with the subsidy generating a mismatch that we have tried to take into account. It is interesting to note that the series for CIGO reflects the actual cyclical behaviour of the Ordinary Scheme in aggregate, with a peak in 1985 and a strong slow down at the end of the sample period. Looking at all the companies under the scheme, we find that they tend to adjust employment downward rather than keep it constant. It is, therefore, difficult to relate the zeros to the functioning of this programme.

Furthermore, the relative variability in employment of companies in CIG over the period and the negligible percentage of subsidized companies among those which do not adjust reinforces our view that there is not much of a relation between the zero changes and the operation of the programme. In summary, the analysis of the CIG dataset suggests that the regulation requiring firms to hold employment constant during the period of benefit does not apply in many cases. Evidently many firms use CIG to modify their skill composition, in which case they may simultaneously hire and lay-off while on the programme.

2.4.2. Turnover

Zero changes in employment may also hide some degree of turnover which occurred during the year. In other words, the zero changes represent the net sum of people who have left and their replacements. These may either coincidentally cancel leaving a zero net-change or may cancel to give a planned net zero adjustment in the case the firm is trying to maintain the number of workers unchanged. Unfortunately we do not have information on turnover and voluntary
quits which would represent an addition of very useful information to the dataset. We, thus, consider cost structures as relating to net changes in employment, an approach taken in some but not all the literature, focusing on decisions on jobs rather than turnover of people. Our approach takes into account not only hiring and firing costs but also technological constraints. We are, nevertheless, aware of the importance of gross adjustment costs and the necessity of comparing, when information is available, the two types of cost (net and gross) in analyzing dynamic factor demand. The importance of distinguishing between gross and net adjustment costs is stressed in Hamermesh (1993b) and Hamermesh, Hassink and van Ours (1994). Both studies provide a clear distinction between these two types of costs and emphasize that macroeconomic fluctuations may have substantial effects beyond those indicated by net employment changes at firms level. Clearly, when we consider net changes in employment we refer to a highly aggregative measure. As the works mentioned above indicate, it is necessary to distinguish between net changes in employment, job creation, job destruction and labour turnover, concepts which are often used in the literature as interchangeable but which imply different decisions and constraints. The first two concepts refer to the jobs while the third to people. If we indicate net changes in employment by $\Delta E$, we have

$$\Delta E \leq [\Delta E^* + \Delta E'] \leq H + F$$

where

$[\Delta E^* + \Delta E']$, the job turnover, may be divided into:

$\Delta E^* = \text{job creation} = \text{new hires + rehires}$; and

$\Delta E' = \text{job destruction} = \text{quits + collective fires + individual dismissals + lay-offs}$; while
H + F = total hires + separations.

Clearly, ΔE = [ΔE⁺ - ΔE⁻] = H - F. It may follow that firms characterized by a zero net variation in employment may have at the same time a high rate of job creation and job destruction and an even higher rate of hiring and firing. Likewise, we may not be able to infer from our data that firms which show a net increase in employment may have undertaken a high number of firings. The more we disaggregate turnover, by including for instance promotions, transfers to and from other plants in the firm etc., the higher the turnover which we obtain. The information we may find on the processes which generate adjustment costs increases with the degree of disaggregation of the flows reported above. Unfortunately, it is very hard to find such detailed information on job turnover.

Clearly, each of the flows which together comprise turnover is affected by the structure of adjustment costs. Firms react to exogenous shocks by activating one channel or another, according to the constrains they face. For instance, in the presence of fixed costs we may observe collective firings being used as a way of reducing the manpower. The perception of the shock as transitory or permanent also determines the size of the adjustment. If fixed costs are high and shocks transitory we may not observe any movement in the firm’s labour input. ³⁹

³⁹ This is an objection to Contini, Gavosto, Revelli and Sestito (1992), and Gavosto and Sestito (1992, 1994). The authors analyze a sample of about 28,000 Italian firms for the period 1985-89. They find a yearly rate of job turnover, measured as ΔE⁺+ΔE⁻, of about 20 per cent. They conclude that the fact that the job turnover in Italy is in line with what we observe in other OECD countries shows how it is not clear that Italy faces, as it is claimed, particularly rigid labour market legislation. Our view is that this is only very indirect evidence of the rigidity (or lack of) of adjustment costs. It is important to know to what flows of adjustment firms choose to resort to in order to vary their employment levels. For instance, in a world of idiosyncratic companies, in the presence of shocks we may observe a high rate of job turnover if companies use collective firings as a way of reducing the
The discussion above also relates to the problem raised by Hamermesh (1992) about what temporally aggregated data can tell us about the structure of adjustment costs. Errors may arise from using data at survey intervals instead of measuring the state and control variables at the time interval that is relevant for the firm. Even if we know that in considering the appropriate degree of spatial aggregation the firm is the relevant decision-making unit we may yet not be sure about the time intervals between firms' decisions on whether to alter employment. Hamermesh claims that if we wish to draw inferences about the structure of adjustment costs we need to obtain microdata that are temporally disaggregated at least to quarterly observations. Annual observations would typically suggest smooth adjustment.

We agree that the frequency of employment decisions is an interesting issue but consider that, in our case, annual data may nevertheless be informative in relation to the structure of costs. Certainly, the tighter are institutional constraints, the longer it takes to alter the labour force. Consider unfair dismissals or apprenticeship contracts. The length of the observation interval deemed necessary may also be related to the size of the company. Large firms are likely to revise employment more frequently than once a year. These companies may forecast and manage product demand better and they are likely not to face the same technological constraints as smaller firms. Therefore, we doubt that there is a definitive evidence that only high frequency data can explain the structure of adjustment costs.

number of employees. This is typically a consequence of high fixed costs of firing. (see Rota, 1994)
2.4.3. Measurement errors

Measurement errors which include coding and response errors are a typical problem with survey data. In our context companies could misreport information, or in the case of the number of employees, they may round the figures or report the same number of employees as the previous year. This may, in some cases, affects the validity of the zeros as indicators of non-adjustment. We will deal with measurement errors in the next chapter when we look at the determinants of the selection process characterizing our sample of companies.

2.5. Transition probabilities

In this section we consider labour adjustment as transitions between states. We may define three states, two of which are defined as adjustment: adjustment upward, or net hiring, and adjustment downward, or net firing and the third defined as zero net adjustment. We thus represent the firm’s decision problem of whether or not to change employment as a time separable Markov decision process. The problem may be stated in the following terms: an employment decision is observed at time points, t,...,T, to be in one of the three possible states S = A, NA, representing the choice of whether or not to adjust. After observing the state of the process, the firm chooses an action. When the process is, for instance, in state $A_{t-1}$, at the beginning of period t, corresponding to the firm’s decision to adjust employment last period, the firm chooses an action $d_t$ and (i) it receives an expected profit $\Pi_t(A_{t-1},d_t)$ and (ii) the firm chooses the next state of the system according to transition probabilities, $p_t^{[A,NA]}(d_t)$ defined as
where i and j denote respectively the origin and the destination state. Transition probabilities are, thus, only functions of the current state and action. The transition probability may thus be written as:

\[ p_i[t = j | L_0, d_0, L_1, d_1, \ldots, L_{t-1} = i, d_t = 0, 1] = p_i(d_t) \]  

That is the conditional probability of any future event, given any past event and the present state, is independent of past events. We are, thus, representing firm's problem in a Markovian framework in which the transition probability and the reward from a given choice depend only on the current state of the system and the current decision; it is not necessary to keep track of the entire previous history of the system. We will further analyze Markov processes in Chapters 5. and 6.

Consider the three adjustment regimes (states) defined above, between which the firm can choose: positive, negative and zero adjustment. Does the direction of employment adjustment in period t-1 affect labour changes in period t? That is, we are interested in finding out whether being in a state one period affects the probabilities associated with a different state in the subsequent period, generating state dependence. As is shown in Figure 32, at (t-1) the company can adjust its employment either upwards (+), downwards (-) or not adjust at all (0).

The overall probability associated with each regime is given by the sum of the conditional probabilities. For instance, the probability of adjusting upwards, \( p_i(+) \), is: \( p_i(0), \) the probability of not adjusting, and \( p_i(-) \), the probability of adjusting downwards, follow accordingly]
The probability of adjusting positively at period $t+1$ is given by the sum of the probabilities of adjusting positively conditional on having adjusted respectively positively, $pr(+,)$, non adjusting, $pr(0,)$ and adjusting negatively, $pr(-,)$, the period before. If there is state dependence, i.e., it matters which regime the company has started from, then we should expect these conditional probabilities to be significantly different from each other. If there is independence, the conditional probability of adjusting, for example upwards is simply

$$Pr(\cdot_{t+1} | i_{t-1}) = Pr(\cdot_{t+1})$$

where the conditioning terms in (3) have cancelled out. $i$ specifies any possible state characterizing the company at $t-1$. In other words the adjustment at $t$ is not influenced by the occurrence of a certain regime during the previous period. In this case just to look at (4) is informative.
We may summarize all the possible combinations of origin and destinations states by means of a 3x3 transition matrix as the following:

\[
\begin{pmatrix}
  p_{i1}(t) & p_{i2}(t) & p_{i3}(t) \\
  p_{1j}(t) & p_{2j}(t) & p_{3j}(t) \\
  p_{ij}(t) & p_{ij}(t) & p_{ij}(t)
\end{pmatrix}
= \begin{pmatrix}
  p_{11}(t) & p_{12}(t) & p_{13}(t) \\
  p_{21}(t) & p_{22}(t) & p_{23}(t) \\
  p_{31}(t) & p_{32}(t) & p_{33}(t)
\end{pmatrix}
\] (5)

The elements of the matrix are the transition probabilities, \( p_{jk}(t) \), and should be interpreted as the probability that the \( i \)th firm is in state \( k \) at time \( t \) given that it was in state \( j \) at time \( t-1 \), where \( k = 1, 2, 3; j = 1, 2, 3; \) and \( i = 1, 2, \ldots, N \). As we will explain in more detail in Chapters 5 and 6, this Markov structure implies that employment changes in the presence of exogenous shocks are simply determined by \( p(.) \). In fact, given current realizations of shocks, future realizations are independent of the past.

The empirical transition matrices over the sample period for the entire sample are given in Table X. Table XI reports the standard error and Table XII the sample frequencies. Each element of the transition matrix corresponds to:

\[
p_{jk} = \frac{n_{jk}}{\sum_{k=1}^{3} n_{jk}}
\] (6)

\( p(t) \) is, thus, a stochastic matrix with nonnegative elements and rows which sum to unity.

From (6) we have

\[
p_{jk} = \frac{n_{jk}}{\sum_{k=1}^{3} n_{jk}} \sim N(\bar{p}_{jk}p_{jk}(1-p_{jk}))
\] (7)

and the standard errors are calculated as
\[ \sqrt{\frac{p_{jk}(1-p_{jk})}{\sum_k n_{jk}}} \]  

We also provide a visualization of these probabilities in Figures 33 and 34. These are 3-dimensional charts where the frequencies of positive, zero and negative variations for each pair of years are represented by vertical bars. Figure 33 shows the transition over the interval 1982-86 and Figure 34 over the interval 1985-89. Each group of nine vertical bars represents exactly the information contained in Table XII. For example: pos82-83 indicates positive changes between year 1982 and 1983, etc. One dimension is given by the origin state, another by the destination and a third by the frequency.

In Appendix 2.4, we report the transition probabilities calculated by industry.
Table X Transition probabilities

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th></th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(83-84)</td>
<td>+ 0 -</td>
<td>+ 0 -</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ 0.42 0.18 0.39</td>
<td>+ 0.47 0.20 0.43</td>
<td></td>
</tr>
<tr>
<td>(82-83)</td>
<td>0 0.33 0.29 0.38</td>
<td>0 0.38 0.32 0.29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- 0.33 0.14 0.52</td>
<td>- 0.35 0.46 0.19</td>
<td></td>
</tr>
<tr>
<td>(85-86)</td>
<td>+ 0 -</td>
<td>+ 0 -</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ 0.44 0.22 0.33</td>
<td>+ 0.53 0.17 0.29</td>
<td></td>
</tr>
<tr>
<td>(84-85)</td>
<td>0 0.35 0.31 0.33</td>
<td>0 0.42 0.28 0.29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- 0.35 0.18 0.46</td>
<td>- 0.36 0.19 0.44</td>
<td></td>
</tr>
<tr>
<td>(87-88)</td>
<td>+ 0 -</td>
<td>+ 0 -</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ 0.51 0.18 0.30</td>
<td>+ 0.52 0.16 0.31</td>
<td></td>
</tr>
<tr>
<td>(86-87)</td>
<td>0 0.42 0.28 0.29</td>
<td>0 0.44 0.26 0.29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- 0.40 0.16 0.42</td>
<td>- 0.46 0.15 0.38</td>
<td></td>
</tr>
</tbody>
</table>
Table XI Standard errors

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th></th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.015</td>
<td>.012</td>
<td>.015</td>
</tr>
<tr>
<td></td>
<td>.015</td>
<td>.015</td>
<td>.016</td>
</tr>
<tr>
<td></td>
<td>.019</td>
<td>.009</td>
<td>.013</td>
</tr>
<tr>
<td></td>
<td>.012</td>
<td>.011</td>
<td>.013</td>
</tr>
<tr>
<td></td>
<td>.013</td>
<td>.010</td>
<td>.012</td>
</tr>
<tr>
<td></td>
<td>.019</td>
<td>.017</td>
<td>.018</td>
</tr>
<tr>
<td></td>
<td>.011</td>
<td>.011</td>
<td>.014</td>
</tr>
<tr>
<td></td>
<td>.013</td>
<td>.010</td>
<td>.012</td>
</tr>
<tr>
<td></td>
<td>.019</td>
<td>.017</td>
<td>.018</td>
</tr>
<tr>
<td></td>
<td>.014</td>
<td>.011</td>
<td>.014</td>
</tr>
</tbody>
</table>

|   | (3)       |       | (4)       |
|   | .014      | .012  | .014      |
|   | .019      | .018  | .018      |
|   | .012      | .013  | .013      |
|   | .018      | .016  | .016      |
|   | .013      | .011  | .014      |

|   | (5)       |       | (6)       |
|   | .013      | .010  | .012      |
|   | .019      | .017  | .018      |
|   | .014      | .011  | .014      |
|   | .013      | .009  | .012      |
|   | .019      | .017  | .019      |
|   | .015      | .011  | .014      |
Table XII Frequencies

<table>
<thead>
<tr>
<th></th>
<th>Tot.</th>
<th>Tot.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>410</td>
<td>181</td>
</tr>
<tr>
<td>(2)</td>
<td>300</td>
<td>266</td>
</tr>
<tr>
<td>(3)</td>
<td>455</td>
<td>191</td>
</tr>
<tr>
<td>(3)</td>
<td>(4)</td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td>580</td>
<td>286</td>
</tr>
<tr>
<td>(5)</td>
<td>255</td>
<td>221</td>
</tr>
<tr>
<td>(6)</td>
<td>439</td>
<td>227</td>
</tr>
<tr>
<td>(5)</td>
<td>(6)</td>
<td></td>
</tr>
<tr>
<td>(6)</td>
<td>750</td>
<td>266</td>
</tr>
<tr>
<td>(7)</td>
<td>279</td>
<td>187</td>
</tr>
<tr>
<td>(8)</td>
<td>462</td>
<td>193</td>
</tr>
</tbody>
</table>
Figure 33

Transitions 1982-86

destination

neg84-85
zero84-85
pos84-85
neg83-84
zero83-84
pos83-84
neg82-83
zero82-83
pos82-83

frequency
Figure 34

Transitions 1985-1989
In order to test for state dependence we have implemented a $\chi^2$ test for the equality of the conditional and unconditional probabilities. The test is based on the comparison between the actual and the predicted probabilities. The null hypothesis implies that all conditional probabilities are equal to the corresponding unconditional probabilities, i.e. $P_{ij}(t)=P_i(t)$, $j=1,2,3$ and $i=1,2,3$. No matter what origin state, the probability of adjusting upwards, downwards or not to adjust at all is the same. The actual probabilities are simply given by the sample frequencies as in the above tables. The test is given by

$$\sum_i \frac{(A_i-P_i)^2}{P_i} \sim \chi^2\text{(4)}$$

where (4) indicates the degrees of freedom. The value of the test are reported in the following table. The numbers in brackets refer to the matrices as they appear in Tables X-XII above. All the values imply a rejection of the null hypothesis of absence of state dependence. The $\chi^2$ tests for the equality of the transitions probabilities at industry levels are presented in Appendix 2.5, Table A.2.18. We find that, also at industry level the test fails to accept the null hypothesis.

We can evaluate the t-steps transition probabilities. $p^{(t)}$ may be calculated by multiplying the matrix $p$ by itself $t$ times. If $\lim p^{(t)}$ exists as $t$ goes to infinity, then
and the elements of $p(\infty)$ are the equilibrium probabilities. The values of the equilibrium probabilities for the transition matrices in Table X are presented in the following table.

Table XIV Equilibrium probabilities

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.36</td>
<td>0.17</td>
<td>0.47</td>
</tr>
<tr>
<td>2</td>
<td>0.39</td>
<td>0.25</td>
<td>0.36</td>
</tr>
<tr>
<td>3</td>
<td>0.38</td>
<td>0.23</td>
<td>0.39</td>
</tr>
<tr>
<td>4</td>
<td>0.44</td>
<td>0.14</td>
<td>0.42</td>
</tr>
<tr>
<td>5</td>
<td>0.45</td>
<td>0.19</td>
<td>0.36</td>
</tr>
<tr>
<td>6</td>
<td>0.48</td>
<td>0.21</td>
<td>0.31</td>
</tr>
</tbody>
</table>

Table XIV describes thus the long-run behaviour of the system. Each row vector defines the invariant distribution of each of the transition matrices shown in Table X. These are also represented in Figure 35. For all initial probability distributions each sequence $\{p'_t\}$ converges to the invariant distribution given by a 3x3 limit matrix with all the rows equal to each other, given by the corresponding vectors shown in the table. This can be seen as following:

$$p[X_0 = i_0, X_1 = i_1, \ldots, X_t = i_t]$$

$$= p[X_t = i_t | X_0 = i_0, \ldots, X_{t-1} = i_{t-1}] \cdot p[X_0 = i_0, \ldots, X_{t-1} = i_{t-1}]$$

$$= p_{(t-1), (t, t)} \cdot p[X_0 = i_0, \ldots, X_{t-1} = i_{t-1}]$$

$$= p_{(t-2), (t-1), (t-1)} \cdot p[X_0 = i_0, \ldots, X_{t-2} = i_{t-2}]$$

$$\vdots$$

$$= p_{(t-1), \ldots, (t-1)} \cdot p_{(t-2), \ldots, (t-1)} \cdot \ldots \cdot p_{(0), (1)} \cdot p[X_0 = i_0]$$

If we represent the probability that the process goes from state $i$ to state $j$ in $t$ steps, $p_{ij}^t$.
as

\[ p_{ij}^t = P[X_{t+m} = j \mid X_m = i] \]  \hspace{1cm} (12)

For any \( r < t \)

\[ p_{ij}^t = \sum_{k=0}^{\infty} p_{ik}^r p_{kj}^{t-r} \]  \hspace{1cm} (13)

If we let \( p^{(t)} \) denote the matrix of \( t \)-step transition probabilities \( p_{ij}^t \), then (13) corresponds to \(^4^0\):

\[ p^{(t)} = p^{(r)} * p^{(t-r)} \]  \hspace{1cm} (14)

and

\[ p^{(t)} = p*p^{(t-1)} = p*p*p^{(t-2)} = \ldots = p^t \]  \hspace{1cm} (15)

Hence, \( p^{(t)} \) may be calculated by multiplying the matrix \( P \) by itself \( t \) times. The behaviour through time of \( X \) is simply defined by the transition probabilities.

\(^4^0\) Eq. (14) is known as the Chapman-Kolmogorov equation.
Equilibrium probabilities

Figure 35
We will consider transition probabilities again in Chapter 5. and 6. when we examine their role in a dynamic structural model. There, transition probabilities embody the main dynamics in the problem and give the probability of movements of the system from one stage to another within a framework of optimal sequential decision making. In that context, transition probabilities represent firms' subjective beliefs about uncertain future states represented by price and wage uncertainty. In the present chapter they provide a more limited role of data description.
Appendix 2.1. Variable specification

The dataset contains information on firms located in Lombardia. When companies are multiplant with some of the plants outside the district data refer to the whole productive unit. Only about 20 companies were quoted in 1982. This number could have slightly increased during the sample period due to the growth of some companies. All data are annual over the period 1982-89.

**Total employment:** number of productive and non productive workers. It includes temporary lay-offs. Companies report the average annual employment (calculated as the sum of monthly employment divided by twelve).

**Output:** real sales (deflated by the value added deflator for manufactures). We prefer sales to value added since they do not include inventories.

**Wages:** the wage-bill divided by total employment deflated by value added price.

**Value added price:** value added deflator for manufactures. Source: Banca d'Italia, 1990, Tav. aB 7.

**Productivity:** real sales divided by total employment.

**Gross profits:** value added minus wages, salaries and all personnel costs (included the mandatory fund for severance payments). This variable is deflated by value added price and scaled by employment in 1982.

**Investments in new machinery:** new investments in equipments and machinery deflated by value added price and scaled by employment in 1982.

**Size:** reciprocal of employment.
Output, wages, gross profits, investment in new machinery are in millions of lira.

All variables except for value added price are at firm level.

**Industries: two digit level (three digit in brackets)**

21 ferrous metals (211-224)  
23 non ferrous metals (231-248)  
25 chemicals and artificial fibres (251-260)  
31 metal manufacturing (311-319)  
32 machinery (321-328)  
33 electrical equipment (330-348)  
35 transport equipment (351-365)  
37 instrument manufacturing (371-374)  
41 food, drink and tobacco (411-429)  
43 textiles (431-439)  
44 leather goods (441-456)  
46 timber and furniture (461-467)  
47 paper and printing (471-474)  
48 rubber manufacturing (481-483)  
49 others (491-496)

one digit level:

industry 1 = 21+23+25  
industry 2 = 31+32+33+35+37  
industry 3 = 41+43+44+46+47+48+49

In order to avoid mergers and take-overs we have omitted from the dataset companies characterized by 500% increases in the rate of change of employment and/or sales.
### Table A.2.3. Average output across manufacturing industries

<table>
<thead>
<tr>
<th>Year</th>
<th>Ind. 1</th>
<th>Ind. 2</th>
<th>Ind. 3</th>
<th>All</th>
<th>No. of Firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1982</td>
<td>94.50</td>
<td>578</td>
<td>53.19</td>
<td>1345</td>
<td>66.59</td>
</tr>
<tr>
<td>1983</td>
<td>97.34</td>
<td>578</td>
<td>52.07</td>
<td>1345</td>
<td>66.54</td>
</tr>
<tr>
<td>1984</td>
<td>106.9</td>
<td>578</td>
<td>56.07</td>
<td>1345</td>
<td>71.56</td>
</tr>
<tr>
<td>1985</td>
<td>111.5</td>
<td>578</td>
<td>60.76</td>
<td>1345</td>
<td>74.43</td>
</tr>
<tr>
<td>1986</td>
<td>105.8</td>
<td>578</td>
<td>62.57</td>
<td>1345</td>
<td>75.27</td>
</tr>
<tr>
<td>1987</td>
<td>107.8</td>
<td>578</td>
<td>66.61</td>
<td>1345</td>
<td>78.23</td>
</tr>
<tr>
<td>1988</td>
<td>124.4</td>
<td>578</td>
<td>73.36</td>
<td>1345</td>
<td>83.01</td>
</tr>
<tr>
<td>1989</td>
<td>131.6</td>
<td>578</td>
<td>79.38</td>
<td>1345</td>
<td>87.08</td>
</tr>
</tbody>
</table>

### Table A.2.4. Average gross profits across manufacturing industries

<table>
<thead>
<tr>
<th>Year</th>
<th>Ind. 1</th>
<th>Ind. 2</th>
<th>Ind. 3</th>
<th>All</th>
<th>No. of Firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1982</td>
<td>0.078</td>
<td>578</td>
<td>0.048</td>
<td>1345</td>
<td>0.049</td>
</tr>
<tr>
<td>1983</td>
<td>0.066</td>
<td>578</td>
<td>0.038</td>
<td>1345</td>
<td>0.042</td>
</tr>
<tr>
<td>1984</td>
<td>0.063</td>
<td>578</td>
<td>0.038</td>
<td>1345</td>
<td>0.043</td>
</tr>
<tr>
<td>1985</td>
<td>0.064</td>
<td>578</td>
<td>0.038</td>
<td>1345</td>
<td>0.042</td>
</tr>
<tr>
<td>1986</td>
<td>0.061</td>
<td>578</td>
<td>0.038</td>
<td>1345</td>
<td>0.044</td>
</tr>
<tr>
<td>1987</td>
<td>0.058</td>
<td>578</td>
<td>0.038</td>
<td>1345</td>
<td>0.043</td>
</tr>
<tr>
<td>1988</td>
<td>0.064</td>
<td>578</td>
<td>0.039</td>
<td>1345</td>
<td>0.041</td>
</tr>
<tr>
<td>1989</td>
<td>0.061</td>
<td>578</td>
<td>0.037</td>
<td>1345</td>
<td>0.038</td>
</tr>
</tbody>
</table>

### Table A.2.5. Average investments in new machinery across manufacturing industries

<table>
<thead>
<tr>
<th>Year</th>
<th>Ind. 1</th>
<th>Ind. 2</th>
<th>Ind. 3</th>
<th>All</th>
<th>No. of Firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1982</td>
<td>0.022</td>
<td>578</td>
<td>0.012</td>
<td>1345</td>
<td>0.014</td>
</tr>
<tr>
<td>1983</td>
<td>0.022</td>
<td>578</td>
<td>0.014</td>
<td>1345</td>
<td>0.017</td>
</tr>
<tr>
<td>1984</td>
<td>0.021</td>
<td>578</td>
<td>0.012</td>
<td>1345</td>
<td>0.016</td>
</tr>
<tr>
<td>1985</td>
<td>0.200</td>
<td>578</td>
<td>0.013</td>
<td>1345</td>
<td>0.015</td>
</tr>
<tr>
<td>1986</td>
<td>0.022</td>
<td>578</td>
<td>0.013</td>
<td>1345</td>
<td>0.016</td>
</tr>
<tr>
<td>1987</td>
<td>0.023</td>
<td>578</td>
<td>0.015</td>
<td>1345</td>
<td>0.019</td>
</tr>
<tr>
<td>1988</td>
<td>0.026</td>
<td>578</td>
<td>0.016</td>
<td>1345</td>
<td>0.020</td>
</tr>
<tr>
<td>1989</td>
<td>0.026</td>
<td>578</td>
<td>0.017</td>
<td>1345</td>
<td>0.020</td>
</tr>
</tbody>
</table>
Appendix 2.2. Summary statistics at industry level

Industry 1 = ferrous metals; non ferrous metals; chemicals and artificial fibres.  
Industry 2 = metal manufacturing; machinery; electrical equipment; transport equipment; instrument manufacturing.  
Industry 3 = food, drink and tobacco; textiles; leather goods; timber and furniture; paper and printing; rubber manufacturing; others.

See Appendix 2.1. for variables definition.

**Table A.2.1. Average employment across manufacturing industries**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>all</th>
<th>no. of firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1982</td>
<td>89.39</td>
<td>577</td>
<td>77.36</td>
<td>1346</td>
<td>79.45</td>
</tr>
<tr>
<td>1983</td>
<td>87.83</td>
<td>580</td>
<td>75.97</td>
<td>1348</td>
<td>77.56</td>
</tr>
<tr>
<td>1984</td>
<td>85.65</td>
<td>580</td>
<td>75.15</td>
<td>1348</td>
<td>76.24</td>
</tr>
<tr>
<td>1985</td>
<td>85.25</td>
<td>581</td>
<td>74.81</td>
<td>1348</td>
<td>76.59</td>
</tr>
<tr>
<td>1986</td>
<td>85.12</td>
<td>582</td>
<td>74.71</td>
<td>1348</td>
<td>76.95</td>
</tr>
<tr>
<td>1987</td>
<td>85.32</td>
<td>582</td>
<td>75.76</td>
<td>1348</td>
<td>78.12</td>
</tr>
<tr>
<td>1988</td>
<td>87.21</td>
<td>582</td>
<td>77.56</td>
<td>1348</td>
<td>79.73</td>
</tr>
<tr>
<td>1989</td>
<td>89.11</td>
<td>581</td>
<td>79.17</td>
<td>1349</td>
<td>80.62</td>
</tr>
</tbody>
</table>

**Table A.2.2. Average real wage across manufacturing industries**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>all</th>
<th>no. of firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1982</td>
<td>11.77</td>
<td>582</td>
<td>11.33</td>
<td>1346</td>
<td>10.35</td>
</tr>
<tr>
<td>1983</td>
<td>11.79</td>
<td>590</td>
<td>11.37</td>
<td>1346</td>
<td>10.47</td>
</tr>
<tr>
<td>1984</td>
<td>11.68</td>
<td>591</td>
<td>11.56</td>
<td>1346</td>
<td>10.54</td>
</tr>
<tr>
<td>1985</td>
<td>12.22</td>
<td>593</td>
<td>12.04</td>
<td>1346</td>
<td>10.80</td>
</tr>
<tr>
<td>1986</td>
<td>12.39</td>
<td>594</td>
<td>12.23</td>
<td>1346</td>
<td>11.02</td>
</tr>
<tr>
<td>1987</td>
<td>13.35</td>
<td>593</td>
<td>13.23</td>
<td>1346</td>
<td>11.77</td>
</tr>
<tr>
<td>1988</td>
<td>14.26</td>
<td>594</td>
<td>14.01</td>
<td>1346</td>
<td>12.34</td>
</tr>
<tr>
<td>1989</td>
<td>15.18</td>
<td>595</td>
<td>15.07</td>
<td>1345</td>
<td>13.06</td>
</tr>
</tbody>
</table>
Appendix 2.3. Rate of change in employment, output and wages: summary statistics

Table A.2.6.

<table>
<thead>
<tr>
<th>Rate of change in employment</th>
<th>82-83</th>
<th>83-84</th>
<th>84-85</th>
<th>85-86</th>
<th>86-87</th>
<th>87-88</th>
<th>88-89</th>
</tr>
</thead>
<tbody>
<tr>
<td>no. of firms</td>
<td>3253</td>
<td>3258</td>
<td>3260</td>
<td>3263</td>
<td>3263</td>
<td>3263</td>
<td>3263</td>
</tr>
<tr>
<td>mean</td>
<td>1.55</td>
<td>2.02</td>
<td>2.26</td>
<td>1.26</td>
<td>2.58</td>
<td>3.27</td>
<td>3.54</td>
</tr>
<tr>
<td>std.dev.</td>
<td>29.5</td>
<td>26.8</td>
<td>20.4</td>
<td>16.7</td>
<td>19.6</td>
<td>20.2</td>
<td></td>
</tr>
<tr>
<td>skewness</td>
<td>10.6</td>
<td>7.2</td>
<td>5.7</td>
<td>2.9</td>
<td>3.0</td>
<td>2.8</td>
<td>4.0</td>
</tr>
<tr>
<td>kurtosis</td>
<td>180.0</td>
<td>90.2</td>
<td>70.2</td>
<td>28.3</td>
<td>31.3</td>
<td>20.7</td>
<td>39.5</td>
</tr>
<tr>
<td>sum</td>
<td>5064</td>
<td>6581</td>
<td>7382</td>
<td>4115</td>
<td>8446</td>
<td>10672</td>
<td>11561</td>
</tr>
<tr>
<td>lowest value</td>
<td>-78.6</td>
<td>-86.7</td>
<td>-67.5</td>
<td>-88.4</td>
<td>-72.5</td>
<td>-91.7</td>
<td>-93.9</td>
</tr>
<tr>
<td>highest value</td>
<td>500.0</td>
<td>520.0</td>
<td>370.2</td>
<td>210.0</td>
<td>240.0</td>
<td>200.0</td>
<td>300.0</td>
</tr>
<tr>
<td>range</td>
<td>578.5</td>
<td>606.7</td>
<td>437.9</td>
<td>298.4</td>
<td>312.5</td>
<td>291.7</td>
<td>393.9</td>
</tr>
<tr>
<td>mode</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>median</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table A.2.7.

<table>
<thead>
<tr>
<th>Rate of change in real output</th>
<th>82-83</th>
<th>83-84</th>
<th>84-85</th>
<th>85-86</th>
<th>86-87</th>
<th>87-88</th>
<th>88-89</th>
</tr>
</thead>
<tbody>
<tr>
<td>no. of firms</td>
<td>3249</td>
<td>3254</td>
<td>3256</td>
<td>3259</td>
<td>3259</td>
<td>3259</td>
<td>3259</td>
</tr>
<tr>
<td>mean</td>
<td>3.0</td>
<td>11.3</td>
<td>9.4</td>
<td>3.9</td>
<td>6.1</td>
<td>10.8</td>
<td>7.2</td>
</tr>
<tr>
<td>std.dev.</td>
<td>34.0</td>
<td>27.6</td>
<td>25.6</td>
<td>23.8</td>
<td>19.4</td>
<td>21.3</td>
<td>20.2</td>
</tr>
<tr>
<td>skewness</td>
<td>5.0</td>
<td>3.9</td>
<td>3.7</td>
<td>3.7</td>
<td>1.9</td>
<td>1.4</td>
<td>2.1</td>
</tr>
<tr>
<td>kurtosis</td>
<td>42.7</td>
<td>35.4</td>
<td>33.2</td>
<td>36.1</td>
<td>18.5</td>
<td>8.7</td>
<td>18.0</td>
</tr>
<tr>
<td>sum</td>
<td>9797</td>
<td>38846</td>
<td>30606</td>
<td>12665</td>
<td>19810</td>
<td>32613</td>
<td>23343</td>
</tr>
<tr>
<td>lowest value</td>
<td>-77.9</td>
<td>-79.9</td>
<td>-70.5</td>
<td>-83.1</td>
<td>-80.3</td>
<td>-85.3</td>
<td>-76.1</td>
</tr>
<tr>
<td>highest value</td>
<td>403.7</td>
<td>362.8</td>
<td>347.1</td>
<td>360.2</td>
<td>284.9</td>
<td>203.5</td>
<td>285.4</td>
</tr>
<tr>
<td>range</td>
<td>481.7</td>
<td>442.7</td>
<td>417.6</td>
<td>443.4</td>
<td>365.3</td>
<td>288.7</td>
<td>361.5</td>
</tr>
<tr>
<td>median</td>
<td>-1.4</td>
<td>8.4</td>
<td>6.3</td>
<td>1.6</td>
<td>4.7</td>
<td>7.9</td>
<td>5.5</td>
</tr>
</tbody>
</table>
Table A.2.8.

<table>
<thead>
<tr>
<th>Rate of change in real wages</th>
</tr>
</thead>
<tbody>
<tr>
<td>82-83</td>
</tr>
<tr>
<td>no. of firms</td>
</tr>
<tr>
<td>mean</td>
</tr>
<tr>
<td>std. dev.</td>
</tr>
<tr>
<td>skewness</td>
</tr>
<tr>
<td>kurtosis</td>
</tr>
<tr>
<td>sum</td>
</tr>
<tr>
<td>lowest value</td>
</tr>
<tr>
<td>highest value</td>
</tr>
<tr>
<td>range</td>
</tr>
<tr>
<td>median</td>
</tr>
</tbody>
</table>
Appendix 2.4. Percentage rate of change in employment by size (number of employees)

Firms with 1-19 employees:

Figure A.2.1. 1982-83

Figure A.2.2. 1983-84

Figure A.2.3. 1984-85

Figure A.2.4. 1985-86

Figure A.2.5. 1986-1987

Figure A.2.6. 1987-88
Firms with 20-49 employees:

Figure A.2.7. 1988-89

Figure A.2.8. 1982-83

Figure A.2.9. 1984-1985

Figure A.2.10. 1984-85

Figure A.2.11. 1986-86
Firms with 50-99 employees:
Firms with 100-199 employees:

Figure A.2.22. 1982-83

Figure A.2.23. 1983-84

Figure A.2.24. 1984-85

Figure A.2.25. 1985-86

Figure A.2.26. 1986-87

Figure A.2.27. 1987-88
Firms with 200-500 employees:

Figure A.2.28. 1988-89

Figure A.2.29. 1982-83

Figure A.2.30. 1983-84

Figure A.2.31. 1984-85

Figure A.2.32. 1985-86
Appendix 2.5. Transition probabilities by industry

Industry 1 = ferrous metals; non ferrous metals; chemicals and artificial fibres.

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(83-84)</strong></td>
<td><strong>(84-85)</strong></td>
</tr>
<tr>
<td>+ 0 -</td>
<td>+ 0 -</td>
</tr>
<tr>
<td>+ 0.38 0.19 0.42</td>
<td>+ 0.50 0.20 0.29</td>
</tr>
<tr>
<td><strong>(82-83)</strong></td>
<td><strong>(83-84)</strong></td>
</tr>
<tr>
<td>0 0.32 0.29 0.38</td>
<td>0 0.39 0.30 0.30</td>
</tr>
<tr>
<td>- 0.35 0.13 0.51</td>
<td>- 0.29 0.21 0.49</td>
</tr>
<tr>
<td><strong>(85-86)</strong></td>
<td><strong>(86-87)</strong></td>
</tr>
<tr>
<td>+ 0 -</td>
<td>+ 0 -</td>
</tr>
<tr>
<td>+ 0.48 0.19 0.32</td>
<td>+ 0.53 0.16 0.30</td>
</tr>
<tr>
<td><strong>(84-85)</strong></td>
<td><strong>(85-86)</strong></td>
</tr>
<tr>
<td>0 0.36 0.27 0.36</td>
<td>0 0.39 0.28 0.32</td>
</tr>
<tr>
<td>- 0.36 0.22 0.41</td>
<td>- 0.36 0.22 0.41</td>
</tr>
<tr>
<td><strong>(87-88)</strong></td>
<td><strong>(88-89)</strong></td>
</tr>
<tr>
<td>+ 0 -</td>
<td>+ 0 -</td>
</tr>
<tr>
<td>+ 0.53 0.18 0.28</td>
<td>+ 0.48 0.18 0.33</td>
</tr>
<tr>
<td><strong>(86-87)</strong></td>
<td><strong>(87-88)</strong></td>
</tr>
<tr>
<td>0 0.43 0.28 0.28</td>
<td>0 0.49 0.29 0.21</td>
</tr>
<tr>
<td>- 0.42 0.19 0.38</td>
<td>- 0.52 0.14 0.33</td>
</tr>
</tbody>
</table>
Table A.2.10. Standard errors

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.036 .029 .036</td>
<td>.034 .028 .031</td>
<td>.033 .026 .031</td>
<td>.032 .023 .029</td>
<td>.031 .024 .028</td>
<td>.034 .027 .034</td>
</tr>
<tr>
<td></td>
<td>.035 .034 .037</td>
<td>.045 .042 .042</td>
<td>.041 .038 .041</td>
<td>.042 .039 .040</td>
<td>.044 .040 .040</td>
<td>.036 .025 .034</td>
</tr>
<tr>
<td></td>
<td>.031 .022 .032</td>
<td>.028 .025 .031</td>
<td>.032 .027 .032</td>
<td>.032 .028 .033</td>
<td>.034 .027 .034</td>
<td>.036 .025 .034</td>
</tr>
</tbody>
</table>

120
Table A.2.11. Frequencies

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>Total</th>
<th>(2)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>70</td>
<td>35</td>
<td>78</td>
<td>183</td>
<td>105</td>
</tr>
<tr>
<td>57</td>
<td>52</td>
<td>66</td>
<td>175</td>
<td>47</td>
</tr>
<tr>
<td>83</td>
<td>32</td>
<td>122</td>
<td>237</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td></td>
<td></td>
<td>(4)</td>
<td></td>
</tr>
<tr>
<td>111</td>
<td>45</td>
<td>75</td>
<td>231</td>
<td>129</td>
</tr>
<tr>
<td>50</td>
<td>37</td>
<td>49</td>
<td>136</td>
<td>53</td>
</tr>
<tr>
<td>83</td>
<td>51</td>
<td>94</td>
<td>228</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5)</td>
<td></td>
<td></td>
<td>(6)</td>
<td></td>
</tr>
<tr>
<td>140</td>
<td>48</td>
<td>74</td>
<td>262</td>
<td>137</td>
</tr>
<tr>
<td>54</td>
<td>36</td>
<td>35</td>
<td>125</td>
<td>61</td>
</tr>
<tr>
<td>89</td>
<td>39</td>
<td>80</td>
<td>208</td>
<td>99</td>
</tr>
</tbody>
</table>
Industry 2 = metal manufacturing; machinery; electrical equipment; transport equipment; instrument manufacturing.

Table A.2.12. Transition probabilities

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(83-84)</td>
<td>+ 0 -</td>
<td>+ 0 -</td>
</tr>
<tr>
<td></td>
<td>0.44 0.16 0.39</td>
<td>0.49 0.19 0.31</td>
</tr>
<tr>
<td>(82-83)</td>
<td>0 0.31 0.26 0.42</td>
<td>(83-84) 0 0.36 0.30 0.33</td>
</tr>
<tr>
<td></td>
<td>- 0.32 0.14 0.53</td>
<td>- 0.36 0.15 0.48</td>
</tr>
<tr>
<td>(3)</td>
<td>(4)</td>
<td></td>
</tr>
<tr>
<td>(85-86)</td>
<td>+ 0 -</td>
<td>+ 0 -</td>
</tr>
<tr>
<td></td>
<td>0.43 0.22 0.34</td>
<td>0.54 0.17 0.28</td>
</tr>
<tr>
<td>(84-85)</td>
<td>0 0.37 0.34 0.28</td>
<td>(85-86) 0 0.44 0.28 0.27</td>
</tr>
<tr>
<td></td>
<td>- 0.33 0.18 0.48</td>
<td>- 0.35 0.16 0.48</td>
</tr>
<tr>
<td>(5)</td>
<td>(6)</td>
<td></td>
</tr>
<tr>
<td>(87-88)</td>
<td>+ 0 -</td>
<td>+ 0 -</td>
</tr>
<tr>
<td></td>
<td>0.53 0.16 0.30</td>
<td>0.53 0.14 0.32</td>
</tr>
<tr>
<td>(86-87)</td>
<td>0 0.47 0.24 0.28</td>
<td>(87-88) 0 0.16 0.23 0.31</td>
</tr>
<tr>
<td></td>
<td>- 0.40 0.14 0.45</td>
<td>- 0.50 0.15 0.34</td>
</tr>
</tbody>
</table>
### Table A.2.13. Standard Errors

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>.025</td>
<td>.018</td>
<td>.024</td>
</tr>
<tr>
<td>.025</td>
<td>.023</td>
<td>.026</td>
</tr>
<tr>
<td>.019</td>
<td>.014</td>
<td>.020</td>
</tr>
<tr>
<td>(3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.021</td>
<td>.017</td>
<td>.020</td>
</tr>
<tr>
<td>.030</td>
<td>.029</td>
<td>.018</td>
</tr>
<tr>
<td>.020</td>
<td>.017</td>
<td>.021</td>
</tr>
<tr>
<td>(5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.028</td>
<td>.015</td>
<td>.018</td>
</tr>
<tr>
<td>.031</td>
<td>.026</td>
<td>.028</td>
</tr>
<tr>
<td>.022</td>
<td>.016</td>
<td>.022</td>
</tr>
</tbody>
</table>

123
Table A.2.14. Frequencies

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>Total</th>
<th>(2)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>174</td>
<td>65</td>
<td>157</td>
<td>396</td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>94</td>
<td>148</td>
<td>352</td>
<td></td>
</tr>
<tr>
<td>196</td>
<td>86</td>
<td>329</td>
<td>611</td>
<td></td>
</tr>
<tr>
<td>244</td>
<td>125</td>
<td>193</td>
<td>562</td>
<td></td>
</tr>
<tr>
<td>97</td>
<td>91</td>
<td>74</td>
<td>262</td>
<td></td>
</tr>
<tr>
<td>179</td>
<td>98</td>
<td>258</td>
<td>535</td>
<td></td>
</tr>
<tr>
<td>325</td>
<td>97</td>
<td>187</td>
<td>609</td>
<td></td>
</tr>
<tr>
<td>125</td>
<td>65</td>
<td>74</td>
<td>264</td>
<td></td>
</tr>
<tr>
<td>197</td>
<td>71</td>
<td>218</td>
<td>486</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>239</td>
<td>92</td>
<td>149</td>
<td>480</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>74</td>
<td>81</td>
<td>245</td>
</tr>
<tr>
<td></td>
<td>233</td>
<td>96</td>
<td>305</td>
<td>634</td>
</tr>
<tr>
<td></td>
<td>282</td>
<td>92</td>
<td>146</td>
<td>520</td>
</tr>
<tr>
<td></td>
<td>140</td>
<td>88</td>
<td>86</td>
<td>314</td>
</tr>
<tr>
<td></td>
<td>187</td>
<td>84</td>
<td>254</td>
<td>525</td>
</tr>
<tr>
<td></td>
<td>343</td>
<td>92</td>
<td>212</td>
<td>647</td>
</tr>
<tr>
<td></td>
<td>106</td>
<td>54</td>
<td>73</td>
<td>233</td>
</tr>
<tr>
<td></td>
<td>241</td>
<td>73</td>
<td>165</td>
<td>479</td>
</tr>
</tbody>
</table>

124
Industry 3 = food, drink and tobacco; textiles; leather goods; timber and furniture; paper and printing; rubber manufacturing; others.

Table A.2.15 Transition probabilities

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(83-84)</td>
<td>+ 0 -</td>
<td>+ 0 -</td>
</tr>
<tr>
<td></td>
<td>+ 0.35 0.14 0.50</td>
<td>+ 0.44 0.20 0.35</td>
</tr>
<tr>
<td>(82-83)</td>
<td>0 0.34 0.31 0.34</td>
<td>(83-84) 0 0.37 0.35 0.27</td>
</tr>
<tr>
<td></td>
<td>- 0.41 0.21 0.38</td>
<td>- 0.34 0.22 0.44</td>
</tr>
<tr>
<td>(3)</td>
<td></td>
<td>(4)</td>
</tr>
<tr>
<td>(85-86)</td>
<td>+ 0 -</td>
<td>+ 0 -</td>
</tr>
<tr>
<td></td>
<td>+ 0.44 0.23 0.32</td>
<td>+ 0.52 0.16 0.31</td>
</tr>
<tr>
<td>(84-85)</td>
<td>0 0.34 0.28 0.37</td>
<td>(85-86) 0 0.41 0.29 0.29</td>
</tr>
<tr>
<td></td>
<td>- 0.37 0.16 0.46</td>
<td>- 0.38 0.20 0.41</td>
</tr>
<tr>
<td>(5)</td>
<td></td>
<td>(6)</td>
</tr>
<tr>
<td>(87-88)</td>
<td>+ 0 -</td>
<td>+ 0 -</td>
</tr>
<tr>
<td></td>
<td>+ 0.48 0.21 0.39</td>
<td>+ 0.52 0.18 0.29</td>
</tr>
<tr>
<td>(86-87)</td>
<td>0 0.38 0.30 0.31</td>
<td>(87-88) 0 0.42 0.27 0.30</td>
</tr>
<tr>
<td></td>
<td>- 0.40 0.17 0.42</td>
<td>- 0.38 0.16 0.45</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td></td>
<td>.020</td>
<td>.015</td>
</tr>
<tr>
<td></td>
<td>.024</td>
<td>.023</td>
</tr>
<tr>
<td></td>
<td>.024</td>
<td>.020</td>
</tr>
<tr>
<td></td>
<td>.022</td>
<td>.018</td>
</tr>
<tr>
<td></td>
<td>.029</td>
<td>.029</td>
</tr>
<tr>
<td></td>
<td>.020</td>
<td>.017</td>
</tr>
<tr>
<td></td>
<td>.026</td>
<td>.025</td>
</tr>
<tr>
<td></td>
<td>.021</td>
<td>.016</td>
</tr>
<tr>
<td></td>
<td>.021</td>
<td>.017</td>
</tr>
<tr>
<td></td>
<td>.020</td>
<td>.017</td>
</tr>
<tr>
<td></td>
<td>.029</td>
<td>.027</td>
</tr>
<tr>
<td></td>
<td>.023</td>
<td>.017</td>
</tr>
<tr>
<td></td>
<td>.022</td>
<td>.017</td>
</tr>
</tbody>
</table>
Table A.2.17. Frequencies

<table>
<thead>
<tr>
<th>(1)</th>
<th>Total</th>
<th>(2)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>194</td>
<td>76</td>
<td>278</td>
<td>548</td>
</tr>
<tr>
<td>133</td>
<td>120</td>
<td>133</td>
<td>386</td>
</tr>
<tr>
<td>168</td>
<td>82</td>
<td>155</td>
<td>405</td>
</tr>
<tr>
<td>229</td>
<td>118</td>
<td>169</td>
<td>516</td>
</tr>
<tr>
<td>111</td>
<td>93</td>
<td>119</td>
<td>323</td>
</tr>
<tr>
<td>185</td>
<td>83</td>
<td>232</td>
<td>500</td>
</tr>
<tr>
<td>288</td>
<td>123</td>
<td>180</td>
<td>591</td>
</tr>
<tr>
<td>106</td>
<td>86</td>
<td>87</td>
<td>279</td>
</tr>
<tr>
<td>189</td>
<td>83</td>
<td>197</td>
<td>469</td>
</tr>
</tbody>
</table>
The following table reports the $\chi^2$ tests for the equality of the conditional and unconditional transition probabilities at industry level. (4) indicates the degrees of freedom (see Section 2.5. for the construction of the test). The numbers in brackets refer to the matrices as they appear in Tables A.2.9.-A.2.17. above. All the values imply a rejection of the null hypothesis of absence of state dependence.
Table A.2.18. $\chi^2$ test for state dependence

<table>
<thead>
<tr>
<th>Industry 1.</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>19.76</td>
<td>53.20</td>
<td>32.49</td>
<td>52.62</td>
<td>68.06</td>
<td>96.88</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Industry 2.</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50.39</td>
<td>124.7</td>
<td>141.4</td>
<td>118.7</td>
<td>183.2</td>
<td>218.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Industry 3.</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>153.8</td>
<td>82.57</td>
<td>77.03</td>
<td>95.00</td>
<td>157.0</td>
<td>161.1</td>
</tr>
</tbody>
</table>
Chapter 3. QUADRATIC ADJUSTMENT COSTS: EULER EQUATION REPRESENTATION

Introduction

3.1. Theoretical model

3.2. Stochastic specification and choice of estimation method

3.3. The dataset

3.4. The selectivity problem

3.5. Estimation of the selection term (step 1)

3.6. Estimation of the Euler equation (step 2)

3.7. Static model

Appendix 3.1. Euler equation

Appendix 3.2. Selection bias: distributional assumptions on the error terms.

Appendix 3.3. Inverse Mills’ ratio: summary statistics

Appendix 3.4. Plots of the probability of adjusting

Appendix 3.5. Program for GMM estimation

Appendix 3.6. Estimation of the Euler equation with time dummies
Introduction

In this chapter we analyze the case of quadratic (or, more generally, strictly convex) adjustment costs. The hypothesis of strict convexity allows us to derive a dynamic labour demand model directly from the first order conditions or Euler equations. The model predicts that the firm's optimal response to exogenous shocks is to vary employment slowly and smoothly over time. The Euler equation representation has many advantages in terms of mathematical tractability. Furthermore, if we assume rational expectations, Euler estimation methods may exploit the orthogonality conditions between the firm's forecast error and its information set. This allows us to use "generalized method of moments" (GMM) techniques [Hansen, 1982 and Hansen and Singleton, 1982] where orthogonality comes from adopting as valid instruments a set of variables from the firm's information set. Nevertheless, as illustrated in the previous chapter, in the case of our dataset the estimation of the Euler equation is complicated by the fact that, on average over the sample period, around 20% of companies do not change employment from one period to the other, even in the presence of changes in wages and output. The existence of this non-negligible proportion of zero variations in employment introduces a selection problem which we need to take into account when estimating the Euler model. We interpret the existence of the peak of frequencies at zero net changes of employment in two ways. The first regards zero changes as due to measurement errors, like recording errors, misreporting, etc., which we take into account by the selecting from the sample the firms characterized
by zero variations in employment. The second implies the existence of fixed costs which induce companies not to adjust in certain periods. This leads to a selection model where we observe adjustment in the case of firms facing quadratic adjustment costs and where we have censoring due to the existence of non strictly convex costs.

In order to estimate the model we, therefore, adopt a two stage procedure. First, we study the selection process by estimating the discrete choice of whether or not to adjust, using a probit maximum likelihood estimation procedure and obtain the vector of inverse Mills' ratios which will be included, at the second stage, in the Euler equation as the selection correction term. An important feature of our estimation procedure is that the selection process is specified in terms of the instruments which define the orthogonality conditions in the GMM estimation of the adjusted Euler equation, using the non-censored sample. We consider four different specifications of the Euler equation in order to take into account fixed effects and to experiment with different timing of the instruments. Each model includes the appropriate selectivity term calculated with regard to the timing of the instruments.

Our approach provides a general procedure for taking into account selectivity bias in panel data estimation. The use of the instrumental variables employed in the GMM estimation procedure makes the probits for selection correction consistent with the specification of the Euler equation. Both the sets of instrumental variables and regressors in the GMM procedure will then be augmented by the selection correction terms calculated from the probits, once the appropriated timing of the instruments has been determined on the basis of the
stochastic specification of the model.

3.1. Theoretical model

Our discussion of the effects of different structures of adjustment costs on employment decisions starts with the conventional assumption of quadratic adjustment costs. Because of its mathematical tractability, this hypothesis has been widely employed in the economic literature on dynamics. The computational advantage of the quadratic costs assumption stems from the fact that the marginal conditions derived from the objective function are linear in the state variables. Thus, if the optimum is described by first order conditions, the optimal policy function is also linear in the state variables. ³¹

In what follows we derive a model of dynamic labour demand where the costs of changing employment are quadratic. Under this hypothesis, the optimal employment path is described by the Euler equation. The model predicts that the firms will react to exogenous shocks by varying their current employment gradually and smoothly over time. Intuitively this optimizing response comes from the fact that it would be very costly for the firm to change employment to the desired level instantaneously and fully, because this causes costs to rise with the square of the change.

Consider a firm which, at the start of every period \( t \), inherits a stock of labour, \( L_{t-1} \). Hiring and firing decisions are made to maximize the expected present

³¹ The optimal policy function tells us how to proceed from any specific initial point in order to attain the firm’s objective as indicated by the value function.
value of future stream of profits in the presence of adjustment costs:

\[ E_t\left[ \sum_{\tau} \beta^{\tau} \Pi_t(L_{t-\tau}) \right] \]  \hspace{1cm} (16)

where \( \beta \) is the firm's discount rate, \( \Pi(.) \) denotes current profits, including strictly convex adjustment costs, and \( E_t \) is the expectations operator conditional on information available at the start of period \( t \), indicated by \( I_t \). For simplicity, profits are considered to be a function only of the stock of labour, \( L_{t-1} \). Uncertainty comes from future product prices, wages and possibly the interest rate through the discount factor. We may express the optimization problem in terms of the following value function:

\[ V_t(L_{t-1}) = \max_{L_t} \left\{ \Pi_t(L_{t-1}) + \beta_t E_t[V_{t+1}(L_t)] I_t \right\} \]  \hspace{1cm} (17)

Maximization results in the following first order conditions (which hold in the presence of any strictly convex adjustment costs structure)

\[ \frac{\partial \Pi_t}{\partial L_t} + \beta_t E_t \left( \frac{\partial V_{t+1}}{\partial L_t} I_t \right) = 0 \]  \hspace{1cm} (18)

\[ \frac{\partial V_t}{\partial L_{t-1}} = \frac{\partial \Pi_t}{\partial L_{t-1}} \]

Taking the second equation in (18) one period ahead and substituting into the first equation we obtain a standard Euler equation representation of the first order conditions \(^{32}\):

\(^{32}\) This formulation is very similar to that in Machin, Manning and Meghir (1991)
Formulation (19) introduces an expectational error due to the fact that the company chooses employment conditionally on information at time $t$.

\[ \frac{\partial \Pi_t}{\partial L_t} + \beta_t E_t \left( \frac{\partial \Pi_{t+1}}{\partial L_t} | I_t \right) = 0 \]  \hspace{1cm} (19)

Under rational expectations, the expectations of the forecast error is zero, $E(v_{t+1} | I_t) = 0$. Euler equation estimation methods essentially exploit this orthogonality condition between the firm's forecast error and elements within its information set. Hence the problem is driven by two sources of stochastic disturbance: forecast errors and technical shocks introduced by the output price, wages and the interest rate. The latter set of disturbances will affect the timing of the instruments in the estimation of the Euler equations as we shall see in Section 3.2.

The stochastic Euler equation (19) is simply a first order necessary condition which characterizes the optimal employment decision. It says that the change in expected discounted profits from a small change in $L_t$ will be zero under the optimal decision rule. The remarkable feature of the Euler equation is that the impact on discounted profits may be evaluated in terms of the value of marginal conditions in period $t$ and the expected discounted value of marginal profits in period $t+1$: it is not necessary to consider the impact on marginal profits on future periods $t+2$, $t+3$, ... This idea forms the basis for the estimation and testing procedure pioneered by Hansen and Singleton (1982) and Hansen (1982): the "generalized method of moments" (GMM) procedure which we shall see later.
Assuming an infinite planning horizon, transversality conditions, which give the terminal condition for the free terminal state problem, are given by

$$\lim_{t \to \infty} E_t \left[ \beta^t \frac{\partial \Pi}{\partial L_t} | I_t \right] = 0$$  \hspace{1cm} (21)

An estimable model may be derived by specifying the structure of the within period profit function $\Pi_t$ including adjustment costs. For example, one might assume a Cobb-Douglas production function in the labour factor and obtain

$$\Pi_t = AL_t^\gamma - \frac{\gamma}{2} (\Delta L_t)^2 - W_t L_t$$  \hspace{1cm} (22)

where $W_t$ is the wage. Since we are interested in determining the path of labour demand focusing on the role of labour adjustment costs and expectations, we maintain, for simplicity, the assumption that profits are functions only of the labour input. In other words, we assume that labour adjustment costs are not affected by other input, in particular capital stock and investment. The term $\frac{\gamma}{2} (\Delta L_t)^2 = C(L_t, L_t)$ represents quadratic adjustment costs; $\gamma$ is a constant parameter, strictly greater than zero.

Following Hamermesh (1989), we consider adjustment costs associate with net rather than gross variations in employment. The distinction between net and gross costs of adjustment is important and it affects the entire subsequent interpretation of employers' behaviour. Some costs originate through gross changes in employment, such as the addition of new workers or departure of current employees. These costs may be relevant even if the level of employment within the firm does not ultimately change; i.e. the net employment change is zero. Clearly, for the same net change in employment, costs will be larger if more workers need
to be hired to replace a larger flow of quits. Conversely, adjustment costs occasioned by net changes in employment do not relate to turnover but to the number of jobs. Net changes may generate costs independent of those produced by the turnover. A good example is given by, the cost of disruption generated by the expansion of the total work force or the costs associated with maintaining intact an entire work shift.

We may write the first order condition (19) explicitly as (see Appendix 3.1. for the derivation):

\[(\beta \gamma)E_t(L_{t+1}) - \gamma(1+\beta)L_t + \gamma L_{t-1} + \alpha\left(\frac{Q}{L}\right)_t - W_t = 0\]  

(23)

where \(Q\) denotes sales. The transversality conditions may now be explicitly written as

\[\lim_{t \to -\infty} \beta^t E_t\left[\alpha\left(\frac{Q}{L}\right)_t - \gamma \Delta L_t - W_t\right] = 0\]  

(24)

Given the first order and transversality conditions, we obtain a unique forward stable optimal solution if one of the two real characteristics roots of the quadratic equation

\[\left(\frac{\beta}{1+\beta}\right)y^2 - y + \left(\frac{1}{1+\beta}\right) = 0\]  

(25)

falls outside the unit circle, and the other falls inside the unit circle. Then, solving the first order conditions (23) for \(L_t\), we obtain an explicit representation for the extremal as
\[ L_t = \psi_1 L_{t+1} + \psi_2 L_{t-1} + \psi_3 \left( \frac{Q}{L} \right)_t + \psi_4 W_t + \nu_{t+1} \]  \hspace{1cm} (26)

The structural parameters in (26) are as follows

\[ \psi_1 = \frac{\beta}{1+\beta} \]
\[ \psi_2 = \frac{1}{1+\beta} \]
\[ \psi_3 = -\frac{\alpha}{\gamma(1+\beta)} \]
\[ \psi_4 = -\frac{1}{\gamma(1+\beta)} \]  \hspace{1cm} (27)

where \( \Psi_0 \) is the intercept term. \(^{33}\)

Current employment decisions depend on next period decisions, as represented by the forward looking term, \( L_{t+1} \). The presence of quadratic adjustment costs slows the response to shocks by companies. If the entire change in employment were made immediately, the marginal cost generated by adding (or dropping) the last employee would be potentially very large, because of the quadratic term. By smoothing adjustment over time, the firm lowers the total costs of adjustment by more than enough to offset the lost profits from failing to set employment to the desired level.

The theory predicts that \( 0 < \psi_1 < 1, 0 < \psi_2 < 1, \psi_3 > 0, \) and \( \psi_4 < 0. \) The theory also predicts \( \psi_1 < \psi_2 \) (since \( 0 < \beta < 1 \)) and \( \psi_3 < -\psi_4 \) (\( 0 < \alpha < 1 \)). However, the inequality relating to \( \psi_1 \) and \( \psi_2 \) is not robust with regard to a simple generalization of the model which leaves the form of the Euler equation (26) unchanged. If we replace

\(^{33}\) Although the theory does not indicate an intercept, we have followed standard practice by including intercept in all estimated equations. The presence of the intercept may be justified by the fact that the productivity and wage variables are not defined in the same unit.
(28) with

$$
\Pi_t = \lambda L_t + \frac{\gamma_1}{2} (\Delta L_{t-1})^2 - \frac{\gamma_2}{2} (\Delta L_t)^2 - \frac{\gamma_3}{2} (\Delta L_{t+1})^2 - w_t L_t
$$

we find that

$$
\psi_1 = \frac{\beta \gamma_1 + \gamma_3}{\beta \gamma_1 + (1 + \beta) \gamma_2 + \gamma_3}
$$

$$
\psi_2 = \frac{\beta \gamma_1 + \gamma_2}{\beta \gamma_1 + (1 + \beta) \gamma_2 + \gamma_3}
$$

$$
\psi_3 = \frac{\alpha}{\beta \gamma_1 + (1 + \beta) \gamma_2 + \gamma_3}
$$

$$
\psi_4 = \frac{1}{\beta \gamma_1 + (1 + \beta) \gamma_2 + \gamma_3}
$$

It follows that

$$
\psi_1 > \psi_2 \quad \text{or} \quad \psi_1 < \psi_2
$$

as

$$
\gamma_3 > \beta \gamma_1 + (1 + \beta) \gamma_2 \quad \text{or} \quad \gamma_3 < \beta \gamma_1 + (1 + \beta) \gamma_2
$$

In this specification, adjustment costs are incurred in relation to both changes in employment in the previous period ($\gamma_1$) and in the successive period ($\gamma_3$), in addition to the current period ($\gamma_2$). The presence of adjustment costs relatively to lagged changes will arise if training costs persist into the following period; those relating to projected changes arise in relation to the need to advertise, check with employment agencies and so forth. The inequality $\gamma_1 < \gamma_2$ is reversed if this latter category of costs is sufficiently large.

Due to its simplicity the Euler equation has represented the dominant approach to analysis of dynamic factor demands. Indeed, in the presence of convex
and continuously differentiable adjustment costs the optimization problem results in an adjustment path which has the desirable features of smoothness, continuity and forward stability.

3.2. Stochastic specification and choice of estimation method

Equation (26) represents the empirical specification we are going to estimate in Section 3.6. As already anticipated, we may directly exploit the Euler equation (23) taken as defining orthogonality conditions and apply the Generalized Method of Moments estimator. GMM estimates the vector of unknown parameters of the model, \( \Psi' \), by finding a vector \( \Psi \) that sets the sample covariances corresponding to the population orthogonality as close to zero as possible.

Write \( X_i'=(L_{it}, L_{it-1}, L_{it-2}, Q_{it}, W_i) \) where \( i=1,\ldots,N \) is a firm index, and let \( z_i \) denote the vector of instruments for firm \( i \) in period \( t \) to be defined below. Define

\[
e_{it} = L_{it} - X'_i \Psi_{it}
\]

the estimated residual corresponding to observation \((i,t)\), \( e'=(e_1',\ldots,e_n') \) where \( p \) is the first year included in the sample (see Figure 36 below) and \( e'=(e_1',e_2',\ldots,e_N') \).

Define the stacked matrices \( X \) and \( Z \) conformably. The basic GMM estimator is then defined by

\[
\min \| e' Z (Z' Z)^{-1} Z' e \|
\]

More generally, however, we consider the estimator defined as

\[
\min \| e' \left[ \sum_{i=1}^{N} Z_i' H Z_i \right]^{-1} Z' e \|
\]
where the Toeplitz matrix $H$ takes into account the possibility of a moving average error. We restrict attention to the case of MA(1) errors and define $H$ as

$$
H = \begin{bmatrix}
1 & h & \ldots & 0 \\
h & 1 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & h
\end{bmatrix}
$$

(34)

where $h$ is estimated at the first stage of a two stage procedure (see Arellano and Bond, 1988). (At the first stage, $h$ is set to -0.5 which corresponds to a differenced white noise error.)

In order to exploit the time-series and cross-sectional information of our dataset, as it will become more clear in the section on the empirical specification of the model, we propose to use panel data which allow us to test for the presence of fixed effects and, if necessary, make allowance for them in the estimation. This implies the estimation of the model both in levels and in first differences. As we may see from Figure 36, four models are involved in the estimation of the Euler equation.

The four models differ not only because regressors are specified either in levels or differences but also because the timing of the instruments is different. The purpose of this is to compare regressions based on different timing of the

---

$^{34}$ We considered the scale of adjustment costs by estimating initially first differences i.e. by looking at changes in employment over the sample period. In fact, first differences represent the natural way of looking at adjustment costs which directly involve the idea of change in employment. This suggests that the specification in levels may not tell us much about the existence of these costs. However by estimating the model both in levels and in first differences we may test for the presence of fixed effects.
### Figure 36 Panel data estimation

<table>
<thead>
<tr>
<th>Estimation in:</th>
<th>LEVELS</th>
<th>DIFFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timing of valid instruments:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$[(t-1),(t-2),(t-3)]$</td>
<td>$[(t-2),(t-3)]$</td>
</tr>
<tr>
<td>data first observation:</td>
<td>$p=2$</td>
<td>$p=3$</td>
</tr>
<tr>
<td>overidentifying restrictions:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sargan test</td>
<td>fixed effect:</td>
<td>Hausman specification test</td>
</tr>
<tr>
<td></td>
<td>$[(t-2),(t-3)]$</td>
<td>$[t-3]$</td>
</tr>
<tr>
<td>data first observation:</td>
<td>$p=3$</td>
<td>$p=4$</td>
</tr>
</tbody>
</table>

instrumental variables. The use of these four specifications has its rationale in the fact that the appropriate timing of the instruments is ultimately, in this problem, an empirical question.

What are the sources of stochastic variations in the model? First of all, we have measurement errors due to the use of actual values of $L_{s+1}$ as approximation of $E_t(L_{s+1})$. Hence, the error term $\nu_{t+1}$ is correlated with the realizations of $L_{s+1}$.\(^{35}\) Another error component is given by the dynamic specification which implies that the fixed effect is correlated with the regressors. To control for fixed effect we

---

\(^{35}\) This error term may also be thought of as reflecting stochastic shocks to the parameters of the model. Such shocks may be serially correlated and theory itself provides no guidance for the structure of such serial correlation.
estimate the model in first differences with the error term $\Delta v_{t+1} = v_{t+1} - v_t$. This suggests that we use as instruments all the available history starting from $t-1$ backwards for the differenced equation, and starting from $t$ for the levels equation. In fact, we choose to omit the variables dated $t$ as instruments in the levels equation and $t-1$ in the differenced equation because of the possibility of technological shocks. An additional source of variation could be a serially uncorrelated effects to the adjustment function, and this would require an extra disturbance term, $u$, in equation (26) which becomes the following

$$L_u = \psi_0 + \psi_1 L_{u+1} + \psi_2 L_{u-1} + \psi_3 \left( \frac{Q}{L} \right)_u + \psi_4 W_u + f_i + u_i + v_{t+1}$$

(35)

where $f_i$ denotes fixed effects we will consider later. The error $u_i$ could thus interact with the non-linear term $Q/L$. This would imply finding instruments orthogonal to $Q/L$. We could, for simplicity, approximate it as additive to $\Psi_3$, i.e. $[\psi_3(Q/L)_i]$ is replaced by $[((\psi_3+u_i)(Q/L)_i]$. In this case also the variables dated $t$ would fail to be weakly exogenous with respect to the parametrization and the differenced error term $[v_{t+1} - v_i + u_i - u_i]$ would be MA(1). This suggests the use of variables up to $t-2$ as instruments in the differenced equation and $t-1$ in the level equation. In what follows we write $\eta_{t+1} = v_{t+1} + u_i$.

If firms take decisions with respect to information at $t-1$ rather than $t$, then $u_i$ is MA(1) and this would suggest use of the entire history up to period $t-2$ in the level equation and up to $t-3$ in the differenced equation as instruments. However, the longer the shortest lag, the poorer the explanation in the reduced form equation for the instruments, and this results in poorer fit, and therefore less precise parameter estimates, in the instrumented Euler equations. In the empirical
part we investigate the validity of instruments using the Sargan test for overidentifying restrictions (Sargan, 1958, 1959 and 1988).

To summarize, if the realization error \( \nu_{i,t} \) were the only source of stochastic variation in the model, we would be able to use variables dated \( t \) and earlier as valid instruments in the levels equation and \( t-1 \), and earlier in the differenced equations. If, in addition, the adjustment function is subject to disturbances \( \nu_a \) so that output and employment are jointly determined, we may only use as valid instruments variables dated \( t-1 \) and earlier in the levels equation, and \( t-2 \) and earlier in the differenced equation. If, furthermore, firms take decisions on the basis of lagged information, we may only use as valid instruments variables dated \( t-2 \) and earlier in the levels equation and \( t-3 \) in the differenced equation.

To allow for an efficient use of the instruments we estimate the four specifications, over a sample of length \( T \), where \( T \) is small, but a large number of firms, \( N \), using the specific form of the Generalized Method of Moments Estimator developed by Arellano and Bond (1988, 1991) and Holtz-Eakin, Newey and Rosen (1988). This estimator utilises all the permissible orthogonality conditions of the form

\[
E(\eta_{it-1} z_{it-k}) = 0, \quad k = j, ..., J \tag{36}
\]

for appropriate timing of the instruments, where the initial lag \( j \) is equal to 1,2,3 (see below), and the final lag \( J \) is set to 3, since remote lags are unlikely to give informative instruments. \( z_{it-k} \) represents a vector of all the variables in the model including the lagged endogenous variables, \( L_{i,t-k} \). Write \( Z_t = (z_{t-1}^t, ..., z_{t-3}^t) \), the complete vector of instruments used for firm \( i \) in period \( t \). The orthogonality condition (36) is then
or, if we estimate in differences

\[ E(\eta_{t+1} Z_{it}) = 0 \]  

(37)

\[ E(\Delta \eta_{t+1} Z_{it}) = 0 \]  

(38)

Intuitively, every year we estimate a different reduced form for each dependent variable using all valid instruments available in that period. For example, if we choose instruments dated t-2 or earlier we are able, in our panel, to start in 1984. In Table XV we describe the choice of instruments.

**Table XV Method of moments**

<table>
<thead>
<tr>
<th>dependent variable</th>
<th>instruments (including lagged dep. var.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( L_{t,1984} )</td>
<td>( Z_{t,1982} )</td>
</tr>
<tr>
<td>( L_{t,1985} )</td>
<td>( Z_{t,1982}; Z_{t,1983} )</td>
</tr>
<tr>
<td>( L_{t,1986} )</td>
<td>( Z_{t,1983}; Z_{t,1984}; Z_{t,1982} )</td>
</tr>
<tr>
<td>( \ldots )</td>
<td>( \ldots )</td>
</tr>
<tr>
<td>( \ldots )</td>
<td>( \ldots )</td>
</tr>
<tr>
<td>( \ldots )</td>
<td>( \ldots )</td>
</tr>
<tr>
<td>( L_{t,1989} )</td>
<td>( Z_{t,1987}; Z_{t,1986}; \ldots; Z_{t,1982} )</td>
</tr>
</tbody>
</table>

Estimation of the Euler equation using the dataset, described in detail in Chapter 2, introduces a selectivity problem due to the presence of a peak of zeros in the distribution of the rate of change in employment. Before turning to
estimation it may, therefore, be useful to summarize the most relevant characteristics of the dataset. We do this in the next section.

3.3. The dataset

The crucial characteristic of our dataset is that, on average over the sample period, about 20 per cent of companies do not alter employment from one year to the next even in the presence of variations in output and wages. The theoretical explanation relies on the hypothesis of fixed costs of adjustment which make optimal for firms to change the labour input less frequently. However, the existence of this censoring point at zero requires further discussion and empirical investigation before we consider an alternative model of dynamic labour demand in the presence of different structures of adjustment costs. For instance, the zeros in the dataset may represent measurement errors. Here we only mention the possible sources of the zero changes in employment while a more detailed discussion can be found in Chapter 2.

1) Zero changes in employment may hide some degree of turnover which occurred during the year. In other words, the zero changes may represent the net sum of people who have left and their replacements, and these may coincidentally cancel leaving a zero net change. This may appear problematic in the context of our model where the focus is on adjustment costs relating to net changes in employment, with attention to the overall level of staffing rather, and not on changes in its detailed composition. This problem was discussed in detail in Section 2.4.2.
2) In a related way, hirings and firings may cancel to give a planned net zero adjustment. Firms may deliberately decide not to change the total labour input, since adjustment may entail high costs. For example, to avoid the costs of reorganizing the production process they may try to maintain the number of workers at each plant unchanged. In other words, firms may organize their personnel such that adjustment costs are minimized by setting the number of new hires equal to the number of quits.

3) Zero adjustment could hide subsidized excess labour rather than a decision not to adjust. The Italian wage supplementation scheme requires that the firms formally account for temporary laid-off workers on the company pay-roll. This allows the possibility that excess labour may be disguised within the firm: a zero change in employment may conceal an otherwise undesired element of labour and result in some smoothing of changes in employment over time. This problem was discussed at greater length in Chapter 1, where we explained the legislation on wage supplementation, individual and collective firings and in Chapter 2, where we matched our data with a dataset on the supplementation programme.

4) There is a problem with regard to what temporally aggregated data can tell us about the structure of adjustment costs. Errors may arise from using data at survey intervals, instead of measuring the state and control variables at the time interval relevant to the firm. Furthermore, there may be a problem in using the appropriate degree of spatial aggregation: the firm rather than the plant may be the relevant decision-making unit (Hamermesh, 1992).

5) Coding and response errors are a general problem in the analysis of survey data. In our context companies could misreport information, or in the case
of the number of employees, they may round the figures. This may, in some cases, affect the validity of the zeros as indicators of non adjustment.

Whatever its origin, the peak at zero introduces a potential selection problem when we estimate the Euler equation since it generates a censoring in the data. In the following section we analyze this problem in detail.

3.4. The selectivity problem

As explained, the existence of a spike in the density function of employment changes at zero adjustment introduces a potential selection problem when estimating the Euler equation (26) in Section 3.1. In the previous section, we considered possible explanations of this censoring point at zero. One explanation relies upon the idea that, in the presence of shocks, firms decide to postpone adjustment when adjustment costs have fixed components, while they would alter the number of employees if those costs were only quadratic. Thus we do not observe any change in employment for a subset of firms in the sample. Another explanation is based on a class of problem relating to the collection of data on employment which we broadly define as measurement errors. In this section, while remaining agnostic with regard to these hypotheses, we consider the selection problem and ask whether the inclusion of the firms characterized by zero adjustment in the sample matters in estimating the Euler specification.

Let $S_i$ be the subset of firms which adjust and $S_o$ the subset of non adjusting firms. Due to the presence of the peak of observations on employment at zero, Ordinary Least Squares performed on $S=(S_o U S_i)$ will give inconsistent estimates.
Alternatively, if we believed that measurement errors were the only source of censoring and that they are completely random, we could simply drop all the observations in $S_o$ from the sample and base the estimates on the selected sample $S_t$. Nevertheless, if this separation is not independent of the disturbance terms $\{\eta_i\}$ we will obtain biased and inconsistent estimates. [see for instance, Maddala (1983) or Amemiya (1985)]. Estimates derived from the selected sample will be biased due to the correlation between the independent variables and the stochastic disturbance induced by the sample selection rule. Hence, before proceeding with the estimation of the Euler equation using our dataset, we must model the process of selection in order to correct for sample selection biases.

At time $t$, we specify the following model

\[
L_u^* = \begin{cases} 
L_u^* = \Psi' X_u + \eta_{it+1} & \text{if } y_u = 1 \\
L_{u-1} & \text{if } y_u = 0 
\end{cases}
\]

(39)

where $L_u^*$ is the latent dependent variable given by the Euler adjustment equation, $X_u$ is the matrix of regressors including the lagged dependent variable, $L_{u-1}$, and the forward term, $L_{u+1}$, as in specification (26), $\Psi$ represents the vector of parameters as in (27). $y_u$ is a binary-valued indicator function which determines the selection. We suppose

\[
Pr(y_u = 1) = f(Z_u)
\]

(40)

where $Z_u$ is the complete set of instruments used in estimating the Euler representation. In Chapter 6 we consider the optimal adjustment choice with fixed costs which will give a selection rule which falls into this class. In this chapter we examine the question of selectivity bias without imposing a particular structure on
the selection process.

Condition (39) says that if costs are strictly convex then those firms which adjust follow the adjustment path resulting from the solution of the Euler equation; however some firms will not adjust. To obtain the Euler representation for the sample of firms which do not adjust we rewrite equation (34) in terms of the exogenous variables $X_a$ conditional on the instruments set $Z_a$ and the selection process

$$E(L_u | Z_a^u y_u > 0) = \psi' E(X_u | Z_a^u y_u > 0) + E(\eta_{u+1} | Z_a^u y_u > 0)$$

Define the indicator function as

$$y_u = \delta' Z_u^a + \epsilon_u$$

where $\delta$ is a vector of parameters, $Z_a^a$ is a matrix of regressors given by the instruments which will be used in the panel estimation of the Euler model, and $\epsilon_u$ is an error term, representing unobserved aspects of the decision problem. We may suppose adjustment takes place if the indicator function exceeds a critical value which, without loss of generality, we may take to be zero, i.e.

$$y_u = 1 \quad \text{iff} \quad \delta' Z_u^a + \epsilon_u > 0$$

Then (41) may be rewritten as

$$E(L_u | Z_u^a) = \psi' E(X_u | Z_u^a) + E(\eta_{u+1} | Z_u^a, \delta' Z_u^a + \epsilon_u > 0)$$

The specification of the selection process in terms of the instruments which define the orthogonality conditions in the GMM estimation of the Euler equation is one of main features of the procedure discussed in this chapter. In order to take into account the selection process which determines our observed sample we use a
GMM technique, in our final estimation, which includes correction of the residuals for selectivity. This allows us to use the inverse Mills' ratio both as a regressor and as a valid instrument in the GMM procedure. In what follows we establish how the selection process is explained by the set of instruments $Z_u$. In order to estimate the parameters in (44) we need to make distributional assumptions on the disturbance terms. The simplest assumption is that of joint normality:

$$\begin{bmatrix} \eta_{t+1}, \epsilon_{it} \end{bmatrix} | Z_u \sim N(0, \Sigma)$$

$$\Sigma = \begin{pmatrix} \sigma^2 & \sigma_{\eta \epsilon} \\ \sigma_{\eta \epsilon} & \sigma^2 \end{pmatrix}$$

In Appendix 3.2, we discuss alternative assumptions for the disturbance terms. We develop (41) in two steps. First of all notice that

$$E(\eta_{t+1} | \epsilon_{it}) = \frac{\sigma_{\eta \epsilon}}{\sigma_\epsilon} \left( \frac{\epsilon_{it} - \mu_\epsilon}{\sigma_\epsilon} \right) = \rho \frac{\sigma_\eta}{\sigma_\epsilon} \left( \frac{\epsilon_{it} - \mu_\epsilon}{\sigma_\epsilon} \right) = \omega \frac{\epsilon_{it} - \mu_\epsilon}{\sigma_\epsilon}$$

$$\omega = \rho \frac{\sigma_\eta}{\sigma_\epsilon}$$

$$\rho = \frac{\sigma_{\eta \epsilon}}{\sigma_\eta \sigma_\epsilon}$$

where $\mu_\epsilon = E(\epsilon_{it})$. Without loss of generality we may set $\mu_\epsilon = 0$ and $\sigma_\epsilon = 1$. Equation (46) then reduces to $E(\eta_{t+1} | \epsilon_{it}) = \omega \epsilon_{it}$. If $\omega = 0$ then there is no selection bias.

Second, from the selection equation we may obtain the expression for the
The OLS estimator and its asymptotic bias are respectively

\[ \hat{\psi} = \psi + \left[ \sum_{i=1}^{N} Z_i'Z_i \right]^{-1} \sum_{i=1}^{N} \left[ Z_i'(\eta_i | e_i > -Z_i\delta) \right] \]

\[ \text{plim}_{N \to \infty} \hat{\psi} = \psi + \text{plim}_{N \to \infty} \left[ \frac{1}{N} \sum_{i=1}^{N} Z_i'Z_i \right]^{-1} \text{plim}_{N \to \infty} \left[ \frac{1}{N} \sum_{i=1}^{N} \left( Z_i'(\eta_i | e_i > -Z_i\delta) \right) \right] \]

where \( Z_i \) is defined below in (53) and \( e_i \) and \( \eta_i \) are defined conformably. Assume that the probability limit of \( N^{-1} \sum Z_i'Z_i \) exists and is non-singular. Write it as \( M_{ZZ} \).

Then, note from (45) that we may write

\[ \eta_i = \frac{\sigma_{\eta e}}{\sigma_e} e_i + \omega_i \]

where

\[ \text{plim}_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \omega_i e_i = 0 \]

and that

\[ E(\epsilon_i | e_i > -Z_i\delta) = \sigma_e \lambda_i \]

where

\[ \lambda_i = \frac{\Phi \left( \frac{\delta'Z_{it}}{\sigma_e} \right)}{\Phi \left( \frac{\delta'Z_{it}}{\sigma_e} \right)} \]

[see, for example, Maddala (1983), p.365.]

Define

\[ M_{Z\lambda} = \text{plim}_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} Z_i' \lambda_i \]

Then

\[ \text{plim}_{N \to \infty} \hat{\psi} = \psi + \frac{\sigma_{\eta e}}{\sigma_e} M_{ZZ}^{-1} M_{Z\lambda} \]

If selection is non-random, \( M_{Z\lambda} \neq 0 \).
where \( \phi \) and \( \Phi \) are respectively the standard normal density function and distribution function. The ratio, \( \lambda_u \), is the mean of a standard normal truncated distribution, i.e. the inverse Mills' ratio. It measures the conditional expectation that a firm \( i \), at time \( t \), with characteristics \( X_i \), remains in the sample after selection. If \( \sigma^2 > 0 \) then \( \omega > 0 \) (\( \lambda_u \) is always positive), the bias tends to give rise to positively valued disturbances for the selected observations. Otherwise the bias has the opposite effect. If there is no correlation between \( \eta_{t+1} \) and \( \Delta \), i.e. \( \rho = 0 \), there is no selection bias. Thus we may write the final specification of the Euler equation for the selected sample as

\[
E(L_u | Z_u, C_u > 0) = \psi E(X_u | Z_u) + \omega \lambda (\delta'Z_u)
\]

The bias from this truncated regression depends on the extent to which \( \lambda_u \) differs from zero and it is correlated with \( X_u \) since they are both functions of \( Z_u \). In order to estimate (41) we have to calculate the inverse Mills' ratio and this is the object of the next section.

3.5. Estimation of the selection term (step 1)

To deal with the problem of selectivity we combine GMM with correction of residuals for selectivity in the following two step procedure.
Step 1. We run a probit on the entire sample where the dependent variable \( y_i \) is equal to 1 if the company adjusts (if there is either positive or negative change in employment) and equal to 0 if the company does not adjust employment. The probability of an uncensored observation is given by

\[
p^A = pr(y_i = 1) = pr(\delta'Z_u + \epsilon_u > 0) = \Phi(-\delta'Z_u)
\]

where \( \Phi \) is the standard normal cumulative distribution.

Regressors are given by the set of instruments \( Z_u \): lagged values, in levels, of employment (number of employees), \( L \), real sales, \( Q \), the real wage-bill, \( WB \), investment in machinery (in real terms), \( INV \), profitability (in real terms), \( P \), and firm size, \( S \) (measured as the reciprocal of employment, \( 1/L \)). This represents exactly the set of instruments we employ in the estimation of the Euler equation at the next stage. In anticipation of what we will then do, we choose output and the wage-bill rather lagged values of productivity and real wages, which are divided by employment, in order to avoid to use variables that may be affected by measurement errors in employment, as explained in Section 3.3.

Estimation of the selection term is carried out consistently with our final objective: panel data estimation of the Euler equation as shown in Figure 36 in Section 3.2. This involves three specifications:

Model "(t-1), (t-2), (t-3)"; we run probits for each year using regressors dated (t-1) at the beginning of the time series, (t-1), (t-2) the following year and then (t-1), (t-2), (t-3) for the rest of the time series. This prevents us from losing too many years’ data in the construction of the inverse Mills’ ratio. In more detail, our time series starts in 1982 and finishes in 1989, and hence we run the first probit for 1983 with regressors dated 1982. In the following year, 1984, we run the probit including
regressors dated (1982) and (1983). For 1985, we may exploit all the lags and use regressors dated (1982), (1983) and (1984). The complete set of lagged regressors is applied to all subsequent years.

Model "(t-2) (t-3)" is run using the same method but, in the first year, it includes only regressors dated (t-2). The estimation starts in 1984 and uses regressors dated 1982. For subsequent years the regressors dated (t-2) and (t-3) are used.

Model "(t-3)" is also run with regressors dated t-3.

The estimates are shown in Tables XVI-XVIII. From these estimates we may calculate the inverse Mills' ratios which represent possible selectivity bias in our estimation. Results are shown in the following tables. Standard errors are given in parentheses. In all models, history stops at t-3. The inclusion of t-4 and earlier periods did not add any explanatory power.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_{t-1}$</td>
<td>0.005</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>$L_{t-2}$</td>
<td>0.002</td>
<td>-0.001</td>
<td>0.003</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td></td>
</tr>
<tr>
<td>$L_{t-3}$</td>
<td></td>
<td></td>
<td>0.003</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.002)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>$Q_{t-1}$</td>
<td>-0.0004</td>
<td>-0.003</td>
<td>0.001</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(0.0004)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>$Q_{t-2}$</td>
<td>0.003</td>
<td>-0.001</td>
<td>-0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td></td>
</tr>
<tr>
<td>$Q_{t-3}$</td>
<td></td>
<td>-0.001</td>
<td>0.003</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td></td>
</tr>
<tr>
<td>$WB_{t-1}$</td>
<td>-0.005</td>
<td>0.006</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.018)</td>
<td>(0.017)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>$WB_{t-2}$</td>
<td>-0.018</td>
<td>0.018</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.020)</td>
<td>(0.025)</td>
<td></td>
</tr>
<tr>
<td>$WB_{t-3}$</td>
<td></td>
<td>-0.011</td>
<td>-0.0002</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.017)</td>
<td>(0.020)</td>
<td></td>
</tr>
<tr>
<td>$INV_{t-1}$</td>
<td>4.496</td>
<td>2.512</td>
<td>2.572</td>
<td>0.727</td>
</tr>
<tr>
<td></td>
<td>(1.183)</td>
<td>(1.591)</td>
<td>(1.652)</td>
<td>(1.598)</td>
</tr>
<tr>
<td>$INV_{t-2}$</td>
<td>0.448</td>
<td>2.034</td>
<td>1.652</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.318)</td>
<td>(1.377)</td>
<td>(1.944)</td>
<td></td>
</tr>
<tr>
<td>$INV_{t-3}$</td>
<td></td>
<td>1.564</td>
<td>4.381</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.225)</td>
<td>(1.852)</td>
<td></td>
</tr>
<tr>
<td>$P_{t-1}$</td>
<td>-1.093</td>
<td>-1.502</td>
<td>0.317</td>
<td>-0.684</td>
</tr>
<tr>
<td></td>
<td>(0.395)</td>
<td>(1.454)</td>
<td>(1.089)</td>
<td>(1.375)</td>
</tr>
<tr>
<td>$P_{t-2}$</td>
<td>-0.181</td>
<td>0.907</td>
<td>0.855</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.234)</td>
<td>(1.399)</td>
<td>(1.457)</td>
<td></td>
</tr>
<tr>
<td>$P_{t-3}$</td>
<td></td>
<td>-0.579</td>
<td>-0.325</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.879)</td>
<td>(1.330)</td>
<td></td>
</tr>
<tr>
<td>$S_{t-1}$</td>
<td>-0.313</td>
<td>-1.209</td>
<td>0.567</td>
<td>-2.517</td>
</tr>
<tr>
<td></td>
<td>(0.423)</td>
<td>(0.891)</td>
<td>(0.893)</td>
<td>(1.613)</td>
</tr>
<tr>
<td>$S_{t-2}$</td>
<td>-1.039</td>
<td>-2.774</td>
<td>0.240</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.718)</td>
<td>(1.299)</td>
<td>(1.109)</td>
<td></td>
</tr>
<tr>
<td>$S_{t-3}$</td>
<td></td>
<td>0.195</td>
<td>-1.119</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.755)</td>
<td>(1.144)</td>
<td></td>
</tr>
<tr>
<td>const</td>
<td>0.318</td>
<td>0.657</td>
<td>-0.627</td>
<td>0.584</td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>(0.051)</td>
<td>(0.047)</td>
<td>(0.057)</td>
</tr>
<tr>
<td>log likelihood</td>
<td>-1842.7</td>
<td>-1511.9</td>
<td>-1638.9</td>
<td>-1627.5</td>
</tr>
<tr>
<td>no. obs.</td>
<td>3247</td>
<td>3247</td>
<td>3247</td>
<td>3247</td>
</tr>
<tr>
<td>McFadden's pseudo R²</td>
<td>0.043</td>
<td>0.060</td>
<td>0.041</td>
<td>0.060</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>168.60</td>
<td>190.98</td>
<td>140.60</td>
<td>208.43</td>
</tr>
<tr>
<td>d.f.</td>
<td>(6)</td>
<td>(12)</td>
<td>(18)</td>
<td>(18)</td>
</tr>
<tr>
<td>prob&gt;$\chi^2$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Table XVI: Model "(t-1) (t-2) (t-3)" cont.

<table>
<thead>
<tr>
<th></th>
<th>1987</th>
<th>1988</th>
<th>1989</th>
</tr>
</thead>
<tbody>
<tr>
<td>( L_{t1} )</td>
<td>0.005</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>( L_{t2} )</td>
<td>-0.010</td>
<td>0.003</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>( L_{t3} )</td>
<td>0.006</td>
<td>-0.004</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>( Q_{t1} )</td>
<td>0.001</td>
<td>-0.0005</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>( Q_{t2} )</td>
<td>-0.003</td>
<td>0.001</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>( Q_{t3} )</td>
<td>-0.002</td>
<td>-0.001</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>( WB_{t1} )</td>
<td>-0.024</td>
<td>-0.001</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.020)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>( WB_{t2} )</td>
<td>-0.017</td>
<td>0.014</td>
<td>0.035</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.027)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>( WB_{t3} )</td>
<td>0.003</td>
<td>-0.002</td>
<td>-0.019</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.019)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>( INV_{t1} )</td>
<td>0.031</td>
<td>-0.096</td>
<td>-0.153</td>
</tr>
<tr>
<td></td>
<td>(1.592)</td>
<td>(1.434)</td>
<td>(1.145)</td>
</tr>
<tr>
<td>( INV_{t2} )</td>
<td>1.227</td>
<td>5.086</td>
<td>0.217</td>
</tr>
<tr>
<td></td>
<td>(1.663)</td>
<td>(1.968)</td>
<td>(1.505)</td>
</tr>
<tr>
<td>( INV_{t3} )</td>
<td>-1.912</td>
<td>-0.582</td>
<td>1.141</td>
</tr>
<tr>
<td></td>
<td>(1.574)</td>
<td>(1.489)</td>
<td>(1.634)</td>
</tr>
<tr>
<td>( P_{t1} )</td>
<td>-0.151</td>
<td>0.753</td>
<td>0.507</td>
</tr>
<tr>
<td></td>
<td>(1.305)</td>
<td>(1.222)</td>
<td>(1.431)</td>
</tr>
<tr>
<td>( P_{t2} )</td>
<td>-0.001</td>
<td>-3.319</td>
<td>1.518</td>
</tr>
<tr>
<td></td>
<td>(1.266)</td>
<td>(1.730)</td>
<td>(1.966)</td>
</tr>
<tr>
<td>( P_{t3} )</td>
<td>0.740</td>
<td>1.330</td>
<td>-2.402</td>
</tr>
<tr>
<td></td>
<td>(1.218)</td>
<td>(1.209)</td>
<td>(1.662)</td>
</tr>
<tr>
<td>( S_{t1} )</td>
<td>-1.147</td>
<td>-1.416</td>
<td>2.375</td>
</tr>
<tr>
<td></td>
<td>(1.711)</td>
<td>(1.515)</td>
<td>(1.994)</td>
</tr>
<tr>
<td>( S_{t2} )</td>
<td>-4.194</td>
<td>-2.336</td>
<td>-6.978</td>
</tr>
<tr>
<td></td>
<td>(2.415)</td>
<td>(2.859)</td>
<td>(2.763)</td>
</tr>
<tr>
<td>( S_{t3} )</td>
<td>2.158</td>
<td>-0.865</td>
<td>1.679</td>
</tr>
<tr>
<td></td>
<td>(1.189)</td>
<td>(2.363)</td>
<td>(2.207)</td>
</tr>
<tr>
<td>const</td>
<td>-0.703</td>
<td>0.852</td>
<td>0.830</td>
</tr>
<tr>
<td></td>
<td>(0.555)</td>
<td>(0.055)</td>
<td>(0.058)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>log likelihood</th>
<th>-1558.6</th>
<th>-1547.0</th>
<th>-1457.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>no. obs.</td>
<td>3247</td>
<td>3247</td>
<td>3247</td>
</tr>
<tr>
<td>McFadden's pseudo R^2</td>
<td>0.047</td>
<td>0.045</td>
<td>0.044</td>
</tr>
<tr>
<td>( \chi^2 )</td>
<td>156.05</td>
<td>146.18</td>
<td>134.92</td>
</tr>
<tr>
<td>d.f.</td>
<td>(18)</td>
<td>(18)</td>
<td>(18)</td>
</tr>
<tr>
<td>prob&gt;( \chi^2 )</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

157
Table XVII Model \"(t-2)(t-3)\"

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_{t-2}$</td>
<td>0.001</td>
<td>-0.001</td>
<td>0.002</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>$L_{t-3}$</td>
<td>0.003</td>
<td>0.001</td>
<td>0.005</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>$Q_{t-2}$</td>
<td>0.0003</td>
<td>0.001</td>
<td>-0.002</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>$Q_{t-3}$</td>
<td>-0.002</td>
<td>0.003</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>$WB_{t-2}$</td>
<td>0.024</td>
<td>0.015</td>
<td>0.001</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.017)</td>
<td>(0.020)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>$WB_{t-3}$</td>
<td>-0.006</td>
<td>0.001</td>
<td>0.005</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.019)</td>
<td>(0.019)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>$INV_{t-2}$</td>
<td>0.777</td>
<td>-1.163</td>
<td>1.688</td>
<td>1.732</td>
</tr>
<tr>
<td></td>
<td>(1.275)</td>
<td>(1.298)</td>
<td>(1.870)</td>
<td>(1.609)</td>
</tr>
<tr>
<td>$INV_{t-3}$</td>
<td>1.809</td>
<td>4.438</td>
<td>-1.991</td>
<td>1.514</td>
</tr>
<tr>
<td></td>
<td>(1.198)</td>
<td>(1.846)</td>
<td>(1.514)</td>
<td>(1.514)</td>
</tr>
<tr>
<td>$P_{t-2}$</td>
<td>-0.859</td>
<td>1.472</td>
<td>-0.519</td>
<td>0.499</td>
</tr>
<tr>
<td></td>
<td>(0.547)</td>
<td>(1.158)</td>
<td>(1.251)</td>
<td>(1.229)</td>
</tr>
<tr>
<td>$P_{t-3}$</td>
<td>-0.678</td>
<td>-0.371</td>
<td>0.608</td>
<td>0.608</td>
</tr>
<tr>
<td></td>
<td>(0.890)</td>
<td>(1.256)</td>
<td>(1.268)</td>
<td>(1.268)</td>
</tr>
<tr>
<td>$S_{t-2}$</td>
<td>-1.831</td>
<td>-2.274</td>
<td>-0.715</td>
<td>-5.384</td>
</tr>
<tr>
<td></td>
<td>(0.468)</td>
<td>(0.976)</td>
<td>(0.855)</td>
<td>(1.646)</td>
</tr>
<tr>
<td>$S_{t-3}$</td>
<td>0.240</td>
<td>-1.923</td>
<td>2.133</td>
<td>1.811</td>
</tr>
<tr>
<td></td>
<td>(0.749)</td>
<td>(0.997)</td>
<td>(1.181)</td>
<td>(1.181)</td>
</tr>
<tr>
<td>const</td>
<td>0.640</td>
<td>0.632</td>
<td>0.548</td>
<td>0.716</td>
</tr>
<tr>
<td></td>
<td>(0.478)</td>
<td>(0.047)</td>
<td>(0.053)</td>
<td>(0.055)</td>
</tr>
</tbody>
</table>

| log likelihood | -1516.1 | -1640.8 | -1629.4 | -1564.3 |
| no. obs.       | 3247    | 3247    | 3247    | 3247    |
| McFadden's pseudo $R^2$ | 0.057 | 0.040 | 0.059 | 0.044 |
| $\chi^2$      | 182.69  | 136.76  | 204.54  | 144.64  |
| d.f.           | (6)     | (12)    | (12)    | (12)    |
| prob$>\chi^2$ | 0.000   | 0.000   | 0.000   | 0.000   |

The dependent variable is 1 if the firm adjusts and 0 otherwise: The regressors are: lagged values, in levels, of employment (number of employees), $L$, real sales, $Q$, the real wage-bill, $WB$, investment in machinery (in real terms), $INV$, profitability (in real terms), $P$, and firm size, $S$ measured as the reciprocal of employment, $1/L$.

$\chi^2$, with the degrees of freedom in parentheses, is a test of the joint significance of all regressors except the constant.
Table XVII Model "+(t-2)(t-3)" cont.

<table>
<thead>
<tr>
<th></th>
<th>1988</th>
<th>1989</th>
</tr>
</thead>
<tbody>
<tr>
<td>L_{t-2}</td>
<td>0.004</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>L_{t-3}</td>
<td>-0.004</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Q_{t-2}</td>
<td>0.0005</td>
<td>-0.0001</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Q_{t-3}</td>
<td>-0.001</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>WB_{t-2}</td>
<td>0.014</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>WB_{t-3}</td>
<td>-0.004</td>
<td>-0.024</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>INV_{t-2}</td>
<td>5.180</td>
<td>0.255</td>
</tr>
<tr>
<td></td>
<td>(1.823)</td>
<td>(1.459)</td>
</tr>
<tr>
<td>INV_{t-3}</td>
<td>-0.631</td>
<td>0.944</td>
</tr>
<tr>
<td></td>
<td>(1.483)</td>
<td>(1.573)</td>
</tr>
<tr>
<td>P_{t-2}</td>
<td>-2.397</td>
<td>2.062</td>
</tr>
<tr>
<td></td>
<td>(1.138)</td>
<td>(1.566)</td>
</tr>
<tr>
<td>P_{t-3}</td>
<td>1.256</td>
<td>-2.431</td>
</tr>
<tr>
<td></td>
<td>(1.201)</td>
<td>(1.598)</td>
</tr>
<tr>
<td>S_{t-2}</td>
<td>-4.029</td>
<td>-5.093</td>
</tr>
<tr>
<td></td>
<td>(2.149)</td>
<td>(2.170)</td>
</tr>
<tr>
<td>S_{t-3}</td>
<td>-0.540</td>
<td>2.171</td>
</tr>
<tr>
<td></td>
<td>(2.227)</td>
<td>(2.094)</td>
</tr>
<tr>
<td>const</td>
<td>0.852</td>
<td>0.829</td>
</tr>
<tr>
<td></td>
<td>(0.054)</td>
<td>(0.058)</td>
</tr>
<tr>
<td>log likelihood</td>
<td>-1547.9</td>
<td>-1460.0</td>
</tr>
<tr>
<td>no. obs.</td>
<td>3247</td>
<td>3247</td>
</tr>
<tr>
<td>McFadden’s pseudo R²</td>
<td>0.044</td>
<td>0.043</td>
</tr>
<tr>
<td>χ²</td>
<td>144.29</td>
<td>130.60</td>
</tr>
<tr>
<td>d.f.</td>
<td>(12)</td>
<td>(12)</td>
</tr>
<tr>
<td>prob&gt;χ²</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The dependent variable is 1 if the firm adjusts and 0 otherwise: The regressors are: lagged values, in levels, of employment (number of employees), L, real sales, Q, the real wage-bill, WB, investment in machinery (in real terms), INV, profitability (in real terms), P, and firm size, S measured as the reciprocal of employment, 1/L.

χ², with the degrees of freedom in parentheses, is a test of the joint significance of all regressors except the constant.
The dependent variable is 1 if the firm adjusts and 0 otherwise: The regressors are: lagged values, in levels, of employment (number of employees), L, real sales, Q, the real wage-bill, WB, investment in machinery (in real terms), INV, profitability (in real terms), P, and firm size, S measured as the reciprocal of employment, 1/L.

$\chi^2$, with the degrees of freedom in parentheses, is a test of the joint significance of all regressors except the constant.

Goodness of fit is measured using McFadden's pseudo $R^2$ statistic. Although low, the tabulated statistics are not out of line with levels typical for models of binary data. In view of the statistical significance of some of the coefficients we may proceed on the basis that selection appears reasonably well explained by the instruments which characterize the orthogonality conditions in the regression model.

In general, selectivity is explained by the variable size, $S_{t_2}$ and $S_{t_3}$, new investment...
in machinery, \( \text{INV}_{t_2} \) and \( \text{INV}_{t_3} \), and, in some cases, profitability \( \text{P}_{t_1} \), \( \text{P}_{t_2} \) and employment \( \text{L}_{t_1} \), \( \text{L}_{t_2} \) and \( \text{L}_{t_3} \). The variable size is in every case estimated to have a negative coefficient indicating that the probability of adjusting decreases for small size companies, which implies that small firms are more constrained in varying the labour input than are large companies. On the one hand, technological constraints may not allow small firms to operate with fewer than a certain number of employees; on the other hand, hiring and firing costs (for instance, training or dismissals) may be relatively large for small companies and induce them thereby to postpone adjustment decisions. It is also possible that very small firms may decide to remain under the threshold of 15 employees in order to avoid stricter employment protection conditions and unionization. Conversely, the coefficient relating to lagged new machinery investments is positive indicating that firms who have invested in new equipment are more likely to adjust. This result is consistent with the study of the behaviour of the yearly averages of this variable across companies reported in Section 2.5 and illustrated in Figure 31. As far as profitability measure is concerned, the comparison of the coefficients across the probits does not show a clear, stable relation between employment variation and profits.

We also test the joint significance of the coefficients in the three models. At the foot of the tables we report the values of the \( \chi^2 \) statistics, with the degree of freedom in parentheses, of the joint significance of all coefficients except the constant. In each model, all the statistics, which are automatically calculated by Stata, when executing probit and logit estimations, suggest the rejection of the null hypothesis that all coefficients are equal to zero.
In Figures 37 and 38 we plot the inverse Mills' ratios and the probabilities of adjusting calculated according to the three probits using regressors dated \((t-1)(t-2)(t-3); (t-2)(t-3);\) and \((t-3)\) respectively. We have plotted the average calculated over companies for each available year.

Figure 37 shows that there are not large differences between the selectivity terms generated by the three probits run using different regressor timings. Only the inverse Mills' ratios calculated using regressors dated \((t-3)\) appear to show a
different behaviour in periods 1986-1987 and more markedly in 1987-1988. Each set of inverse Mills' ratios, averaged across firms, seems to be quite constant across years and this suggests that, on average, this term is very small when we take first differences. However, the behaviour of the inverse Mills' ratios may be very different across firms where they exhibit considerable variability. Summary statistics are provided in Appendix 3.3.

Figure 38 Probability of selection

The probability of adjusting according to the three probit specifications, plotted in
Figure 38 is given in (46). Its behaviour closely resembles the inverse of the Mills’ ratio in Figure 37. This is given by the fact that the probability of adjusting \( \Pr(y=1) = \Phi(-s'Z) \) represents the denominator of the Mills’ ratio. In Appendix 3.4, we plot the probability of adjusting calculated on the basis of probits "(t-1)(t-2)(t-3)"; "(t-2)(t-3)"; and "(t-3)" by year and by adjustment regimes companies followed the previous period: decrease, increase and no change in employment.

In the remainder of this section we concentrate on the ability of the regressors in the probits, defined to have the timing consistent with the panel data estimation, in explaining the process of endogenous selection. In fact, these represent the instrumental variables whose timing will determine each of four the models illustrated in Figure 36. We undertake various tests of joint significance.

The first is a test of the joint significance, in the explanation of the selectivity, of groups of same regressors taken at different time. The test is only possible for models "(t-1)(t-2)(t-3)" and "(t-2)(t-3)" since model "(t-3)" only contains variables at time t-3. For example, for model "(t-1)(t-2)(t-3)" we test the joint significance of \([L_{t-1}, L_{t-2}, L_{t-3}]; [Q_{t-1}, Q_{t-2}, Q_{t-3}]; \) etc. The null hypothesis is that the coefficients of the regressors in the groups are all equal to zero. We use a Wald F test based on the inverse of the information matrix and therefore on a quadratic approximation to the likelihood function. Let \( Rb=r \) be the set of linear hypotheses to be tested jointly, where \( b \) indicates the vector of coefficients. Then the F statistics may be written as
\[ F = \frac{1}{\text{mse}} [(b-q)'A(b-q)] \tag{50} \]

where \[ q = b + A^{-1}R'(RA^{-1}R')^{-1}(r-Rb) \]

In the case of maximum likelihood, \( A \) is the matrix of second partial derivatives and mse=1. A more powerful test might be the likelihood ratio test, suitable for non-linear restrictions. However, our sample is sufficiently large that we may take the two tests as to be equivalent. In fact, asymptotically the Wald and the likelihood ratio test are equal as the two \( \chi^2 \) statistics both tend to \( rF \). [see for instance Davidson and MacKinnon, 1993] The values of the statistics are presented in the following tables.
Table XIX refers to probit "(t-1)(t-2)(t-3)" the estimates of which are presented in Table XVI in the earlier pages. The statistics is \( F(\gamma, n-k) \), where \( \gamma=\)number of restrictions, \( n=\)number of observations and \( k=\)number of coefficients estimated.

The group of variables representing size shows a high level of joint significance in all the years. The other groups follow a more differentiated pattern of significance.

For year 1984 we do not have regressors dated (t-3) since the sample period starts at 1982. The F test for regressors dated (t-1)(t-2) is presented in Table XX.
Again, the group formed by the size variable taken at different time shows a high level of significance. The null hypothesis, that the coefficients are equal to zero, is rejected also in the case of the wage-bill, WB, and profitability, P.

The same test has been carried out for variables in the probit model ",(t-2)(t-3)" the estimates of which are shown in Table XVII in the previous pages. In this model, the groups of variables are formed by regressors dated t-2 and t-3, chosen consistently with the timing of the instruments in the Generalized Method of Moments estimation for model "(t-2)(t-3)" both in levels and in differences, as illustrated in Figure 36 in Section 3.2. The results of the $F_{(n, n-k)}$ statistics are presented in the following table.

<table>
<thead>
<tr>
<th></th>
<th>$F(2,3234)$:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1984</td>
<td></td>
</tr>
<tr>
<td>$L_{t-1}$</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>$L_{t-2}$</td>
<td>[0.56]</td>
<td></td>
</tr>
<tr>
<td>$Q_{t-1}$</td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td>$Q_{t-2}$</td>
<td>[0.38]</td>
<td></td>
</tr>
<tr>
<td>$WB_{t-1}$</td>
<td>2.11</td>
<td></td>
</tr>
<tr>
<td>$WB_{t-2}$</td>
<td>[0.12]</td>
<td></td>
</tr>
<tr>
<td>$INV_{t-1}$</td>
<td>1.50</td>
<td></td>
</tr>
<tr>
<td>$INV_{t-2}$</td>
<td>[0.22]</td>
<td></td>
</tr>
<tr>
<td>$P_{t-1}$</td>
<td>2.52</td>
<td></td>
</tr>
<tr>
<td>$P_{t-2}$</td>
<td>[0.08]</td>
<td></td>
</tr>
<tr>
<td>$S_{t-1}$</td>
<td>8.48</td>
<td></td>
</tr>
<tr>
<td>$S_{t-2}$</td>
<td>[0.00]</td>
<td></td>
</tr>
</tbody>
</table>

Prob > $F$ in brackets
Table XXI Test of joint significance of groups of same regressors with different timing: model "(t-2)(t-3)"

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Lₜ₂</td>
<td>2.10</td>
<td>1.85</td>
<td>2.04</td>
<td>1.11</td>
<td>0.14</td>
</tr>
<tr>
<td>Lₜ₃</td>
<td>[0.12]</td>
<td>[0.16]</td>
<td>[0.13]</td>
<td>[0.33]</td>
<td>[0.87]</td>
</tr>
<tr>
<td>Qₜ₂</td>
<td>3.09</td>
<td>1.05</td>
<td>1.21</td>
<td>0.13</td>
<td>0.01</td>
</tr>
<tr>
<td>Qₜ₃</td>
<td>[0.04]</td>
<td>[0.35]</td>
<td>[0.30]</td>
<td>[0.88]</td>
<td></td>
</tr>
<tr>
<td>WBₜ₂</td>
<td>0.63</td>
<td>0.01</td>
<td>0.50</td>
<td>0.66</td>
<td>3.43</td>
</tr>
<tr>
<td>WBₜ₃</td>
<td>[0.53]</td>
<td>[0.99]</td>
<td>[0.61]</td>
<td>[0.51]</td>
<td>[0.03]</td>
</tr>
<tr>
<td>INVₜ₂</td>
<td>1.27</td>
<td>4.41</td>
<td>1.11</td>
<td>4.30</td>
<td>0.28</td>
</tr>
<tr>
<td>INVₜ₃</td>
<td>[0.28]</td>
<td>[0.01]</td>
<td>[0.33]</td>
<td>[0.01]</td>
<td>[0.75]</td>
</tr>
<tr>
<td>Pₜ₂</td>
<td>0.98</td>
<td>0.51</td>
<td>0.87</td>
<td>3.31</td>
<td>1.16</td>
</tr>
<tr>
<td>Pₜ₃</td>
<td>[0.37]</td>
<td>[0.60]</td>
<td>[0.42]</td>
<td>[0.04]</td>
<td>[0.31]</td>
</tr>
<tr>
<td>Sₜ₂</td>
<td>6.72</td>
<td>8.41</td>
<td>8.82</td>
<td>17.01</td>
<td>6.76</td>
</tr>
<tr>
<td>Sₜ₃</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
</tr>
</tbody>
</table>

Prob > F in square brackets

In Table XXI the level of joint significance of the groups of regressors taken at time t-2 and t-3, in explaining the selection process increases, in particular in the case of profitability, P, size, S and to a lesser extent, wage-bill, WB and new machinery investments, INV. This suggests that selectivity is better explained by considering variables dated t-2 and t-3 and this should be reflected in the panel estimation we will carry out in the next section.

Another test concerns the ability of the variables employed in the probit estimation but which do not appear directly as regressors in the Euler equation in explaining the selection process. We run three sets of probits characterized by different timing of the regressors as in Tables XVI-XVIII using only INV (new
investment in machinery), P (profits) and S (size) as explanatory variables. As we
notice from equation (26) these variables do not appear directly as regressors in the
Euler equation. They will actually belong to the set of instruments in the GMM
estimation as we will show in the estimation of the Euler equation in Step 2. In the
following tables we show the results from the probits.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>INV&lt;sub&gt;t-1&lt;/sub&gt;</td>
<td>0.119</td>
<td>0.073</td>
<td>0.063</td>
<td>0.021</td>
</tr>
<tr>
<td>(0.020)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INV&lt;sub&gt;t-2&lt;/sub&gt;</td>
<td>0.032</td>
<td>0.005</td>
<td>0.045</td>
<td></td>
</tr>
<tr>
<td>(0.023)</td>
<td></td>
<td></td>
<td></td>
<td>(0.033)</td>
</tr>
<tr>
<td>INV&lt;sub&gt;t-3&lt;/sub&gt;</td>
<td></td>
<td>0.042</td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td>(0.021)</td>
<td></td>
<td></td>
<td></td>
<td>(0.031)</td>
</tr>
<tr>
<td>P&lt;sub&gt;t-1&lt;/sub&gt;</td>
<td>0.001</td>
<td>-0.93</td>
<td>0.55</td>
<td>-0.054</td>
</tr>
<tr>
<td>(0.015)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P&lt;sub&gt;t-2&lt;/sub&gt;</td>
<td>0.113</td>
<td>0.018</td>
<td>0.045</td>
<td></td>
</tr>
<tr>
<td>(0.049)</td>
<td></td>
<td></td>
<td></td>
<td>(0.060)</td>
</tr>
<tr>
<td>P&lt;sub&gt;t-3&lt;/sub&gt;</td>
<td></td>
<td>-0.027</td>
<td>0.045</td>
<td></td>
</tr>
<tr>
<td>(0.049)</td>
<td></td>
<td></td>
<td></td>
<td>(0.048)</td>
</tr>
<tr>
<td>S&lt;sub&gt;t-1&lt;/sub&gt;</td>
<td>-0.047</td>
<td>-0.070</td>
<td>0.19</td>
<td>-0.116</td>
</tr>
<tr>
<td>(0.047)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S&lt;sub&gt;t-2&lt;/sub&gt;</td>
<td>-0.049</td>
<td>-0.119</td>
<td>0.018</td>
<td></td>
</tr>
<tr>
<td>(0.023)</td>
<td></td>
<td></td>
<td></td>
<td>(0.039)</td>
</tr>
<tr>
<td>S&lt;sub&gt;t-3&lt;/sub&gt;</td>
<td></td>
<td>0.003</td>
<td>-0.065</td>
<td></td>
</tr>
<tr>
<td>(0.024)</td>
<td></td>
<td></td>
<td></td>
<td>(0.037)</td>
</tr>
<tr>
<td>const</td>
<td>0.551</td>
<td>0.877</td>
<td>0.077</td>
<td>0.077</td>
</tr>
<tr>
<td>(0.033)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log likelihood</td>
<td>-1888.1</td>
<td>-1456.5</td>
<td>-1660.4</td>
<td>-1647.1</td>
</tr>
<tr>
<td>no. obs.</td>
<td>3247</td>
<td>3247</td>
<td>3247</td>
<td>3247</td>
</tr>
<tr>
<td>d.f.</td>
<td>(3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>prob&gt;χ²</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The dependent variable is 1 if the firm adjusts and 0 otherwise: The regressors are: lagged values, in levels, of investment in machinery (in real terms), \( \text{INV} \), profitability (in real terms), \( \text{P} \), and firm size, \( \text{S} \) measured as the reciprocal of employment, \( 1/L \).

\( \chi^2 \), with the degrees of freedom in parentheses, is a test of the joint significance of all regressors except the constant.
Table XXII: Model ".(t-1) (t-2) (t-3)" cont.

<table>
<thead>
<tr>
<th></th>
<th>1987</th>
<th>1988</th>
<th>1989</th>
</tr>
</thead>
<tbody>
<tr>
<td>INV_1</td>
<td>0.024</td>
<td>0.015</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.024)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>INV_2</td>
<td>0.020</td>
<td>0.108</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.033)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>INV_3</td>
<td>-0.019</td>
<td>0.002</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.025)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>P_1</td>
<td>0.053</td>
<td>0.026</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>(0.052)</td>
<td>(0.048)</td>
<td>(0.054)</td>
</tr>
<tr>
<td>P_2</td>
<td>-0.034</td>
<td>-0.115</td>
<td>0.106</td>
</tr>
<tr>
<td></td>
<td>(0.054)</td>
<td>(0.070)</td>
<td>(0.072)</td>
</tr>
<tr>
<td>P_3</td>
<td>0.097</td>
<td>0.070</td>
<td>-0.057</td>
</tr>
<tr>
<td></td>
<td>(0.052)</td>
<td>(0.048)</td>
<td>(0.060)</td>
</tr>
<tr>
<td>S_1</td>
<td>-0.064</td>
<td>-0.054</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td>(0.053)</td>
<td>(0.061)</td>
</tr>
<tr>
<td>S_2</td>
<td>-0.156</td>
<td>-0.108</td>
<td>-0.253</td>
</tr>
<tr>
<td></td>
<td>(0.073)</td>
<td>(0.089)</td>
<td>(0.089)</td>
</tr>
<tr>
<td>S_3</td>
<td>0.075</td>
<td>-0.015</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.073)</td>
<td>(0.069)</td>
</tr>
<tr>
<td>const</td>
<td>0.864</td>
<td>0.095</td>
<td>0.966</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
<td>(0.046)</td>
<td>(0.047)</td>
</tr>
</tbody>
</table>

log likelihood  -1580.8  -1555.2  -1473.61
no. obs.        3247    3247    3247

χ²
- 111.65  1129.79  103.49
d.f.          (98)    (9)    (9)
prob>χ²        0.000   0.000   0.000
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>INV&lt;sub&gt;t-2&lt;/sub&gt;</td>
<td>0.050</td>
<td>0.018</td>
<td>0.054</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.021)</td>
<td>(0.032)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>INV&lt;sub&gt;t-3&lt;/sub&gt;</td>
<td>0.049</td>
<td>0.111</td>
<td>-0.011</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.031)</td>
<td>(0.026)</td>
<td></td>
</tr>
<tr>
<td>P&lt;sub&gt;t-2&lt;/sub&gt;</td>
<td>0.068</td>
<td>0.075</td>
<td>0.011</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.046)</td>
<td>(0.046)</td>
<td>(0.052)</td>
</tr>
<tr>
<td>P&lt;sub&gt;t-3&lt;/sub&gt;</td>
<td>-0.021</td>
<td>0.392</td>
<td>0.111</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.047)</td>
<td>(0.052)</td>
<td></td>
</tr>
<tr>
<td>S&lt;sub&gt;t-2&lt;/sub&gt;</td>
<td>-0.099</td>
<td>-0.104</td>
<td>-0.029</td>
<td>-0.223</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.031)</td>
<td>(0.027)</td>
<td>(0.049)</td>
</tr>
<tr>
<td>S&lt;sub&gt;t-3&lt;/sub&gt;</td>
<td>0.005</td>
<td>-0.104</td>
<td>0.076</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.031)</td>
<td>(0.036)</td>
<td></td>
</tr>
<tr>
<td>const</td>
<td>0.087</td>
<td>0.079</td>
<td>0.736</td>
<td>0.874</td>
</tr>
<tr>
<td></td>
<td>(0.039)</td>
<td>(0.038)</td>
<td>(0.043)</td>
<td>(0.044)</td>
</tr>
</tbody>
</table>

|                |        |        |        |        |
| log likelihood | -1554.6| -1664.6| -1650.6| -1582.7|
| no. obs.       | 3247   | 3247   | 3247   | 3247   |
| χ²             | 105.79 | 89.16  | 162.11 | 107.88 |
| d.f.           | (3)    | (6)    | (6)    | (6)    |
| prob>χ²        | 0.000  | 0.000  | 0.000  | 0.000  |

The dependent variable is 1 if the firm adjusts and 0 otherwise: The regressors are: lagged values, in levels, of investment in machinery (in real terms), INV, profitability (in real terms), P, and firm size, S measured as the reciprocal of employment, 1/L. 

χ², with the degrees of freedom in parentheses, is a test of the joint significance of all regressors except the constant.
<table>
<thead>
<tr>
<th></th>
<th>1988</th>
<th>1989</th>
</tr>
</thead>
<tbody>
<tr>
<td>INV_{t2}</td>
<td>0.116</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>INV_{t3}</td>
<td>0.004</td>
<td>0.035</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>P_{t2}</td>
<td>-0.082</td>
<td>0.111</td>
</tr>
<tr>
<td></td>
<td>(0.046)</td>
<td>(0.060)</td>
</tr>
<tr>
<td>P_{t3}</td>
<td>0.067</td>
<td>-0.054</td>
</tr>
<tr>
<td></td>
<td>(0.048)</td>
<td>(0.060)</td>
</tr>
<tr>
<td>S_{t2}</td>
<td>-0.166</td>
<td>-0.199</td>
</tr>
<tr>
<td></td>
<td>(0.069)</td>
<td>(0.070)</td>
</tr>
<tr>
<td>S_{t3}</td>
<td>-0.008</td>
<td>0.060</td>
</tr>
<tr>
<td></td>
<td>(0.070)</td>
<td>(0.066)</td>
</tr>
<tr>
<td>const</td>
<td>0.949</td>
<td>0.969</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
<td>(0.047)</td>
</tr>
<tr>
<td>log likelihood</td>
<td>-1556.1</td>
<td>-1474.4</td>
</tr>
<tr>
<td>no. obs.</td>
<td>3247</td>
<td>3247</td>
</tr>
<tr>
<td>χ²</td>
<td>127.89</td>
<td>101.96</td>
</tr>
<tr>
<td>d.f.</td>
<td>(6)</td>
<td>(6)</td>
</tr>
<tr>
<td>prob&gt;χ²</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Table XXIV Model "(t-3)"

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>INV_{t3}</td>
<td>0.063</td>
<td>0.133</td>
<td>0.010</td>
<td>0.052</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.029)</td>
<td>(0.024)</td>
<td>(0.023)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>P_{t3}</td>
<td>0.042</td>
<td>0.065</td>
<td>0.140</td>
<td>0.021</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.028)</td>
<td>(0.031)</td>
<td>(0.021)</td>
<td>(0.029)</td>
</tr>
<tr>
<td>S_{t3}</td>
<td>-0.063</td>
<td>-0.132</td>
<td>-0.062</td>
<td>-0.177</td>
<td>0.124</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.019)</td>
<td>(0.013)</td>
<td>(0.023)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>const</td>
<td>0.757</td>
<td>0.075</td>
<td>0.777</td>
<td>0.971</td>
<td>0.961</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.042)</td>
<td>(0.037)</td>
<td>(0.043)</td>
<td>(0.043)</td>
</tr>
<tr>
<td>log likelihood</td>
<td>-1673.5</td>
<td>-1652.9</td>
<td>-1596.7</td>
<td>-1568.6</td>
<td>-1483.18</td>
</tr>
<tr>
<td>no. obs.</td>
<td>3247</td>
<td>3247</td>
<td>3247</td>
<td>3247</td>
<td>3247</td>
</tr>
<tr>
<td>\chi^2</td>
<td>71.40</td>
<td>157.48</td>
<td>79.99</td>
<td>103.03</td>
<td>84.35</td>
</tr>
<tr>
<td>d.f.</td>
<td>(3)</td>
<td>(3)</td>
<td>(3)</td>
<td>(3)</td>
<td>(3)</td>
</tr>
<tr>
<td>prob &gt; \chi^2</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The dependent variable is 1 if the firm adjusts and 0 otherwise. The regressors are: lagged values, in levels, of investment in machinery (in real terms), INV, profitability (in real terms), P, and firm size, S measured as the reciprocal of employment, 1/L.

\chi^2, with the degrees of freedom in parentheses, is a test of the joint significance of all regressors except the constant.

We consider these probits as representing the unrestricted model while the probits run on the complete set of regressors, in Tables XIV-XVIII as the restricted model. We may, thus, test the joint significance of the variables INV, P and S in explaining selection using a likelihood ratio test. This is equal to twice the difference between the restricted and the unrestricted values of the log-likelihood functions. The null hypothesis implies that restrictions are true. Under the null LR follows a \chi^2 distribution with the degree of freedom given by the number of restrictions. We report the values of the test in the following table.
**Table XXV Log-likelihood ratio test**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2$</td>
<td>90.8</td>
<td>69.2</td>
<td>43.0</td>
<td>39.2</td>
<td>44.4</td>
<td>16.4</td>
<td>31.4</td>
</tr>
<tr>
<td>d.f. (3)</td>
<td>(6)</td>
<td>(9)</td>
<td>(9)</td>
<td>(9)</td>
<td>(9)</td>
<td>(9)</td>
<td>(9)</td>
</tr>
<tr>
<td>p&gt;$\chi^2$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.20</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2$</td>
<td>77.0</td>
<td>47.6</td>
<td>42.4</td>
<td>36.8</td>
<td>16.4</td>
<td>28.0</td>
</tr>
<tr>
<td>d.f.</td>
<td>(3)</td>
<td>(6)</td>
<td>(6)</td>
<td>(6)</td>
<td>(6)</td>
<td>(6)</td>
</tr>
<tr>
<td>p&gt;$\chi^2$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.025</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2$</td>
<td>54.4</td>
<td>42.8</td>
<td>48.0</td>
<td>51.4</td>
<td>22.4</td>
</tr>
<tr>
<td>d.f.</td>
<td>(3)</td>
<td>(3)</td>
<td>(3)</td>
<td>(3)</td>
<td>(3)</td>
</tr>
<tr>
<td>p&gt;$\chi^2$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

where $LR = 2(\hat{l} - \overline{l}) - \chi^2(r)$

where $l = \text{unrestricted log-likelihood}$

$\overline{l} = \text{restricted log-likelihood}$

and $r$ is the number of restrictions

The values of the test all reject the null hypothesis accepting the unrestricted model. This demonstrates the availability of the variables which do not enter directly as regressors in the Euler equation as identifying variables in the selection process.

A further test concerns whether a different selection process is implied for each year. For this purpose we run three pooled probits employing the same choice
of regressors as in the previous models, including time dummies indicated with d86, d87, d88 and d89. Results are reported in the following table. In all models, the test $F_{(c=4)}$ for the joint significance of the time dummies leads to a rejection of the null hypothesis according to which the coefficients of the dummies are zero. The joint significance of the year dummies implies, thus, the existence of a "year effect" in the selection process. This supports our approach to the construction of the inverse Mills’ ratio in order to preserve its time series variations. Results from the pooled probits and F tests of homogeneity across years are presented in Table XXVII and its continuation.
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(150°0)</td>
<td>(250°0)</td>
<td>(350°0)</td>
</tr>
<tr>
<td>15°0</td>
<td>25°0</td>
<td>35°0</td>
</tr>
<tr>
<td>(150°0)</td>
<td>(250°0)</td>
<td>(350°0)</td>
</tr>
<tr>
<td>15°0</td>
<td>25°0</td>
<td>35°0</td>
</tr>
<tr>
<td>(150°0)</td>
<td>(250°0)</td>
<td>(350°0)</td>
</tr>
<tr>
<td>15°0</td>
<td>25°0</td>
<td>35°0</td>
</tr>
<tr>
<td>(150°0)</td>
<td>(250°0)</td>
<td>(350°0)</td>
</tr>
<tr>
<td>15°0</td>
<td>25°0</td>
<td>35°0</td>
</tr>
<tr>
<td>(150°0)</td>
<td>(250°0)</td>
<td>(350°0)</td>
</tr>
<tr>
<td>15°0</td>
<td>25°0</td>
<td>35°0</td>
</tr>
<tr>
<td>(150°0)</td>
<td>(250°0)</td>
<td>(350°0)</td>
</tr>
<tr>
<td>15°0</td>
<td>25°0</td>
<td>35°0</td>
</tr>
<tr>
<td>(150°0)</td>
<td>(250°0)</td>
<td>(350°0)</td>
</tr>
<tr>
<td>15°0</td>
<td>25°0</td>
<td>35°0</td>
</tr>
<tr>
<td>(150°0)</td>
<td>(250°0)</td>
<td>(350°0)</td>
</tr>
<tr>
<td>15°0</td>
<td>25°0</td>
<td>35°0</td>
</tr>
<tr>
<td>(150°0)</td>
<td>(250°0)</td>
<td>(350°0)</td>
</tr>
<tr>
<td>15°0</td>
<td>25°0</td>
<td>35°0</td>
</tr>
<tr>
<td>(150°0)</td>
<td>(250°0)</td>
<td>(350°0)</td>
</tr>
<tr>
<td>15°0</td>
<td>25°0</td>
<td>35°0</td>
</tr>
<tr>
<td>(150°0)</td>
<td>(250°0)</td>
<td>(350°0)</td>
</tr>
<tr>
<td>15°0</td>
<td>25°0</td>
<td>35°0</td>
</tr>
<tr>
<td>(150°0)</td>
<td>(250°0)</td>
<td>(350°0)</td>
</tr>
<tr>
<td>15°0</td>
<td>25°0</td>
<td>35°0</td>
</tr>
<tr>
<td>(150°0)</td>
<td>(250°0)</td>
<td>(350°0)</td>
</tr>
<tr>
<td>15°0</td>
<td>25°0</td>
<td>35°0</td>
</tr>
</tbody>
</table>

**Table XXVI. Pooled profiles**
Table XXVI  Pooled probits cont.

<table>
<thead>
<tr>
<th></th>
<th>(t-1)(t-2)(t-3)</th>
<th>(t-2)(t-3)</th>
<th>(t-3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log likelihood</td>
<td>-7882.0</td>
<td>-7885.0</td>
<td>-7909.3</td>
</tr>
<tr>
<td>no. obs.</td>
<td>16,235</td>
<td>16,235</td>
<td>16,235</td>
</tr>
<tr>
<td>F</td>
<td>5.03</td>
<td>5.20</td>
<td>6.05</td>
</tr>
<tr>
<td>(r,n-k)</td>
<td>(4,16212)</td>
<td>(4,16218)</td>
<td>(4,16224)</td>
</tr>
<tr>
<td>prob&gt;F</td>
<td>0.0005</td>
<td>0.0003</td>
<td>0.000</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>708.78</td>
<td>702.87</td>
<td>654.14</td>
</tr>
<tr>
<td>d.f.</td>
<td>(22)</td>
<td>(16)</td>
<td>(10)</td>
</tr>
<tr>
<td>prob&gt;$\chi^2$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The dependent variable is 1 if the firm adjusts and 0 otherwise. The regressors are: lagged values, in levels, of employment (number of employees), L, real sales, Q, the real wage-bill, WB, investment in machinery (in real terms), INV, profitability (in real terms), P, and firm size, S measured as the reciprocal of employment, 1/L.

$\chi^2$, with the degrees of freedom in parentheses, is a test of the joint significance of all regressors except the constant.

3.6. Estimation of the Euler equation (step 2)

Step 2. We estimate (48), over the sample of firms which do adjust. In order to take into account the possible selection bias introduced by eliminating the companies characterized by zero adjustment from the sample, we include in the Euler equation the inverse Mills’ ratio, calculated in the first stage of this estimation procedure, both as a regressor and as an instrument. We may rewrite the selectivity corrected Euler equation (48) in levels explicitly as

\[ L_{t} = \Psi_{0} + \Psi_{1}L_{t-1} + \Psi_{2}L_{t-1} + \Psi_{3}(Q/L)_{t} + \Psi_{4}W_{t} + \Psi_{5} \lambda_{t} + f_{t} + u_{t} + v_{t+1} \quad (52) \]

which corresponds to equation (35) in Section 3.2; $\Psi_{0}$ is the intercept and $f_{t}$ indicates
fixed effects.

In order to estimate (52) only on companies which adjust we adopt the following method of selecting the employment variable: we keep or drop $L_i$ according to whether it is different from or equal to $L_{i,t}$. If $L_i = L_{i,t}$, the observation corresponds to the lower branch of (39) and is therefore uninformative in relation to the parameter vector $\psi$. It is arguable that we should exclude additionally observations for which $L_{i,t+1} = L_i$. However, the correct sample selection would, in that case, require that observations where $L_i = E_i L_{i,t}$ be discarded. Note, however, that the Euler equation (35) does not exclude this equality and also that, because of the realization error, $v_{i,t+1}$, the selection condition $L_i \neq E_i L_{i,t+1}$ may be quite different from the condition $L_i \neq L_{i,t+1}$.

The selection method adopted creates gaps in the dataset corresponding to the equalities between $L_i$ and $L_{i,t}$. We then estimate the Euler equation in levels using all the remaining observations. In the differenced equation we consider at least two consecutive adjustment periods. In fact, in order to eliminate the fixed effects $f_i$ we need to take differences and this requires that we use only observations

---

37 A complication in considering limited dependent variables in panel data is given by the presence of the fixed effect as conditioning variable and by the fact that we only observe sample moments conditional on selection. In particular, in a dynamic context, we would have

$$E(y^*_{it} | y^*_{it-1}, \ldots, y^*_{i,t-1}) = \alpha y^*_{i,t-1} + E(f_i | y^*_{it}, \ldots, y^*_{i,t-1})$$

where $y^*_{it}$ is observable subject to endogenous selection:

$$y^*_{it} = d_{it} y^*_{it}$$

If selectivity interacts with the fixed effect then the selection model is unidentified in the absence of additional prior restrictions on the distribution of the latent variable. We do not deal with this problem here. For a complete account see Labeaga (1992) and Arellano, Bover and Labeaga (1993).
corresponding to successive changes i.e. $L_t$ different from $L_{t-1}$ and $L_{t-2}$ different from $L_{t-3}$. This selection of the observations creates many gaps within the time-series corresponding to each firm. These gaps must also be taken into account when constructing the instruments.

We need to modify the definition of the GMM estimator to take into account selection. Write $S_i (i=1,...,N)$ as the diagonal selection matrix ($s_{it}$) where $s_{it}=1$ if observation $t$ is included for firm $i$ and $s_{it}=0$ if this observation is omitted. Write

$$Z_i = \begin{bmatrix} Z_{1i}^t & 0' & \ldots & 0' \\ \vdots & \ddots & \ddots & \vdots \\ 0' & 0' & \ldots & Z_{ti}^t \end{bmatrix} \quad Z = \begin{bmatrix} Z_1 \\ \vdots \\ Z_N \end{bmatrix} \quad \text{and} \quad S = \begin{bmatrix} S_1 & 0 \\ \vdots & \vdots \\ 0 & S_N \end{bmatrix}$$ (53)

Then the GMM estimator (33), modified for selection becomes

$$\min_{\phi} e' S Z \left[ \sum_{i=1}^{N} Z_i \phi S_i Z_i \right]^{-1} Z' Se$$ (54)

where the matrix of regressor variables $X$ is expanded to include the inverse Mills' ratio.

We have written a program in Gauss to implement the GMM estimator (54) for our problem. The estimates we report adopt a slightly more general procedure than that implemented in the Arellano and Bond program DPD. DPD imposes a common intercept on the implied reduced form regression for the instruments for each year in the sample, despite the fact that the number of instruments in these regressions increases through the sample. We drop this restriction which is
unnecessary in view of the large number of observations at our disposal. In fact, however, the effect on the estimated coefficient values is minor. The program is reported in Appendix 3.5.

We include as instruments the inverse Mills' ratio at time t, calculated according to the timing of the instruments, lagged values of employment, output, the wage-bill, investment, profitability and size. All instruments are dated according to the specifications shown in Figure 36 in Section 3.2. We use the wage-bill and employment separately as instruments, rather than the lagged per-capita wage, because the wage-bill variable is less affected by possible measurement errors problems affecting employment, as discussed above. We follow the same approach for productivity, using lagged values of output and employment. The selection term, $\lambda_\pi$ is included both as regressor and as instrument calculated according to the timing of the instruments as explained in Section 3.5. The timing of the inverse Mills' ratio is consistent with that in the decision rule adopted in the selection of the observations in the sample in line with the Euler equation. For the equation in levels we use the inverse Mills' ratio in levels with the same timing is the same as the dependent variable. In the case of the differenced equation, the selection we operate takes into account two consecutive periods. We then take the inverse Mills' ratio in first differences ($\lambda_{it-1}$).

To see this return to Section 3.4. where we discussed the selection process. For clarity, rewrite equation (39)

\[ L_u = \begin{cases} L^*_u = \psi L_u + \eta_{it+1} & \text{if} \quad y_u = 1 \\ L_{i-1} & \text{if} \quad y_u = 0 \end{cases} \]  \hspace{1cm} (55)

The Euler representation expressed in first differences is given by

181
\[ \Delta L_u = \psi' \Delta X_u + \Delta \eta_u \]

s.t. \[ y_u = 1 \iff \delta' \Delta Z_u + \epsilon_u > 0 \] (56)

where \[ E(\eta_{u+1} \epsilon_u) = \sigma_{\epsilon} \]

By selecting two consecutive periods of non-adjustment we need to evaluate

\[ E(\Delta \eta_{u+1} \mid y_u = y_{u-1} = 1) = E(\Delta \eta_{u+1} \mid \epsilon_u > -\delta' \Delta Z_u \text{ and } \epsilon_{u-1} > -\delta' \Delta Z_{u-1}) \]

Assume that

\[ E(\eta_u \epsilon_u) = E(\eta_{u+1} \epsilon_{u-1}) = 0 \] (58)

Then

\[ E(\Delta \eta_{u+1} \mid y_u = y_{u-1} = 1) = E(\eta_{u+1} \mid \epsilon_u > -\delta' \Delta Z_{u-1}) - E(\eta_u \mid \epsilon_{u-1} > -\delta' \Delta Z_{u-1}) = \omega (\lambda_u - \lambda_{u-1}) = \omega \Delta \lambda_u \] (59)

where \[ \omega = \frac{\sigma_\epsilon}{\sigma_\eta} \quad \text{and} \quad \rho = \frac{\sigma_{\eta \epsilon}}{\sigma_\eta \sigma_\epsilon} \]

where \( \omega \) is constant over time. Equation (59) explains the use of the inverse Mills' ratio in differences in the difference equation.

The estimates of the four models illustrated in Figure 36 are reported in the following table.
Table XXVII Euler equation

<table>
<thead>
<tr>
<th>Timing of the instruments:</th>
<th>LEVELS</th>
<th>DIFFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>(t-1)(t-2)(t-3)</td>
<td>(t-2)(t-3)</td>
<td>(t-2)(t-3)</td>
</tr>
<tr>
<td><strong>const.</strong></td>
<td>-0.042 (0.027)</td>
<td>-0.020 (0.036)</td>
</tr>
<tr>
<td>L_{4,t}</td>
<td>0.553 (0.044)</td>
<td>0.524 (0.046)</td>
</tr>
<tr>
<td>L_{4,t-1}</td>
<td>0.443 (0.045)</td>
<td>0.472 (0.045)</td>
</tr>
<tr>
<td>(Q/L)_t</td>
<td>-0.001 (0.002)</td>
<td>-0.001 (0.003)</td>
</tr>
<tr>
<td>W_t</td>
<td>0.042 (0.025)</td>
<td>0.022 (0.030)</td>
</tr>
<tr>
<td>λ_t</td>
<td>0.003 (0.006)</td>
<td>0.0003 (0.010)</td>
</tr>
<tr>
<td>no. obs.</td>
<td>15,186</td>
<td>12,850</td>
</tr>
<tr>
<td>std. error of est.</td>
<td>0.150</td>
<td>0.146</td>
</tr>
<tr>
<td>first order serial correl.</td>
<td>-0.514</td>
<td>-0.548</td>
</tr>
<tr>
<td>Sargan test</td>
<td>634.41 (132)</td>
<td>384.71 (94)</td>
</tr>
<tr>
<td>Probab. value</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

1. Heteroscedastic consistent standard errors in parentheses.
2. The Sargan tests are tests for the validity of instruments with degrees of freedom given in.
Standard errors are corrected in order to take into account the presence of possible heteroscedasticity. The first two columns are specified in levels and the last two in differences. The sign and the magnitude of the coefficients of the lagged dependent variable, \( L_{t-1} \) and the forward term, \( L_{t+1} \), are, in all the four models, as predicted by the theory. As shown in (27) in Section 3.1, both \( \psi_1 \) and \( \psi_2 \) should be less than one and greater than zero. In all models the two regressors are significant. The specification in differences using instruments dated \( (t-2)(t-3) \), gives the correct sign for the coefficient of productivity but neither this coefficient nor that of wages variable appears to be significant.

The basic property of the Euler equation implying that employment is stable forwards is satisfied. For the four models we report the characteristic roots, \( \mu_1 \) and \( \mu_2 \), calculated on the basis of equation (25) in Section 3.1.

**Table XXVIII Characteristic roots**

<table>
<thead>
<tr>
<th></th>
<th>Levels</th>
<th>Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>( (t-1)(t-2)(t-3) )</td>
<td>1.039</td>
<td>1.063</td>
</tr>
<tr>
<td>( (t-2)(t-3) )</td>
<td>1.146</td>
<td>2.282</td>
</tr>
<tr>
<td>( (t-2)(t-3) )</td>
<td>0.675</td>
<td>0.721</td>
</tr>
<tr>
<td>( (t-3) )</td>
<td>0.769</td>
<td>0.849</td>
</tr>
</tbody>
</table>

In all the specifications one of the two characteristic roots falls outside the unit

---

38 Specifically we estimate

\[
\text{var}(\psi) = X'\Sigma Z_i S_i H S_i Z_i^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[ s_i^{i}Z_i Z_i' e_u e_u + s_i^{i} S_i S_i,1_{t+1} \right]
\]

\[
\left(Z_i Z_i' + Z_i S_i H S_i Z_i\right) e_u e_u^{-1} \left(\sum_{i=1}^{N} Z_i S_i H S_i Z_i\right)^{-1} Z'SX
\]

184
circle and the other falls inside the unit circle indicating that a forward stable optimal solution has been obtained.

With regard to the fixed effect the differences between the estimates in levels and first differences strongly suggest that fixed effects may be important. In order to check for the presence of fixed effects we have compared i) model (t-1)(t-2)(t-3) in levels with model (t-2)(t-3) in differences and ii) model (t-2)(t-3) in levels with model (t-3) in differences implementing the Hausman specification test. We compare the estimator in levels, $\psi^{lev}$ and the estimator in differences, $\psi^{dif}$. The former is asymptotically efficient under the null hypothesis of fixed effect uncorrelated with all regressors, $H_0: E(f_t | X_t) = 0$, but loses consistency under the alternative hypothesis, and the latter is asymptotically less efficient than the first under the null hypothesis but remains consistent under the alternative. The Hausman test statistic is defined by

$$\frac{(\psi^{lev} - \psi^{dif})' [\hat{V}^{-1}(\psi^{lev} - \psi^{dif})](\psi^{lev} - \psi^{dif})}{(60)}$$

where $V$ is a consistent estimator of the variance-covariance matrix of $\psi^{dif} - \psi^{lev}$. Under the null hypothesis the test is asymptotically distributed as chi-squared with degrees of freedom equal to the dimension of the vector $\psi$. We report the values of the test in Table XXXIX.

The model in levels is accepted (i.e. fails to be rejected) by the first statistic, relating to the comparison between model (t-1)(t-2)(t-3) in levels and (t-3)(t-3) in differences, and rejected by the second value of the test, relating to the comparison between the model in (t-2)(t-3) in levels and (t-3) in differences. However, when we look at the Sargan tests for the overidentifying restrictions, given at the foot of Table XXVII, we see that the chi-squared values suggest a rejection of the
Table XXIX Hausman test

<table>
<thead>
<tr>
<th>Estimators:</th>
<th>( \psi^{iv}: (t-1)(t-2)(t-3) )</th>
<th>( \psi^{iv}: (t-2)(t-3) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(type of model)</td>
<td>( \psi^{ad}: (t-2)(t-3) )</td>
<td>( \psi^{ad}: (t-3) )</td>
</tr>
<tr>
<td>Hausman test:</td>
<td>( \chi^2 )</td>
<td>0.84</td>
</tr>
<tr>
<td>prob. value</td>
<td></td>
<td>0.975</td>
</tr>
</tbody>
</table>

Instruments for both the equations in levels. In the case of the first differenced models, by contrast, the values of the test suggest the acceptance of the null hypothesis of orthogonality of the instruments. The Sargan test is given by the following quadratic form (see Arellano and Bond, 1991):

\[
Sargan = e' S Z \left( \sum_{i=1}^{N} Z_i' S_i H S_i Z_i' \right)^{-1} Z' S e \sim \chi_m^2
\]

where \( m \) is the number of over-identifying restrictions.

The inverse Mills’ ratio does not appear to be significant. This may be because, as pointed out earlier, since this variable does not vary much through time it becomes very close to zero when we take first differences.

From Table XXVII we may see that in all the specifications the dominant characteristic of the model is its dynamic structure: the forward and the lagged employment terms, \( L_{t-1} \) and \( L_{t-2} \), exhibit high and significant coefficients while, on the contrary, the coefficients associated with real wages and productivity, \( W_t \) and \( (Q/L)_t \), are small and insignificant (and, less importantly, given their size and significance, tend to have the wrong sign with respect to the theoretical predictions). The effect of these two last variables is clearly dominated by the dynamics. In
Chapter 7. we report the coefficients estimates also as point elasticities and we could also follow the same procedure here. However, because the coefficients are estimated as insignificant and incorrectly signed, there appears little purpose in this exercise. (The elasticities with respect to the wage rate corresponding to (52), holding productivity constant, are, in the differenced model (t-2)(t-3), of 0.04 in the short run and of 0.14 in the long run).

It seems, therefore, that employers respond to shocks on wages and productivity by varying the number of workers only if these shocks are very large. The explanation relies on the existence of high adjustment costs which make employers very cautious in adjusting employment unless shocks to wages and productivity are very large. This effect of high adjustment costs on the coefficients of wages and productivity may be understood if we go back to (27) were the coefficients of the Euler equation are expressed in terms of the parameters of the model. We assume that $\beta$ and $\alpha$, the discount rate and the parameter in the production function are well-behaved, i.e. $0 < \alpha, \beta < 1$. We may thus see that if adjustment costs are very high, $\gamma$, the parameter in the adjustment costs specification, is very high then both $\psi$, and $\psi_a$, the parameters associated with productivity and wages respectively, will be very small. In the extreme case where $\gamma$ approaches infinity, $\psi$, and $\psi_a$ will be close to zero. Hence, we either obtain no adjustment as the result of the size of adjustment costs and the lack of large shocks to wages and productivity over the period, or a very slow adjustment as the consequence of the cautiousness of employers in varying the number of workers. The latter result could, for instance, be explained by the difficulty in firing a person which leads to a very small and cautious level of hiring.
In Appendix 3.6, we report the estimates of the Euler equation with time dummies. We may consider the inclusion of time dummies as a way of taking into account macroeconomic shocks. In this case, then, idiosyncratic firms face the same macro shocks. However, in none of the four models, do the time dummies show a high level of significance. This indicates that the model does not require augmentation by variables representing the overall level of activity in the economy, or by other variables defined at the aggregate level.

3.7. The static model

In this section we implement a test for the presence of adjustment costs. We consider the following reparametrization of specification (26)

\[
\left( \frac{Q}{L} \right)_t = \theta_0 + \theta_1 L_{t-1} + \theta_2 L_{t} + \theta_3 L_{t-1} + \theta_4 W_{t} + \zeta_{t+1}
\]

where

\[
\begin{align*}
\theta_1 &= -\frac{\beta \gamma}{\alpha} \\
\theta_2 &= \frac{\gamma (1 + \beta)}{\alpha} \\
\theta_3 &= -\frac{\gamma}{\alpha} \\
\theta_4 &= \frac{1}{\alpha}
\end{align*}
\]

\(\theta_0\) is the intercept (see footnote 33 in Section 3.1.). This reparametrization allows us to test the joint significance of \(L_{\gamma+1}, L_{\alpha}, L_{\beta}\) and hence for the presence of adjustment costs. It also nests a static model given by a standard marginal productivity relation of the type
\[
\left( \frac{Q}{L} \right)_t = \theta_0 + \theta_4 W_t + \zeta_t
\]  \hspace{1cm} (64)

where \( W_t \) is real wage.

The test is the following

\[
H_0: \left( \frac{Q}{L} \right)_t = \theta_0 + \theta_4 W_t \hspace{1cm} \text{i.e.} \hspace{0.5cm} \theta_1 = \theta_2 = \theta_3 = 0
\]

\[
H_A: \left( \frac{Q}{L} \right)_t = \theta_0 + \theta_1 L_{t+1} + \theta_2 L_t + \theta_3 L_{t-1} + \theta_4 W_t \hspace{0.5cm} \text{i.e.} \hspace{0.5cm} \neg(\theta_1 = \theta_2 = \theta_3 = 0)
\]  \hspace{1cm} (65)

If \( L_{t+1}, L_t, L_{t-1} \) are jointly significant, then we reject the validity of the static model in the absence of adjustment costs.

To estimate (62) we adopt the same procedure discussed in the previous sections. We summarize the principal features:

i) we take into account the selection bias introduced by eliminating the companies with zero adjustment from the sample. We, thus, include in equation (57) the inverse Mills’ ratio calculated in Section 3.6.

ii) we estimate the model both in levels and in differences, in order to test for the presence of fixed effects.

iii) we allow for serial correlation of some order implied by different timing of the instruments. We also test for the overidentifying restrictions as illustrated in Figure 36.

iv) we estimate the model using GMM.

We, thus, obtain the following empirical specification

\[
\left( \frac{Q}{L} \right)_t = \theta_0 + \theta_1 L_{t+1} + \theta_2 L_t + \theta_3 L_{t-1} + \theta_4 W_t + \theta_5 \lambda_t + f_t + \zeta_{t+1}
\]  \hspace{1cm} (66)

and for the static model
\[
\left( \frac{Q}{L} \right)_u = \theta_0 + \theta_1 W_u + \theta_2 \lambda_u + \zeta_u
\]  

(67)

where \( \lambda_u \) is the inverse Mills' ratio and \( f_i \) denotes the fixed effect. Results are shown in the following tables: Table XXX reports the estimates of model (66) and Table XXXI presents the estimates relative of the static model (67).
Table XXX Euler equation: reparametrization

Dependent variable: \((Q/L)_t\)

<table>
<thead>
<tr>
<th></th>
<th>LEVELS</th>
<th>DIFFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Estimation in:</strong></td>
<td>LEVELS</td>
<td>DIFFERENCES</td>
</tr>
<tr>
<td>Timing of the instruments:</td>
<td>(t-1)(t-2)(t-3)</td>
<td>(t-2)(t-3)</td>
</tr>
<tr>
<td>(L_{t+1})</td>
<td>0.866</td>
<td>0.969</td>
</tr>
<tr>
<td></td>
<td>(0.368)</td>
<td>(0.600)</td>
</tr>
<tr>
<td>(L_t)</td>
<td>-0.485</td>
<td>-0.621</td>
</tr>
<tr>
<td></td>
<td>(0.638)</td>
<td>(1.033)</td>
</tr>
<tr>
<td>(L_{t-1})</td>
<td>-0.347</td>
<td>-0.222</td>
</tr>
<tr>
<td></td>
<td>(0.319)</td>
<td>(0.499)</td>
</tr>
<tr>
<td>(W_t)</td>
<td>1.467</td>
<td>1.847</td>
</tr>
<tr>
<td></td>
<td>(0.152)</td>
<td>(0.160)</td>
</tr>
<tr>
<td>(\lambda_t)</td>
<td>0.939</td>
<td>1.487</td>
</tr>
<tr>
<td></td>
<td>(0.281)</td>
<td>(0.390)</td>
</tr>
<tr>
<td><strong>no. obs.</strong></td>
<td>15,186</td>
<td>12,850</td>
</tr>
<tr>
<td><strong>std. error of est.</strong></td>
<td>0.229</td>
<td>1.251</td>
</tr>
<tr>
<td><strong>First order serial correl.</strong></td>
<td>0.809</td>
<td>0.768</td>
</tr>
<tr>
<td><strong>Sargan test</strong></td>
<td>5368.56</td>
<td>4032.41</td>
</tr>
<tr>
<td></td>
<td>(132)</td>
<td>(94)</td>
</tr>
<tr>
<td><strong>Prob. value</strong></td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

1. Heteroscedastic consistent standard errors in parentheses.
2. The Sargan tests are tests for the validity of instruments with degrees of freedom given in brackets.
Table XXXI Static model

<table>
<thead>
<tr>
<th>Estimation in:</th>
<th>LEVELS</th>
<th>DIFFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timing of the instruments:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(t-1)(t-2)(t-3)</td>
<td>(t-2)(t-3)</td>
<td>(t-2)(t-3)</td>
</tr>
<tr>
<td>const.</td>
<td>-1.588 (0.239)</td>
<td>-2.244 (0.272)</td>
</tr>
<tr>
<td>$W_t$</td>
<td>1.661 (0.113)</td>
<td>1.999 (0.123)</td>
</tr>
<tr>
<td>$\lambda_t$</td>
<td>0.911 (0.172)</td>
<td>1.230 (0.197)</td>
</tr>
<tr>
<td>no. obs.</td>
<td>15,186</td>
<td>12,850</td>
</tr>
<tr>
<td>std. error of est.</td>
<td>1.215</td>
<td>1.241</td>
</tr>
<tr>
<td>First order serial correl.</td>
<td>0.833</td>
<td>0.834</td>
</tr>
<tr>
<td>Sargan test</td>
<td>5535.88 (135)</td>
<td>4197.56 (97)</td>
</tr>
<tr>
<td>Prob. value</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

1. Heteroscedastic consistent standard errors in parentheses.
2. The Sargan tests are tests for the validity of instruments with degrees of freedom given in

To test for the joint significance of $L_{t+1}$, $L_t$, and $L_{t-1}$ we use a Wald test. Under the null hypothesis, the statistics, $W_t$, is asymptotically distributed as a $\chi^2$ and it has the following form
The values of the Wald test for the four models are given the following table.

Table XXXII Wald test of joint significance of the coefficients of \( L_{4,1}, L, L_{4,1} \).

<table>
<thead>
<tr>
<th>Type of model:</th>
<th>Estimation in:</th>
<th>LEVELS</th>
<th>DIFFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>(t-1)(t-2)(t-3)</td>
<td>(t-2)(t-3)</td>
<td>(t-2)(t-3)</td>
<td>(t-3)</td>
</tr>
<tr>
<td>( \chi^2(3) ):</td>
<td>330.78</td>
<td>201.78</td>
<td>15.05</td>
</tr>
<tr>
<td>prob. value</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

All the values of the test lead to the rejection of the null hypothesis that all coefficients of \( L_{4,1}, L, \) and \( L_{4,1} \) are equal to zero. The joint significance of the three regressors relating to employment also implies the rejection of the static model according to which employment is determined the marginal productivity relationship given in (67).
Appendix 3.1. Euler equation

The profit function at time $t$ and $t+1$ is respectively

$$
\Pi_t = AL_t^\alpha - \frac{\gamma}{2} (\Delta L_t)^2 - W_t L_t
$$

and

$$
\Pi_{t+1} = \beta \left[ A L_{t+1}^\alpha - \frac{\gamma}{2} (\Delta L_{t+1})^2 - W_{t+1} L_{t+1} \right]
$$

The derivatives with respect to $L_t$ are

$$
\frac{\partial \Pi_t}{\partial L_t} = \alpha \left( \frac{Q}{L} \right)_t - \gamma L_t + \gamma L_{t-1} - W_t = 0
$$

and

$$
\frac{\partial \Pi_{t+1}}{\partial L_t} = \beta \gamma L_{t+1} - \beta \gamma L_t
$$

The Euler first order conditions are

$$
\beta \gamma L_{t+1} - \gamma [1 + \beta] L_t + \gamma L_{t-1} + \alpha \left( \frac{Q}{L} \right)_t - W_t = 0
$$
Appendix 3.2. Selectivity bias: distributional assumptions on the error terms.

We have assumed that the random elements $v_s$ and $\epsilon_s$ are jointly normally distributed in the selection equation and in the regression equation. This allows us to obtain the selectivity term, the inverse Mills' ratio, as the ratio between the standard normal density and the standard normal distribution function. Nevertheless, Lee (1982) shows that this is a restrictive distributional assumption and suggests ways to specify selectivity models without the hypothesis of joint normality, adopting flexible function forms for the correction of the selectivity bias in the regression equation. This generalizes the formulae based on normal distributions, and in particular it takes into account non-symmetry and skewness in the distributions. Although we do not implement these procedures in the results reported in this thesis, this approach shows how, in principle, our model may be generalized to avoid restrictive distributional assumptions.

Following Lee's discussion, we may rewrite equation (46) in Section 3.4. in an alternative way as

$$v_{it} = \rho \sigma_v \frac{\epsilon_{it} - \mu_t}{\sigma} + \delta_{it}$$  \hspace{1cm} (A4)

where $\xi_s$ is the regression residual of $v_s$ on $\epsilon_s$ and $E(\xi_s | \delta'Z_x > \epsilon_s) = 0$. Now, in order to introduce functional forms different from normal we specify a strictly increasing transformation $J$ such that

$$\delta'Z_x > \epsilon_s \rightarrow J(\delta'Z_x) > J(\epsilon_s)$$  \hspace{1cm} (A5)

We may think of $v_{it}$ of a convolution of two independent random variables with one of them proportional to $J(\epsilon_s)$,
\[ v_{it} = \alpha (J(\epsilon_{it} - \mu_J) + v_{it} \tag{A6} \]

where \( v \) and \( J(\epsilon) \) are independent and \( \mu_J = \text{E}(J(\epsilon)) \). The disturbances \( \epsilon \) and \( v \) are correlated if \( \alpha \neq 0 \) and uncorrelated if \( \alpha = 0 \). Let \( \sigma_j^2 \) and \( \rho \) be the variance of \( J(\epsilon) \) and the correlation coefficient of \( v \) and \( J(\epsilon) \). Then we may rewrite equation (A4) as

\[ v_{it} = \rho \sigma_v \frac{J(\epsilon_{it}) - \mu_J}{\sigma_j} + v_{it} \tag{A7} \]

and our model becomes

\[ L_{it} = \psi_{1t} + \rho \sigma_v \frac{J(\epsilon_{it}) - \mu_J}{\sigma_j} + v_{it} \tag{A8} \]

\[ y_{it} = \delta Z_{it} + \epsilon_{it} \]

Let \( f_j \) be the known density function of \( J(\epsilon) \) which is assumed to exist under transformation \( J \) and set the random variable \( J(\epsilon) = \epsilon^* \). Denote the incomplete first moment of \( \epsilon^* \) evaluated at \( J(\delta Z) \) as

\[ \mu(J(\delta Z)) = \int_{\epsilon(\delta Z)} f_\epsilon(\epsilon^*) d\epsilon^* \tag{A9} \]

Let \( F(\delta Z) = \text{Pr}(\delta Z \geq \epsilon) \) be the probability that the company adjusts. The regression equation, conditional on the observed sample of firms which adjust, and thus corrected for the selectivity bias becomes

\[ L_{it} = \psi_{it} + \rho \sigma_v \frac{(\mu(J(\delta Z_{it}))}{\sigma_j} - \mu_j + \mu_{it} \tag{A10} \]

where
\[
\begin{align*}
\frac{(J(\delta'Z_u))}{F(\delta'Z_u)}
\end{align*}
\]

has zero conditional mean, \(E(u|X,Z,\text{adjust})=0\). The conditional variance of \(u\) is

\[
\text{var}(u_{it}|X_{it},Z_{it},\text{adjust}) = \frac{\rho^2\sigma_u^2}{\sigma_j^2} \left[ E(J(\epsilon_y)^2|\delta'Z_u) - E(J(\epsilon_y)|\delta'Z_u)\epsilon_y^2 \right] + \sigma_u^2(1-\rho^2)
\]

(A12)

where

\[
\mu_2(J(\delta'Z_u)) = \int_{\mathbb{R}^d} v_e^2 f_j(v_x)dv_x
\]

(A13)
is the incomplete second moment around zero of \(\epsilon^*\) and the transformation \(J\) have been completely specified and \(\mu_j\) and \(\sigma_j\) are known parameters and the remaining unknown parameters of the model are \(\iota_j\), \(\rho\), \(\delta\)'s and \(\sigma^*\). The non-linear equation (A10) may be estimated by two stage method, for example Amemiya (1974) and Heckman (1979). In the first stage we obtain the \(\delta\)'s by maximum likelihood estimation for the implied probability choice model given the distribution \(F(.)\) for the disturbance \(\epsilon_{it}\). In the second stage we estimate the modified regression equation by OLS using the non-censored observations. This approach provides a way to generate a large class of models with selectivity. By specifying different transformations \(J(.)\), we may allow different implicit distributions on \(v_{it}\) without
having to impose joint normality. In considering the relation of union membership and wages, Lee (1982) finds that the approach reveals that, if the distributions of the disturbances are incorrectly specified as normal, we may fail to detect the presence of selectivity bias in the regression equations.

---

1 The regression of \( v_h \) on \( J(\epsilon_h) \) needs not to be linear as assumed in the discussion above. Lee extends the approach also to a polynomial functional form.
Appendix 3.3. Inverse Mills’ ratio: summary statistics

Inverse Mills’ ratio calculated from Probit \((t-1)(t-2)(t-3)\)

Table A.3.1 \(\lambda_t\) (levels)

<table>
<thead>
<tr>
<th>year</th>
<th>no. of obs</th>
<th>mean</th>
<th>std. dev.</th>
<th>min.</th>
<th>max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1983</td>
<td>3247</td>
<td>0.330</td>
<td>0.121</td>
<td>0.0082</td>
<td>2.001</td>
</tr>
<tr>
<td>1984</td>
<td>3247</td>
<td>0.333</td>
<td>0.192</td>
<td>0.0008</td>
<td>4.726</td>
</tr>
<tr>
<td>1985</td>
<td>3247</td>
<td>0.344</td>
<td>0.139</td>
<td>0.0006</td>
<td>4.335</td>
</tr>
<tr>
<td>1986</td>
<td>3247</td>
<td>0.336</td>
<td>0.194</td>
<td>0.0001</td>
<td>6.077</td>
</tr>
<tr>
<td>1987</td>
<td>3247</td>
<td>0.319</td>
<td>0.104</td>
<td>0.0001</td>
<td>1.289</td>
</tr>
<tr>
<td>1988</td>
<td>3247</td>
<td>0.323</td>
<td>0.122</td>
<td>0.0009</td>
<td>3.231</td>
</tr>
<tr>
<td>1989</td>
<td>3247</td>
<td>0.319</td>
<td>0.164</td>
<td>0.0001</td>
<td>3.223</td>
</tr>
</tbody>
</table>

Table A.3.2 \(\lambda_t - \lambda_{t-1}\) (differences)

<table>
<thead>
<tr>
<th>year</th>
<th>no. of obs</th>
<th>mean</th>
<th>std. dev.</th>
<th>min.</th>
<th>max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1984</td>
<td>3247</td>
<td>0.011</td>
<td>0.290</td>
<td>-1.887</td>
<td>8.278</td>
</tr>
<tr>
<td>1985</td>
<td>3247</td>
<td>0.033</td>
<td>0.581</td>
<td>-14.127</td>
<td>11.843</td>
</tr>
<tr>
<td>1986</td>
<td>3247</td>
<td>-0.026</td>
<td>0.581</td>
<td>-11.104</td>
<td>18.434</td>
</tr>
<tr>
<td>1987</td>
<td>3247</td>
<td>0.051</td>
<td>0.527</td>
<td>-18.377</td>
<td>2.692</td>
</tr>
<tr>
<td>1988</td>
<td>3247</td>
<td>0.010</td>
<td>0.271</td>
<td>-2.315</td>
<td>8.398</td>
</tr>
<tr>
<td>1989</td>
<td>3247</td>
<td>-0.011</td>
<td>0.404</td>
<td>-4.789</td>
<td>8.339</td>
</tr>
</tbody>
</table>
Inverse Mills' ratio calculated from Probit (t-2)(t-3)

Table A.3.3 $\lambda_t$

<table>
<thead>
<tr>
<th>year</th>
<th>no. of obs.</th>
<th>mean</th>
<th>std. dev.</th>
<th>min.</th>
<th>max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1984</td>
<td>3247</td>
<td>0.333</td>
<td>0.144</td>
<td>0.0007</td>
<td>2.634</td>
</tr>
<tr>
<td>1985</td>
<td>3247</td>
<td>0.343</td>
<td>0.128</td>
<td>0.0003</td>
<td>3.447</td>
</tr>
<tr>
<td>1986</td>
<td>3247</td>
<td>0.336</td>
<td>0.158</td>
<td>0.0001</td>
<td>4.397</td>
</tr>
<tr>
<td>1987</td>
<td>3247</td>
<td>0.325</td>
<td>0.101</td>
<td>0.0001</td>
<td>1.237</td>
</tr>
<tr>
<td>1988</td>
<td>3247</td>
<td>0.328</td>
<td>0.126</td>
<td>0.0009</td>
<td>3.456</td>
</tr>
</tbody>
</table>

Table A.3.4 $\lambda_{t-1}$

<table>
<thead>
<tr>
<th>year</th>
<th>no. of obs.</th>
<th>mean</th>
<th>std. dev.</th>
<th>min.</th>
<th>max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1985</td>
<td>3247</td>
<td>0.031</td>
<td>0.381</td>
<td>-0.790</td>
<td>6.463</td>
</tr>
<tr>
<td>1986</td>
<td>3247</td>
<td>-0.023</td>
<td>0.415</td>
<td>-8.784</td>
<td>11.887</td>
</tr>
<tr>
<td>1987</td>
<td>3247</td>
<td>-0.033</td>
<td>0.410</td>
<td>-13.267</td>
<td>2.730</td>
</tr>
<tr>
<td>1988</td>
<td>3247</td>
<td>0.009</td>
<td>0.243</td>
<td>-1.910</td>
<td>6.710</td>
</tr>
<tr>
<td>1989</td>
<td>3247</td>
<td>-0.026</td>
<td>0.315</td>
<td>-3.046</td>
<td>3.746</td>
</tr>
</tbody>
</table>
Inverse Mills' ratio calculated from Probit (t-3)

Table A.3.5 $\lambda_i$

<table>
<thead>
<tr>
<th>year</th>
<th>no. of obs.</th>
<th>mean</th>
<th>std. dev.</th>
<th>min.</th>
<th>max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1985</td>
<td>3247</td>
<td>0.327</td>
<td>0.136</td>
<td>0.0001</td>
<td>2.429</td>
</tr>
<tr>
<td>1986</td>
<td>3247</td>
<td>0.322</td>
<td>0.117</td>
<td>0.0067</td>
<td>2.426</td>
</tr>
<tr>
<td>1987</td>
<td>3247</td>
<td>0.322</td>
<td>0.126</td>
<td>0.0082</td>
<td>2.429</td>
</tr>
<tr>
<td>1988</td>
<td>3247</td>
<td>0.317</td>
<td>0.102</td>
<td>0.0011</td>
<td>1.224</td>
</tr>
<tr>
<td>1989</td>
<td>3247</td>
<td>0.316</td>
<td>0.111</td>
<td>0.0003</td>
<td>2.426</td>
</tr>
</tbody>
</table>

Table A.3.6 $\lambda_t - \lambda_{t+1}$

<table>
<thead>
<tr>
<th>year</th>
<th>no. of obs.</th>
<th>mean</th>
<th>std. dev.</th>
<th>min.</th>
<th>max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1986</td>
<td>3247</td>
<td>-0.016</td>
<td>0.220</td>
<td>-5.9604</td>
<td>3.752</td>
</tr>
<tr>
<td>1987</td>
<td>3247</td>
<td>-0.0002</td>
<td>0.191</td>
<td>-3.750</td>
<td>6.176</td>
</tr>
<tr>
<td>1988</td>
<td>3247</td>
<td>-0.016</td>
<td>0.180</td>
<td>-6.260</td>
<td>0.538</td>
</tr>
<tr>
<td>1989</td>
<td>3247</td>
<td>-0.002</td>
<td>0.104</td>
<td>-0.930</td>
<td>3.746</td>
</tr>
</tbody>
</table>
Appendix 3.4. Plots of the probability of adjusting

1. Probability of adjusting after having adjusted upwards the previous period

Figure A.3.1.
2. Probability of adjusting after not having adjusted the previous period

Figure A.3.2.

\[ (t-1)C(t-2)C(t-3) + (t-2)C(t-3) \cdot (t-3) \]
3. Probability of adjusting after having adjusted downwards the previous year

All probabilities are calculated each year as average over companies from probits estimations as explained in the chapter.
Table A.3.7 Probability of adjusting after having adjusted upwards at t-1

<table>
<thead>
<tr>
<th>Year</th>
<th>No. of Observations</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1984</td>
<td>982</td>
<td>0.801</td>
<td>0.110</td>
<td>0.001</td>
<td>0.997</td>
</tr>
<tr>
<td>1985</td>
<td>1166</td>
<td>0.797</td>
<td>0.085</td>
<td>0.001</td>
<td>0.999</td>
</tr>
<tr>
<td>1986</td>
<td>1297</td>
<td>0.810</td>
<td>0.092</td>
<td>0.001</td>
<td>0.999</td>
</tr>
<tr>
<td>1987</td>
<td>1273</td>
<td>0.824</td>
<td>0.070</td>
<td>0.301</td>
<td>0.999</td>
</tr>
<tr>
<td>1988</td>
<td>1449</td>
<td>0.821</td>
<td>0.075</td>
<td>0.076</td>
<td>0.999</td>
</tr>
<tr>
<td>1989</td>
<td>1482</td>
<td>0.828</td>
<td>0.073</td>
<td>0.010</td>
<td>0.999</td>
</tr>
</tbody>
</table>

Probability from model (t-2)(t-3)

<table>
<thead>
<tr>
<th>Year</th>
<th>No. of Observations</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1984</td>
<td>982</td>
<td>0.802</td>
<td>0.095</td>
<td>0.011</td>
<td>0.995</td>
</tr>
<tr>
<td>1985</td>
<td>1166</td>
<td>0.800</td>
<td>0.074</td>
<td>0.001</td>
<td>0.999</td>
</tr>
<tr>
<td>1986</td>
<td>1297</td>
<td>0.811</td>
<td>0.008</td>
<td>0.001</td>
<td>0.999</td>
</tr>
<tr>
<td>1987</td>
<td>1273</td>
<td>0.822</td>
<td>0.063</td>
<td>0.447</td>
<td>0.999</td>
</tr>
<tr>
<td>1988</td>
<td>1449</td>
<td>0.820</td>
<td>0.069</td>
<td>0.267</td>
<td>0.999</td>
</tr>
<tr>
<td>1989</td>
<td>1482</td>
<td>0.829</td>
<td>0.068</td>
<td>0.059</td>
<td>0.999</td>
</tr>
</tbody>
</table>

Probability from model (t-3)

<table>
<thead>
<tr>
<th>Year</th>
<th>No. of Observations</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1985</td>
<td>1166</td>
<td>0.807</td>
<td>0.078</td>
<td>0.020</td>
<td>0.995</td>
</tr>
<tr>
<td>1986</td>
<td>1297</td>
<td>0.814</td>
<td>0.074</td>
<td>0.019</td>
<td>0.998</td>
</tr>
<tr>
<td>1987</td>
<td>1273</td>
<td>0.822</td>
<td>0.066</td>
<td>0.019</td>
<td>0.997</td>
</tr>
<tr>
<td>1988</td>
<td>1449</td>
<td>0.822</td>
<td>0.064</td>
<td>0.270</td>
<td>0.999</td>
</tr>
<tr>
<td>1989</td>
<td>1482</td>
<td>0.826</td>
<td>0.060</td>
<td>0.547</td>
<td>0.999</td>
</tr>
</tbody>
</table>
Table A.3.8 Probability of adjusting after not having adjusted at t-1

<table>
<thead>
<tr>
<th>Year</th>
<th>No. of obs.</th>
<th>Probability (t-1)(t-2)(t-3)</th>
<th>Probability (t-2)(t-3)</th>
<th>Probability (t-3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1984</td>
<td>911</td>
<td>0.794</td>
<td>0.076</td>
<td>0.013</td>
</tr>
<tr>
<td>1985</td>
<td>637</td>
<td>0.779</td>
<td>0.072</td>
<td>0.017</td>
</tr>
<tr>
<td>1986</td>
<td>713</td>
<td>0.789</td>
<td>0.069</td>
<td>0.101</td>
</tr>
<tr>
<td>1987</td>
<td>731</td>
<td>0.794</td>
<td>0.064</td>
<td>0.024</td>
</tr>
<tr>
<td>1988</td>
<td>658</td>
<td>0.791</td>
<td>0.072</td>
<td>0.002</td>
</tr>
<tr>
<td>1989</td>
<td>646</td>
<td>0.792</td>
<td>0.099</td>
<td>0.005</td>
</tr>
</tbody>
</table>
Table A.3.9 Probability of adjusting after having adjusted downwards at t-1

<table>
<thead>
<tr>
<th>Year</th>
<th>No. of obs.</th>
<th>Mean</th>
<th>Std. dev.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1984</td>
<td>1354</td>
<td>0.833</td>
<td>0.055</td>
<td>0.606</td>
<td>0.999</td>
</tr>
<tr>
<td>1985</td>
<td>1444</td>
<td>0.820</td>
<td>0.060</td>
<td>0.073</td>
<td>0.999</td>
</tr>
<tr>
<td>1986</td>
<td>1237</td>
<td>0.824</td>
<td>0.058</td>
<td>0.264</td>
<td>0.999</td>
</tr>
<tr>
<td>1987</td>
<td>1243</td>
<td>0.828</td>
<td>0.059</td>
<td>0.425</td>
<td>0.999</td>
</tr>
<tr>
<td>1988</td>
<td>1140</td>
<td>0.827</td>
<td>0.060</td>
<td>0.037</td>
<td>0.999</td>
</tr>
<tr>
<td>1989</td>
<td>1119</td>
<td>0.829</td>
<td>0.058</td>
<td>0.504</td>
<td>0.999</td>
</tr>
</tbody>
</table>

Probability from model (t-2)(t-3)

<table>
<thead>
<tr>
<th>Year</th>
<th>No. of obs.</th>
<th>Mean</th>
<th>Std. dev.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1984</td>
<td>1354</td>
<td>0.831</td>
<td>0.060</td>
<td>0.234</td>
<td>0.999</td>
</tr>
<tr>
<td>1985</td>
<td>1444</td>
<td>0.818</td>
<td>0.065</td>
<td>0.043</td>
<td>0.999</td>
</tr>
<tr>
<td>1986</td>
<td>1237</td>
<td>0.822</td>
<td>0.060</td>
<td>0.201</td>
<td>0.999</td>
</tr>
<tr>
<td>1987</td>
<td>1243</td>
<td>0.823</td>
<td>0.064</td>
<td>0.265</td>
<td>0.999</td>
</tr>
<tr>
<td>1988</td>
<td>1140</td>
<td>0.821</td>
<td>0.073</td>
<td>0.079</td>
<td>0.998</td>
</tr>
<tr>
<td>1989</td>
<td>1119</td>
<td>0.827</td>
<td>0.062</td>
<td>0.494</td>
<td>0.999</td>
</tr>
</tbody>
</table>

Probability from model (t-3)

<table>
<thead>
<tr>
<th>Year</th>
<th>No. of obs.</th>
<th>Mean</th>
<th>Std. dev.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1985</td>
<td>1444</td>
<td>0.835</td>
<td>0.060</td>
<td>0.270</td>
<td>0.999</td>
</tr>
<tr>
<td>1986</td>
<td>1237</td>
<td>0.835</td>
<td>0.058</td>
<td>0.546</td>
<td>0.995</td>
</tr>
<tr>
<td>1987</td>
<td>1243</td>
<td>0.831</td>
<td>0.065</td>
<td>0.271</td>
<td>0.996</td>
</tr>
<tr>
<td>1988</td>
<td>1140</td>
<td>0.833</td>
<td>0.064</td>
<td>0.271</td>
<td>0.995</td>
</tr>
<tr>
<td>1989</td>
<td>1119</td>
<td>0.831</td>
<td>0.062</td>
<td>0.276</td>
<td>0.993</td>
</tr>
</tbody>
</table>
Appendix 3.5. Gauss program for GMM estimation

The program is written in Gauss. The version presented refers to the estimation of model \((t-2)(t-3)\) in differences.

PROGRAM

```
clear u1,u2,u3,v,u4,v5,zuuziv1,zuuzgmm;
emode = 1; @ 0 for levels, 1 for differences @
estage = 3; @ 1 for single stage estimator, 2 for 2 stage, 3 for stage @
if emode == 0;
estage = minc(estage|2); @ 3 stage not necessary for levels @
endif;
output file = results.out reset; @ output written to this file @
output on;
nobs = 8319; @ known number of observations, else set to zero @
im = 13; @ position of Mills’ ratio @
ip1 = 3; @ initial observation actually used @
ivy = 2; @ first lag used in instrumenting @
nx = 6; @ # regressors including constant @
nx = nx-emode; @ reduce if estimation in differences @
np = 8; @ years per firm @
c = rows(w)/np; @ # firms @
ie = 6; @ position of employment variable @
iw = 3; @ position of wages variable @
iq = 9; @ position of productivity variable @
ib = 4; @ position of wage-bill variable @
iop = 5; @ position of output variable @
inif = 7; @ position of nif variable @
imol = 8; @ position of mol variable @
isiz = 11; @ position of size variable @
nid = 6; @ # instrument distributed lags @
nio = 2; @ # other instruments @
ip1 = ip1+emode; @ increase if estimation in differences @
ip2 = 7; @ final observation actually used @
iv1 = iv1+emode; @ increase if estimation in differences @
nit =((ip2-iv1)*(ip2-iv1+1)/2-(ip1-iv1-1)*(ip1-iv1)/2)*nid
+ (ip2-ip1+1)*nio; @ total number of instruments @
if emode == 0;
let dnam = empl; @ dependent variable name @
let vnam = constant empl+1 empl-1 wages prdty invmil;
let "regressor variable names @
else;
let dnam = dempl; @ dependent variable name @
```
let vnam = dempl+1 dempl-1 dwages dprdty dinvmil;
   @ regressor variable names @
endif;
print "Number of observations used "; nob;
print "Position of inverse Mills’ ratio "; im;
print "Initial observation used "; ip1;
print "Initial lag used in instrumenting "; iv1;
output off;

   @ for Hausman test save b and varb after execution @
   @ computation starts @

   ipq = ip2-ip1+1;  @ max number of observations usable @
   zx = zeros(nit,nx);  @ crossproduct matrix @
   zy = zeros(nit,1);  @ ditto @
   zz = zeros(nit,nit);  @ instrument cross-product matrix @
   hh = -.5;  @ 1st order autocorrelation coefficient, 1st stage @
   istage = 0;  @ Current estimation stage @
   a2:  @ jump back for second or third stage @
   istage = istage+1;  @ increment stage indicator @
if emode==1 and istage<3;
   hv = 1|hh;
if ip2-ip1-1>0;
   hv = hv|zeros(ip2-ip1-1,1);
endif;
   h = toeplitz(hv);  @ H matrix for difference estimator @
   zhz = zeros(nit,nit);  @ GMM cross-product matrix @
endif;
if istage==1;
   if nob==0;
      y = zeros(1,1);  @ dependent variable @
      x = zeros(1,nx);  @ regressor matrix @
      v = zeros(1,1);  @ vector to indicate presence of lags @
   else;
      y = zeros(nobs,1);
      x = zeros(nobs,nx);
      v = zeros(nobs,1);
   endif;
elseif istage==2;
   zuuziv = zeros(nit,nit);  @ cross product matrix for iv residuals and instrs @
else;
   clear zuuziv;
   zuuzgmm = zeros(nit,nit);  @ cross product matrix for gmm residuals and instrs @
   zu = zeros(nit,1);  @ Z'u @
endif;
li = 0;  @ indicator for lags @
ic = 1;
ixy = 0;  @ observation counter @
do until ic == nc+1;  @ loop over companies @
  ip = ip1;
  zq = zeros(ipq,nit);  @ instrument matrix @
  if emode == 1;
    hq = h;  @ H matrix for GMM @
  endif;
  ivq = 0;  @ initialize instrument position counters @
do until ip == ip2+1;  @ loop over years @
    istage; ic; ip; ixy;  @ print position @
    i = (ic-1)*np+ip;  @ observation number @
    @ create entire set of instuments @
    if emode == 0;
      zq[ip-ip1+1, ivq+1:ivq+nio] = w[i,im];
    else;
      zq[ip-ip1+1, ivq+1:ivq+nio] = (w[i,im]-w[i-1,im]);
    endif;
    ivq = ivq+nio;
  iv = iv1;
  do until liv = = ip;
    zq[ip-ip1+1, ivq+1:ivq+nid] = w[i,ie] ~ w[i,ib] ~ w[i,iop] ~ w[i,isiz];  @ create instruments @
    ivq = ivq+nid;  @ update instrument counter @
    iv = iv+1;
  endo;
  if w[i,ie] == w[i-1,ie] or (emode == 1 and w[i-1,ie] == w[i-2,ie]);
    if emode == 1;
      hq[,ip-ip1+1] = zeros(ipq,1);  @ modify H matrix @
      hq[ip-ip1+1,] = zeros(1,ipq);  @ for missing values @
    endif;
    li = 0;  @ reset lag indicator @
    goto a1;  @ discard observation if employment unchanged @
  endif;
  ixy = ixy+1;  @ augment observation counter @
  if istage == 1;
    if nobs == 0;
      if emode == 0;
        y = y|w[i,ie];  @ augment dependent variable vector @
        x = x|(1 ~ w[i+1,ie] ~ w[i-1,ie] ~ w[i,iw] ~ w[i,iq] ~ w[i,im]);  @ augment regressor variable matrix @
      else;  @ data for estimation in differences @
        y = y|(w[i,ie]-w[i-1,ie]);
        x = x|((w[i+1,ie]-w[i,ie]) ~ (w[i-1,ie]-w[i-2,ie]) ~ (w[i,iw]-w[i-1,iw]) ~ (w[i,iq]-w[i-1,iq]) ~ (w[i,im]-w[i-1,im]));
      endif;
      v = v|li;  @ augment vector indicating if lag present @
    else;
ifemode == 0;
y[ixy] = w[i,ie];  @ augment dependent variable vector @
x[ixy,.] = (1 ~ w[i+1,ie] ~ w[i-1,ie] ~ w[i,iw] ~ w[i,iq] ~ w[i,im]);  @ augment regressor variable matrix @
else;  @ data for estimation in differences @
y[ixy] = w[i,ie]-w[i-1,ie];
x[ixy,.] = (w[i+1,ie]-w[i,ie])~(w[i-1,ie]-w[i-2,ie])
~(w[i,iw]-w[i-l,iw]) ~ (w[i,iq]-w[i-l,iq])
~(w[i,im]-w[i-l,im]);
endif;
v[ixy] = li;  @ augment vector indicating if lag present @
endif;
endif;
if ixy*istage == l and nobs == 0;
y = trimr(y,l,0);  @ remove leading dummy observation @
x = trimr(x,l,0);  @ ditto @
v = trimr(v,l,0);  @ ditto @
endif;
zp = zq[ip-ip1+1,];  @ row vector for instruments @
if istage == l;
  zz = zz + zp'*zp;  @ increment crossproduct matrix @
  zx = zx + zp*x[ixy,];
  zy = zy + zp*y[ixy];
  if ip<ip2;
    li = 1;  @ reset lag indicator @
  else;
    li = 0;
  endif;
else;
  zuuziv = zuuziv + (uiv[ixy] ^ 2).* (zp'*zp);
  @ White cross-product matrix for robust s.e.s @
else;
  if li==1;
    zp1 = zq[ip-ip1,];  @ lagged instruments @
  else;
    zp1 = zeros(l,cols(zq));
  endif;
  zuuzgmm = zuuzgmm + (u[ixy] ^ 2).* (zp'*zp)
  + (u[ixy]*u1[ixy]*v[ixy]).*(zp'*zp1+zp1'*zp);
  @White cross-product matrix for robust s.e.s @
  zu = zu + u[ixy]*zp';  @ cross-product of instruments & GMM resids @
endif;
a1:
ip = ip+1;
endo;  @ close year loop @
ifemode == 1;
if istage<3;
  zhz = zhz + zq'*hq*zq;  @ increment crossproduct matrix @

211
endif;
edif;
ic = ic+1;
endo;
if nobs==0;
nobs = rows(y);
endif;
output on;
if istage==2;
print; print; print;
print "Second stage estimates";
print;
elseif istage==3;
print; print; print;
print "Third stage estimates";
print;
endif;

@ close company loop @

@ OLS section @
df = nobs-nx;
if istage==1;
xx = x'*x;
xy = x'*y;
xxinv = invpd(xx);
b = xxinv*xy;
mny = meanc(y);
sdy = stdc(y);
tss = y'*y;

@ cross-product matrix @
@ ditto @
@ (X'X)-1 @
@ coefficient vector @
@ dependent variable mean @
@ dependent variable s.d. @
@ total sum of squares @
@ explained sum of squares @
@ residual sum of squares @
@ equation standard error @
@ coefficient standard errors @
@ sum of squares about mean @
@ R2 @

fstat = (rss/(1-rsq))*(df/(nx-1));

pvf = cdffc(fstat,nx-l,df);

u = y-x*b;
u1 = trimr(0|u,0,1);
rho = u*(u1.*v)/(u*(u.*v));

wmat = zeros(cols(x),cols(x));
i=1;
do until i==nobs+1;
    wmat = wmat + (u[i]^2).*x[i,]*x[i,];
    i=i+1;
endo;

varb = xxinv*wmat*xxinv;
srob = sqrt(diag(varb));
ireg = 1;
call lsout;  @ print results @
endif;
@ IV section @

if istage<3;

invzz = invpd(zz);  @ (Z'Z)-1 @
xxzx = zx'*invzz*zx;  @ X'Z(Z'Z)-1Z'X @
invvxxz = invpd(xzzx);  @ (X'Z(Z'Z)-1Z'X)-1 @
xxzy = zx'*invzz*zy;  @ X'Z(Z'Z)-1Z'Y @
b = invvvxzx*xxzy;  @ IV coefficient estimates @
u = y - x*b;  @ residuals @

rss = u'*u;
u1 = trimr(0|u,0,1);  @ lagged residuals @
rho = u'*(u1.*v)/(u'*(u.*v));  @ autocorrelation @
se = sqrt(rss/df);  @ equation standard error @

if istage==1;
s = sqrt(diag(invxzzx))*se;  @ coefficient standard errors @
else;

wmat = zx'*invzz*zuuziv*invzz*zx;  @ White matrix for robust s.e.s @

varb = invvvxzx*wmat*invvvxzx;  @ variance matrix of estimates @
s = sqrt(diag(varb));  @ robust standard errors @
endif;
sargan = zy'*invzz*zy + b'*xxzx*b - 2*b'*xxzy;
@ Sargan test statistic @
sargan =  sargan*df/rss;
sdf = cols(zz) - cols(xzzx);  @ degrees of freedom @
pvsg = cdfchic(sargan,sdf);  @ probability value @
ireg = ireg+1;  @ regression reference number @

call lsout;  @ print results @
if estage>1 and istage==1;
uiv = u;  @ save residuals for second stage @
endif;
endif;

@ GMM section @

if emode==1;

invzz = invpd(zhz);  @ (Z'HZ)-1 @
xxzx = zx'*invzz*zx;  @ X'Z(Z'Z)-1Z'X @
invvxxz = invpd(xzzx);  @ (X'Z(Z'Z)-1Z'X)-1 @
xxzy = zx'*invzz*zy;  @ X'Z(Z'Z)-1Z'Y @
b = invvvxzx*xxzy;  @ GMM coefficient estimates @
u = y - x*b;  @ residuals @

rss = u'*u;
u1 = trimr(0|u,0,1);  @ lagged residuals @
rho = u'*(u1.*v)/(u'*(u.*v));  @ autocorrelation @
se = sqrt(rss/df);  @ equation standard error @

if istage<3;
s = sqrt(diag(invvxzx))*se;  @ coefficient standard errors @
else;

213
wmat = zx' * invzz * zuuzgmm * invzz * zx;
     @ White matrix for robust s.e.s @
varb = invxzzx * wmat * invxzzx;
s = sqrt(diag(varb)); @ robust standard errors @
zuuzinv = invpd(zuuzgmm); @ inverse matrix for Sargan test @
sargan = zu' * zuuzinv * zu; @ Saragan test statistic @
pvsg = cdfchic(sargan, sdf); @ probability value @
endif;
ireg = ireg + 1; @ regression reference number @
call lsout; @ print results @
endif;
if estage == 3 and istage == 1;
    hh = rho; @ revise serial correlation parameter @
    if rho < -.5;
        print;
        print "Warning: estimated autocorrelation over-ridden";
        hh = -.5; @ -.5 is minimum value allowed @
    endif;
endif;
if emode == 0;
    ireg = ireg + 1; @ advance regression counter to omit GMM @
endif;
if istage < estage;
    output off;
goto a2;
endif;
if emode == 0;
    varb = varb[2:rows(b), 2:rows(b)];
    b = trimr(b, 1, 0); @ eliminate terms relating to intercept @
    @ for Hausman test @
endif;
end;
@ estimation output routine @
@ proc(O) = lsout;
local fmt jr, mask, njr, omat, pvt, str, t;
mask = 0 ~ 1 ~ 1 ~ 1 ~ 1;
let fmt[5, 3] = "*.*s" 9 8 "*.*lf' 12 6 "*.*lf' 12 6 "*.*lf' 12 6 "*.*lf' 12 6 "*.*lf' 10 3; 
print ftos(ireg, "Regression: %*.*lf', 20, 0);
if ireg == 1;
    print " OLS";
elseif ireg == 2 or ireg == 4;
    print " IV";
else;
    print ftos(hh, " GMM - using rho = %*.*lf', 20, 5);
print ftos(nobs,"Observations: %*.*lf",20,0);
print ftos(dnam," Dependent variable:%*.*s",20,8);
print ftos(mny, "Mean: %*.*lf",20,5);
print ftos(sdy, " Standard deviation:%*.*lf",20,5);
print ftos(tv, "Total SS: %*.*lf",20,5);
print ftos(df, " Degrees of freedom:%*.*lf",20,0);
if ireg==1;
    print ftos(rsq,"R-squared: %*.*lf",20,5);
endif;
print ftos(rss,"Residual SS: %*.*lf",20,5);
print ftos(se," Std error of est: %*.*lf",20,5);
if ireg==1;
    str = ftos(nx,"F(%*.*lf,,1,0) $+ ftos(df,"%*.*lf": ",1,0);
    str = strsect(str,1,15) $+ ftos(fstat,"%*.*lf",19,5);
    print str;
    print ftos(pvf," Probability of F: %*.*lf",20,5);
else if ireg==2 or ireg==6;
    print;
    print ftos(sdf,"Sargan test of over-identifying restrictions chi2(%*.*lf",5,0);
    print ftos(sargan,") = %*.*lf",10,5);
    print ftos(pvsg," probability value: %*.*lf",10,5);
endif;
print ftos(rho,"First order residual autocorrelation coefficient: %*.*lf" ,28,5);
jr = 1;
if ireg==1;
    njr = 3;
else;
    njr = 2;
endif;
do until jr==njr;
    if jr==2;
        s = srob; @ select robust standard errors @
    endif;
    if jr==2 or ireg==4 or ireg==6;
        print;
        print "Robust standard errors";
    endif;
    t = b./s; @ t values @
pvt = 2*cdftc(abs(t),df); @ p values for t statistics @
print;
print " Variable Estimate Error t-value > |t|";
print "-----------------------------------------------";
omat = vnam ~ b ~ s ~ t ~ pvt;
call printfm(omat,mask,fmt);
jr = jr+1;
endo;
print;print;print;
retp;
endp;
Appendix 3.6. Estimation of the Euler equation with time dummies

Table A.3.10. Euler equation with time dummies

<table>
<thead>
<tr>
<th>Dependent variable: $L_t$</th>
<th>Estimation in:</th>
<th>LEVELS</th>
<th>DIFFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timing of the instruments:</td>
<td>(t-1)(t-2)(t-3)</td>
<td>(t-2)(t-3)</td>
<td>(t-2)(t-3)</td>
</tr>
<tr>
<td>const.</td>
<td>-0.040</td>
<td>-0.017</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.037)</td>
<td>(0.028)</td>
</tr>
<tr>
<td>$L_{t+1}$</td>
<td>0.556</td>
<td>0.522</td>
<td>0.495</td>
</tr>
<tr>
<td></td>
<td>(0.045)</td>
<td>(0.047)</td>
<td>(0.179)</td>
</tr>
<tr>
<td>$L_{t+1}$</td>
<td>0.440</td>
<td>0.474</td>
<td>0.440</td>
</tr>
<tr>
<td></td>
<td>(0.045)</td>
<td>(0.046)</td>
<td>(0.139)</td>
</tr>
<tr>
<td>$(Q/L)_t$</td>
<td>-0.001</td>
<td>-0.001</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.033)</td>
</tr>
<tr>
<td>$W_t$</td>
<td>0.047</td>
<td>0.016</td>
<td>-0.382</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.033)</td>
<td>(0.494)</td>
</tr>
<tr>
<td>$\lambda_t$</td>
<td>0.003</td>
<td>-0.001</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.011)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>$D84$</td>
<td>-0.006</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>$D85$</td>
<td>-0.009</td>
<td>0.004</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>$D86$</td>
<td>-0.013</td>
<td>0.002</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.004)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>$D87$</td>
<td>-0.007</td>
<td>0.001</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.005)</td>
<td>(0.050)</td>
</tr>
<tr>
<td>$D88$</td>
<td>-0.008</td>
<td>0.010</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.006)</td>
<td>(0.065)</td>
</tr>
</tbody>
</table>
Table A.3.10 Euler equation with time dummies, cont.

<table>
<thead>
<tr>
<th>Estimation in:</th>
<th>LEVELS</th>
<th>DIFFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timing of the instruments:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(t-1)(t-2)(t-3)</td>
<td>(t-2)(t-3)</td>
<td>(t-2)(t-3)</td>
</tr>
<tr>
<td>no. obs.</td>
<td>15,186</td>
<td>12,850</td>
</tr>
<tr>
<td>std. error of est.</td>
<td>0.150</td>
<td>0.145</td>
</tr>
<tr>
<td>Sargan test</td>
<td>622.41</td>
<td>381.57</td>
</tr>
<tr>
<td>Probab. value</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

1. Heteroscedastic consistent standard errors in parentheses.
2. The Sargan tests are tests for the validity of instruments with the degrees of freedom given in brackets.
Chapter 4. FIXED ADJUSTMENT COSTS AND DYNAMICS

Introduction

4.1. Hamermesh's (1989) model

4.2. Discrete dynamic process: introduction
Introduction

In the previous chapter we assumed strictly convex adjustment costs, and considered the existence of the peak of zero net changes in employment, in the data, as a selection problem which needs to be taken into account in the estimation of the Euler equation. In other words, we maintained the hypothesis of quadratic adjustment costs and considered the proportion of zero variations in employment as an observability problem. For certain firms we were not able to see adjustment but we implicitly assumed that if we had been able to see changes in employment, these would have been as predicted by the Euler equation. We suggested that the same selectivity problem could have had two different origins. The first was through measurement errors, for example errors in recording the actual employment changes, which imply that the zeros may be incorrect. The second source was the existence of fixed costs which create an adjustment path which is characterized by discrete jumps, in contrast to the continuous and smooth path as implied by quadratic costs. Whatever the source of the peaks at zero in the rate of change in employment, the implied censoring in the data needed to be taken into account when estimating the Euler equation.

In this chapter we start to consider explicitly the possibility of fixed costs of adjustment. We discussed in Chapter 1. how legal constraints imposed on the companies' employment decisions imply costs that are decreasing rather than increasing at the margin. Therefore, adjustment costs may be better approximated by a fixed rather than a quadratic structure and, in that case, they will generate
large and infrequent changes of employment in the face of shocks.

In this chapter and those following we model firms’ optimal employment decisions assuming the existence of fixed costs. This chapter is mainly an introduction to the problems which arise when we attempt to model intertemporal decisions in the presence of fixed costs. This requires the use of dynamic programming within the framework of discrete decision processes. We emphasize the intuition behind the methods which we will present at a more formal level in the next chapter. Dynamic programming is a very general tool for dynamic optimization, and is particularly useful in treating uncertainty. It breaks a whole sequence of decisions into just two components: the immediate decision, and a valuation function that encapsulates the consequences of all subsequent decisions, starting from the position that results from the immediate choice. One works backwards from the final decision all the way to the initial state. Discrete decision processes allow us to take into account discontinuities, jumps and periods of inaction in the sequence of choices characterizing firm’s optimal behaviour in the presence of fixed adjustment costs. In fact, the infrequent adjustments have the implication that the Euler-GMM approach is no longer appropriate as we will illustrate in the following pages.

We start considering fixed costs by discussing, in the following section, the seminal work by Hamermesh (1989) where the problem of fixed costs in dynamic labour demand was treated for the first time in a formal way and analyzed empirically at plant level. In the subsequent sections we introduce firms’ optimal employment decisions as a result of a sequential decision process over firm’s time horizon under uncertainty, in the context of a discrete decision process model. The
complete formal treatment of the model of dynamic labour demand with fixed adjustment costs is given in Chapter 6.

4.1. Hamermesh's (1989) model

When adjustment costs are independent of the size of employment changes, the first order conditions given by the Euler equation break down. An example of fixed costs is given by the costs of hiring: the costs of maintaining a hiring office, advertising, filling out forms, etc. imply that after hiring one worker the marginal cost of attracting additional workers may be zero. The optimizing response in the presence of fixed costs is quite intuitive: if firms incur the same costs whenever they alter their labour demand, it does not pay to spread the adjustment over many periods as in the case of quadratic adjustment costs, since, if it were to do this, the company would incur fixed costs in each period. Thus, it is optimal for firms to complete the adjustment quickly rather than slowly, concentrating it in a single period. Because of this instantaneous jump to the desired level of employment, and the discontinuity in the demand for labour implied by this cost structure, the problem of optimal decision-making may be incorrectly seen as static rather than dynamic. To illustrate this result we follow Hamermesh (1989) and assume only fixed costs. Adjustment costs, C(·), may described as

$$C(L_{t-1}, L_t) = \begin{cases} 0 & \text{if } L_{t-1} = L_t \\ K & \text{if } L_{t-1} \neq L_t \end{cases}$$

(69)

where the fixed costs, K, are incurred in each period that the firm changes its labour input, otherwise they are zero. It is 'K' that plays the critical role in the
firm's employment decisions by discouraging over too frequent adjustments. Obviously expectations are also crucial in determining the adjustment. In particular, it is important how employers forecast the path of shocks that alter the desired level of employment.

It is useful to write (69) as

\[ C(L_{t-1}, L_t) = 1(L_{t-1} * L_t) K \]  

(70)

using an indicator function defined as

\[ 1(L_{t-1} * L_t) = \begin{cases} 0 & \text{if } L_{t-1} = L_t \\ 1 & \text{if } L_{t-1} * L_t \end{cases} \]  

(71)

It is clear that condition (71) implies a discontinuity arising from the decision of the firm of either to adjust, if it thinks it is worth changing employment and paying $K$, or not to adjust, if the deviations from desired employment do not justify the cost of varying labour. Also, the magnitude of the shock which alters the desired level of employment, and the perception by the employers of whether the shock is transitory or permanent, will affect the choice between adjusting and not adjusting. The perception of the magnitude and the duration of the exogenous shock is very important in this fixed costs context. In the case of quadratic adjustment costs it always pays the employer to change the labour input at least by a small amount, as the desired level is altered, even when the shock is brief. By contrast, in the presence of fixed costs, the optimal response to a shock may be do nothing if the shock is small and brief, and if the cost of altering labour is not outweighed by the increase in expected profits from a change in employment. As a result, given fixed costs and the forecasted shock path, the firm's optimal employment decision will be either to choose not to change employment, leaving $L_t = L_{t-1}$, or to jump
instantaneously to the desired level of employment, attaining $L_t^*$. 

In order to represent profits we adopt the same function as assumed in the discussion on the Euler equation

$$ \Pi_t = AL_t^\varepsilon - 1(L_{t-1} * L_t)K - W_tL_t $$

(72)

except that $C(\Delta L), \Delta L_t = L_t - L_{t-1}$, is described by (71). The employer again maximizes the discounted stream of profits as in Section 3.1. If we condition on the firm deciding to adjust, we may derive the first order conditions and compare them with the Euler equation in Section 3.1.

$$ \frac{\partial \Pi_t}{\partial L_t} = \frac{\partial Q_t}{\partial L_t} - W_t = 0 $$

(73)

Equation (73) looks like a static equilibrium condition given by the equality between the wage and the marginal value product of labour. The dynamics appear to vanish because, in contrast to the case of quadratic costs, the term representing adjustment costs drops from the first order conditions. By omitting the quadratic term due to convex adjustment costs in the first order conditions for profit maximization, we seem to lose the dynamics implied by the presence of adjustment costs. In Hamermesh's model, the company either maintains employment at $L_{v,t}$ or sets its time horizon, $T$, to zero and jumps immediately to the desired, long-run equilibrium value of labour demand, $L_t^*$, according to the rule

$$ K < \frac{[\Pi(L_t^*) - \Pi(L_{t-1})]}{\beta} $$

(74)

where $\beta$ is the firm's discount rate. This result is very similar to the one obtained in (s,s) inventory policies where the cost of restocking is independent of the size of the order. Every time inventories fall below some minimum level, s, a purchase
restores them to their maximum level \( S \). The decision rule (74) shows that the company will adjust according to whether the difference between the pay-offs relative to the two alternative actions is greater than the fixed costs, otherwise it will not alter the existing stock of workers. In particular, under fixed adjustment costs, the firm will only make those changes in the labour input which are justified by large departures of \( L^*_t \) (relative to \( K \)) from the most recent choice of the number of employees. Small fluctuations in desired employment leave current labour demand unchanged, while large ones lead to instantaneous reactions which complete the adjustment. This formulation does not account completely for the dynamics underlying fixed costs. The description of the dynamics is not exhausted by the instantaneous adjustment or by the firm's inaction. Even if companies only face fixed costs, the decision of whether or not to change the number of employees in the current period affects choices in the future. If the company adjusts too much today it may have to suffer from this in future periods, in the face of exogenous shocks.

Hamermesh, thus, describe labour demand as

\[
\begin{align*}
L_t &= L_{t-1} + \mu_{1t} \quad \text{if} \quad |L_{t-1} - L^*_t| \leq k \\
L_t &= L^*_t + \mu_{2t} \quad \text{if} \quad |L_{t-1} - L^*_t| > k
\end{align*}
\]  

(75)

where \( \mu_{1t} \) and \( \mu_{2t} \) are disturbances with \( \text{E}(\mu_{1t}, \mu_{2t}) = 0 \). The source of disturbances is not clearly specified in his article; disturbances seem to stem from errors to meet the target levels of employment. "k" is the percentage deviation of last period's employment from desired employment that is necessary to overcome fixed adjustment costs. It is an increasing function of the fixed costs of adjustment.

To estimate (75) it is crucial to specify \( L^*_t \). The long term equilibrium
employment, as perceived by the firm, is largely dominated by how employers form their expectations. This is, in fact, the way expectations enter Hamermesh's analysis. He specifies $L_t^*$ as:

$$L_t^* = aX_t + \epsilon_t$$  \hspace{1cm} (76)

where "a" is a vector of parameters linking the forcing variables $X_t$ to the desired stock of employment, $X_t$ is a vector of variables that affect $L_t^*$ such as expected output, change in expected output and time trend and $\epsilon_t$ is a forecasting error such that

$$E(\mu_{1\rho}\epsilon) = E(\mu_{2\rho}\epsilon) = 0$$ \hspace{1cm} (77)

The firm will operate on (75a) if

$$\epsilon_t \leq K + [L_{t-1} - aX_t]$$

and

$$\epsilon_t \geq -K + [L_{t-1} - aX_t]$$ \hspace{1cm} (78)

and on (75b) if

$$\epsilon_t > K + [L_{t-1} - aX_t]$$ \hspace{1cm} or

$$\epsilon_t < -K + [L_{t-1} - aX_t]$$ \hspace{1cm} (79)

In this second case, the firm will jump to its new long run equilibrium level of employment if it is sufficiently shocked by changes in $X_t$ or if forecasting errors overstate $|L_{t-1} - L_t^*|$.

The fact that the dynamics inherent in the fixed costs model and the expectations formation process are not completely dealt with in the model creates a gap between the theoretical and empirical models. Dynamics enter in the empirical model through a switching regression model represented essentially by
(75a) and (75b). The model specifies the probability that the firm will reach the
equilibrium level of employment $L_1^*$, and the complementary probability it will
optimally be immobile. However, $L_1^*$ is not a target towards which the firm adjusts
employment but it is part of the optimization problem. It is affected by the choice
of $L_t$ and it cannot be determined independently of it. Since the choice the
employer selects at time $t$ affects all the future actions, he/she must take into
account all the consequences in the future, in an environment dominated by output
and factor price uncertainty. Hence, optimal employment behaviour is governed by
a sequence of decisions - whether or not to adjust - and states - the level of
employment inherited by the firm at the beginning of each period - in the presence
of output price and wages shocks. The incomplete nature of the analysis of the
dynamics of fixed costs is also the result of the inappropriate use of a continuous
decision process which characterizes the Euler equation. To obtain the Euler
equation we need, among other conditions, that the optimal decision be an interior
point of the choice set, but, as is clear from this discussion, in the presence of fixed
costs, companies may be at corners with regard to their employment decisions. This
creates the impression that the intertemporal content of the fixed costs model is
exhausted in a one shot action as opposite to the case of quadratic adjustment
costs.

In the next Section we start our analysis of the considering firm’s optimal
employment decisions as a result of a sequential decision process over the firm’s
time horizon under uncertainty, in the more appropriate context of a discrete
decision process.
4.2. Discrete dynamic process

In the previous discussion, we noticed how the introduction of fixed costs, the only form of costs we assume in the present section, requires a more elaborate framework within which to model firms' optimal employment choices. First of all, the existence of fixed costs makes the use of a continuous decision process, where the decision variable may assume a continuum of values, inappropriate despite the tractability of the first order conditions. As already mentioned in the previous section, in the presence of fixed costs, companies may be at corners with regard their employment decisions. Labour demand is, then, governed by threshold rules, implying an optimal policy which consists of doing nothing when the stock of workers is larger (smaller) than a critical value and to hire (fire) up to the desired level if the stock is below (above) the threshold. As we have shown, this optimal decision rule makes the Euler-GMM approach any longer appropriate, given the non-differentiability of the objective function at the points corresponding to the threshold values. In fact, again as explained in the previous section, if we maintain the Euler equation framework, the presence of fixed costs tends to hide the dynamics of decision making and give the impression that the problem reduces to a static optimization where the firm selects an isolated decision, setting \( t=T \), rather than a sequence of decision rules, \( \delta=(\delta_{t_1},\ldots,\delta_{t_T}) \); each \( \delta \), specifying the firm's best choice as a function of its information at time \( t \) and state variables. It is an "action versus strategy" distinction. The problem becomes one of complex multiperiod decision making and requires dynamic programming techniques for its solution.

In this section we show how, in the presence of fixed costs of adjustment,
labour demand indeed remains dynamic, and decisions taken in one period have consequences for future choices. Rather than considering decisions in isolation, the employer has to take into account that any present chosen level of profits affects future profits. At each stage, he/she selects a decision that maximises the current period profits and the profits that are expected from future periods. Decisions are taken in stages based on information as it becomes available. In order to develop this idea we take a more "structural" approach with respect to the model discussed in the previous section. Following Rust (1992a), we define a dynamic structural model as one in which the most important features are i) the explicit treatment of time and uncertainty and ii) rational behaviour by agents: their decisions may be represented by well-defined objective functions on the basis of which they make decisions sequentially, given current information and their beliefs about the future. The preferences and beliefs are the primitives of the model. Given specific functional forms for the primitives, the structural modelling approach attempts to derive behavioral predictions from an explicit solution to an underlying optimization problem. In the case of fixed costs, the presence of discontinuities in labour demand, corresponding to the firm's decision in each period of whether or not to alter the number of workers may be best taken into account by modelling firm's optimal employment rules as a (stochastic) dynamic discrete decision process (DDP) [see for instance Eckstein and Wolpin, 1989, Rust 1987, 1991, 1992a, 1992b]. The DDP methodology is characterized by a decision variable that is restricted to a countable set of alternatives. In our case there are two alternatives: firms decide whether to change employment, which results in them bearing fixed costs of hiring and firing, or whether to postpone the adjustment to the future. The set up of a
discrete decision process for labour demand in the presence of fixed costs will be the aim of the next chapters.

A simple two-period example presented in Dixit and Pindyck (1994) provides a clear setting for the idea of optimal employment decision making. The focus, there, is on firms' capital investment choices but we may directly apply it to hiring and firing decisions. The authors provide an interpretation of adjustment costs, or at least of a part of them, as sunk costs i.e. costs causing irreversibility of the expenditure on a certain input. In the case of investment, each choice entails sunk costs, must be taken in an uncertain environment, and allows some freedom of timing. The authors point out how investment decisions, including the decision of investing in new workers, entail sunk costs due to the irreversibility of investment expenditure. For example, most expenditures on advertising or interviewing cannot be recovered and are clearly sunk costs. Irreversibility may also arise because of government regulations or institutional arrangements. For example, hiring in new workers may be partly irreversible because of the high cost of hiring, training and firing. Irreversibility implies ability to delay an investment, waiting for new information, and the comparison between the cost of delaying and the benefits from waiting for new information.

The following exercise is a comparison of the present values that result from adjusting today and from waiting. Let $K$ be the fixed cost of changing the level of employment, say investing in new workers, in a firm which then produces one widget per period forever. Suppose the price of a widget in the current period, $0$, is $P_0$. From period 1 onward, it will be $(1+u)P_0$ with probability $q$ and $(1-d)P_0$ with
probability (1-q). Let r (r>0) be the interest rate. Wages are set to be equal to one. First, suppose that the investment opportunity is available only in period 0; if the firm decides not to invest in new workers in period 0, it cannot change its mind in period 1. Let \( V_o \) denote the expected present value of the revenues the firm gets if it hires new workers. Weighting the two alternative possibilities for widget prices by their respective probabilities, discounting and adding, we obtain

\[
V_0 = P_0 + [q(1+u)P_0 + (1-q)(1-d)P_0] \left[ \frac{1}{1+r} + \frac{1}{(1+r)^2} + \ldots \right] \\
= P_0 + [1 + q(u+d)-d]P_0 \frac{1}{1+r} \frac{1}{1+\frac{1}{1+r}} \\
= P_0 \frac{1+r+q(u+d)-d}{r} \\
\tag{80}
\]

If \( V_o > K \), hiring takes place and the firm obtains \( V_o - K \); if \( V_o \leq K \) hiring does not take place and the firm obtains 0; if \( V_o = K \), the firm is indifferent between hiring and not hiring and gets zero in either case. Now, let \( \Omega_o \) denote the net pay-off to the firm if it is forced in period 0 to decide whether to hire, on a now-or-never basis. \(^{39}\)

Thus

\[
\Omega_o = \max \left[ V_o - K, 0 \right] \\
\tag{81}
\]

Now consider the actual situation, where the investment opportunity remains available in future periods. Here the period-0 decision involves a different trade-off: invest now or wait and do what is best when period 1 arrives. To assess this, the firm must look ahead to its own actions in different future eventualities. From

---

\(^{39}\) Note that, in this model, we are assuming risk-neutrality. Under risk-aversion there is a modified set of risk-adjusted probabilities, \( q' \) and \( 1-q' \), which define the decision.
period 1 onward the conditions will not change, so there is no point postponing any profitable projects beyond period 1. Hence we need look ahead only as far as period 1.

Suppose the firm does not hire in period 0, but instead waits. In period 1 the price will be

\[ P_1 = \begin{cases} (1+u)P_0 & \text{with probability } q \\ (1-d)P_0 & \text{with probability } 1-q \end{cases} \tag{82} \]

It will stay at this level for periods 2,3,... The present value of this stream of revenues, discounted back to period 1, is

\[ V_1 = P_1 + \frac{P_1}{1+r} = \frac{P_1}{(1+r)^2} + \ldots = P_1 \frac{(1+r)}{r} \tag{83} \]

For each of two possibilities (the price going up or down between periods 0 and 1), the firm will invest in new workers if \( V_1 > K \), realizing the net pay-off

\[ F_1 = \max\{V_1 - K, 0\} \tag{84} \]

We call this outcome of future decisions the \textit{continuation value}. From the perspective of period 0, the period-1 price \( P_0 \) and therefore the values \( V_1 \) and \( F_1 \) are all random variables. Let \( E_0 \) denote the expectation (probability-weighted average) calculated using the information available at period 0. Then we have

\[ E_0[F_0] = q \max\left( (1+u)P_0 \frac{(1+r)}{r} - K, 0 \right) + (1-q) \max\left( (1-d)P_0 \frac{(1+r)}{r} - K, 0 \right) \tag{85} \]

Return now to the decision at period 0. The firm has two choices. If it invests immediately, it gets the expected present value of the revenues minus the cost of investment, \( V_0 - K \). If it does not, it gets the continuation value \( E_0[F_1] \) derived above,
which starts in period 1 and must be discounted by the factor $1/(1+r)$ to express it in period-0 units. The optimal choice is obviously the one that yields the larger value. Therefore the net present value of the whole hiring opportunity optimally deployed, which we denote by $F_\infty$, is

$$F_0 = \max \left\{ V_0 - K, \frac{1}{1+r} E_0[F_1] \right\}$$ \hspace{1cm} (86)

The firm's optimal decision is the one that maximizes this net present value. This captures the essential idea of dynamic programming. We split the whole sequence of decisions into two parts: the immediate choice, and the remaining decisions, all of whose effects are summarized in the continuation value. To find the optimal sequence of decisions we work backwards. At the last relevant decision point we may make the best choice and thereby find the continuation value ($F_1$ in our example). Then at the decision point before that one, we know the expected continuation value and therefore may optimize the current choice.

Dixit and Pyndick explicitly talk about the existence of the option to postpone the decision. The decision where the hiring opportunity remains available at period 1 is less constrained than the one where it must be made on a now-or-never basis in period 0. Equation (81) shows the net pay-off $\Omega^*$ for this latter case; since that situation terminates the decision process at time 0, let us call it the termination value. 

\footnote{For the authors, a firm with an opportunity to invest is holding an "option" analogous to a financial call option - it has the right but not the obligation to buy an asset at some future time of its choosing. When a firm makes an irreversible investment expenditure, it exercises its option to invest. It gives up the possibility of waiting for new information to arrive that might affect the desirability or timing of the expenditure; it cannot disinvest should market conditions change adversely. This lost option value is an opportunity cost that must be included as part of the cost of investment. (Dixit and Pyndick, 1994)}
value at time 0. Now we have the net worth $F_0$ of the less constrained decision problem from equation (86). The difference $(F_0 - \Omega_0)$ is just the value of the extra freedom, namely the option to postpone the decision. This example well captures the intertemporal content of a choice and the discreteness characterizing the decision process.

In this example there were just two periods and that was the end of the story. Our problem covers a number of periods longer than two but the same procedure applies repeatedly. The use of dynamic programming allows us to make comparisons of this type for more general dynamic decisions with finite and infinite time horizon.

A good way of visualizing the dynamics of the optimal decision problem is given by the decision tree in Figure 39. A firm starts at the beginning of period $t$, with stock of workers $L_{t-1}$. We summarize the entire history of past decisions, $d$, and states, $L$, by $I_t = [L_{t-1}, d_{t-1}, L_{t-2}, \ldots, L_{t-n}, d_{t-n}]$, where $I_t$ indicates the history or information set that characterizes the firm at time $t$. The choice made at any time $t$ affects the outcome and the subsequent choices. Because of uncertainty, the pay-off of each decision is not fully predictable when the choice is made, but it may be observed before the next decision is made. If, at the beginning of period $t$, the company decides to change employment, this choice is indicated by $d_t = A$, it will obtain stock $L_t$. Otherwise it will maintain its employment unchanged at the same level as last

Note that using continuous time, the binomial lattice defined by the triple $(u,d,q)$ will tend in the limit to a diffusion process for the price (i.e. log-price follows a random walk). This describes the binomial method of approximating option prices. (Hull, 1989)
period, $L_{t+1}$. $d_{t} = \text{NA}$ indicates the choice of non adjusting. In period $t+1$, with the stock of workers inherited from period $t$, or history $I_t$, the firm has to face another analogous choice: whether or not to alter employment, with consequences for future periods $t-2$, $t-3$, etc. Thus, firm’s optimal employment policy results in a sequential decision process. Only at $T-1$, the final time period, is the decision problem static and, given any realized history, $I_{T-1}$, and state variable, $L_{T-1}$, the firm will select action $d_{T-1}$ in order to maximize profits. Note that, the maximization must be undertaken
for all the possible histories or states of information $I_{t,n}$, and states. The consequences of choosing an action from the set of available alternative is twofold: the firm receives an immediate reward, current profits, but also specifies a probability distribution for the subsequent system state. The firm's objective may be described as choosing a sequence of actions, a "policy", that will optimize the performance of the system over the decision making horizon. As we have indicated, since the action selected at the present time affects the future evolution of the system, the firm cannot make this choice without taking into account its future consequences. This idea will remain with us in all the subsequent chapters: at each decision point, the decision sequence must be split in two parts: its effects on the immediate period and those on the entire continuation.

As already mentioned we may take into account all the branches of the tree and to solve this stochastic control problem by using dynamic programming (Bellman, 1957). We provide an intuitive presentation of this method. According to the theory of dynamic programming, in order to characterize the profitability of a current choice we must assess the expected profitability of subsequent choices, assumed to be made optimally. Hence, at any time $t$ and in any state, the decision rule has the property that it must also be optimal for the continuation process, treating the current state as starting point. This principle, known as Principle of Optimality, allows us to compute the optimal decision rule by backward induction, starting at a finite terminal point, $T$. The logic of backward induction may be extended to infinite horizon problems. The Principle of Optimality or dynamic programming principle is not a mere computational reduction of the intertemporal optimization problem, as is sometimes claimed; it formulates the natural structural
evolution, backwards through time, of policy optimization. We will emphasize this aspect in the next section.

We notice from Figure 39 that the number of future feasible actions inevitably becomes large, both in terms of the feasible number of possibilities at each point in time and the number of periods remaining up to the firm's horizon, T. In other words, the decision tree associated with each current action tends to have many branches (Bellman's "curse of dimensionality") and this may limit our ability to solve and estimate the model without introduction of restrictions on the processes to be considered.

In the next chapter we outline the basic ingredients for dealing with the problem of intertemporal employment decisions introduced. We need to make the concepts illustrated above operational in order to determine an optimal decision rule for the maximization problem.
Chapter 5. THE PRINCIPLE OF OPTIMALITY AND THE DYNAMIC PROGRAMMING ALGORITHM: basic ingredients of the model

Introduction

5.1. Finite horizon problems

5.1.1. The firm's problem: perfect state information

5.1.2. The Principle of Optimality

5.1.3. The DP algorithm

5.1.4. Imperfect state information

5.1.5. Reformulation of the problem of imperfect state information into a problem with perfect information

5.1.6. Sufficient statistics

5.1.7. Conditional probability measures

5.1.9. Markov decision processes

5.2. Infinite horizon problems

5.2.1. Formulation of the infinite-horizon problem

5.2.2. Principle of Optimality

5.2.3. The DP algorithm

5.2.4. Contraction mapping

5.2.5. Convergence of the DP algorithm, necessary and sufficient conditions for the Bellman equation

Appendix 5.1. Measurability problems
Introduction

Before proceeding to set up a dynamic discrete decision process for modelling firms' optimal employment choices, we need to outline the basic ingredients for the optimal control problem. In the previous chapter we began to introduce a dynamic structural model as characterized by the explicit treatment of time, uncertainty and rational behaviour by agents. Rational behaviour meant that agents have well-defined objective functions and take decisions sequentially based on current information, and on their beliefs about the state of the world. In order to implement the model of dynamic labour demand we need to describe, less intuitively than in the last chapter, the technique which allows us to determine optimal policies within this intertemporal context, namely dynamic programming. We restrict ourselves to the propositions which we will directly employ in the formulation of the economic problem. We adopt a gradual approach introducing one by one the ingredients necessary to reach the final theoretical framework for the maximization problem which we will discuss in detail in the next chapter. While we maintain the assumption of fixed costs in order to link this technical presentation to our ultimate target of the formulation of a model of labour demand with fixed costs, the dynamic programming techniques we discuss are applicable to a much wider range of problems and could include quadratic costs as special case. We start by setting up the problem in the context of a finite horizon and perfect state information. This will occupy Sections 5.1.1.-5.1.4. The finite horizon case is useful for introducing the concept of backwards recursion, i.e. maximization which
proceeds backward in time from the terminal to the initial period, without encountering the complications introduced by infinite number of periods. Perfect information here refers to knowledge of the state variables relevant for the decision process. The assumption of perfect information allows us to concentrate on the role of exogenous shocks in the model as the single source of random disturbances. In this way, we may focus on the Principle of Optimality and the search for an optimal policy.

Imperfect state information, i.e. imperfect knowledge of some of the relevant states variables, adds an extra source of uncertainty to our model. Hence there are two sources of uncertainty which make the model stochastic: one is given by the disturbing random variable w, unpredictable or imperfectly predictable, which represents unexpected states occurring during the evolution of the system, e.g. exogenous shocks; the other is an "ignorance effect", which arises in the case in which some state variables, also current, may be observable only with error. The problematic aspects of this second source of disturbance is due to the fact that the lack of knowledge or imperfect knowledge on a state variable in period t is likely to persist in period t+1. In other words, there is serial correlation in this second error term. Unlike the case of the first random component, here the presence of

\[ \text{Both frameworks, with and without perfect information, are stochastic. Stochastic control is distinguished from its deterministic counterpart by the concern with when information becomes available. In deterministic control, there corresponds a sequence of control variables, which may be specified beforehand, to each initial state and policy. In contrast, if the control variables are specified beforehand for a stochastic system, the decision maker may realize in the course of the system evolution that unexpected states have appeared and the specified control variables are no longer appropriate. Rather than choosing a sequence of control variables, the decision maker attempts to choose a policy which maximizes the total expected reward.} \]
serial correlation requires additional (simplifying) assumptions. We deal with this problem, for the case of finite horizon, in Sections 5.1.5.-5.1.8., where we show how a problem of imperfect information may be reformulated into one of perfect knowledge by making maximization conditional on the decision maker's information set.

A widely used type of discrete dynamic system is the discrete-time Markov decision process. Markov processes have the defining property that all the information relevant to the determination of the probability distribution of future values is summarized in the current state. In Section 5.1.9. we discuss this in the finite horizon context.

The infinite horizon case introduces complications since an infinite number of periods implies an infinite number of decisions. This requires the study of the convergence property of the dynamic programming algorithm. The infinite horizon case is discussed in Sections 5.2.1.-5.2.5., assuming perfect state information in order to maintain simplicity. The infinite horizon formulation is more complicated than its finite horizon counterpart and requires use of more sophisticated technical tools; but it has the advantage that the resulting dynamic system is stationary and it therefore yields a stationary optimal policy. This means that the optimal employment rule is the same in each period.

Our discussion terminates with the infinite horizon problem. In the next chapter we will set up a dynamic programming problem in which companies adopt infinite horizon optimal employment policies in a context of exogenous shocks and imperfect information on certain relevant variables.

In this brief presentation we will principally follow the discussion in
Bertsekas and Shreve (1978) and Bertsekas (1987). We also refer to Ross (1983), Stokey and Lucas (1989), Puterman (1990) and Rust (1992a). For a full treatment of dynamic programming, Markov decision processes and discrete decision processes one should refer to these works.

5.1. Finite horizon problems

The main characteristics of our problem are the following:

we assume

assumption 1) a discrete time dynamic system, also known as plant equation or law of motion

\[ s_{t+1} = f_t(s_t, d_t, w_t) \]  

(87)

where \( t = 0, 1, ..., T \) indexes discrete time and \( T \) is the time horizon; \( s_t \) is state of the system and effectively summarizes past information that is relevant for future optimization making previous history irrelevant \(^1\); \( d_t \) is the control or decision variable to be selected at time \( t \), with knowledge of the state \( s_t \); and \( w_t \) is a random disturbance. Equation (87) describes how \( s_t \) evolves: it obeys a known first-order

\(^1\) Such a state variable is termed sufficient. Given the value of \( s_t \) one may calculate the optimal decision and also the future value \( s_{t+1} \). It represents a very economical description of the state of the system making previous history irrelevant as far as future profits are concerned. Otherwise, the optimal decision would be a function of the previous history of the system determined by all the values of previous decisions and states. Previous history is itself sufficient too but it takes values in an increasing space as \( t \) increases, while \( s_t \) takes values in a fixed space. It simplifies the problem to use variables which take values in a fixed low dimensional space, and indeed it is desirable that this dimension be as low as possible.

242
forward recursion in time;

*assumption 2*) an additively separable reward functional, $\Pi$, relative to our maximization problem. It is additive over time, i.e. a reward is obtained at each time $t$ and the total reward along any system sample trajectory is  

$\mathbb{E}\left\{\Pi_T(s_T) + \sum_{t=0}^{T-1} \Pi_t(s_t, a_t, w_t)\right\}$  

(88)

where $T$, the time horizon, represents the number of times a decision is taken; $T-1$ is the stage at which the last decision is taken and $T-t$ is the time to go. $\Pi_T(s_T)$ is the terminal reward obtained at the end of the process. Because of the presence of the random disturbance term, $w$, profits $\Pi$ are a random variable; we have thus to consider their expected value. The expectation, $\mathbb{E}(\cdot)$, is taken with respect to the joint distribution of the random variables involved in the process;

*assumption 3*) the mode of operation involves information gathering and sequential decision making based on information as it becomes available. The firm’s knowledge upon which decisions are based increases with the passage of time.

5.1.1. The firm’s problem: perfect state information

The firm’s objective is to find an optimal rule for choosing a number of workers to hire/fire in each period $t$, for each possible stock $L_t$ that may occur.

---

"This assumption has been called by Rust (1987, 1992a) *additive separability* (AS). We separate the final profit $\Pi_T$ in (83) because there is no decision to be taken in that period, and neither is there a shock $w_T$.  

243
Mathematically, the problem is to find a sequence of functions $\delta_i$ mapping the stock $L_t$ into the decision $d_t$ such as to maximize expected profits. Decisions are made periodically and the set $T, \{1,\ldots,T\}$ consists of all the time points, years, at which the system is observed and decisions are required. We call $t\in\{0,1,\ldots,T\}$ a stage. At each stage $t$, the employer observes the system being in state $L_t\in S_t$ and chooses an action, $d_t$ from the set of admissible actions at time $t$, $D_t$. Thus $d_t$ indicates the number of workers that should be hired/fired at time $t$ if the stock is $L_t$. The sequence $\delta=(\delta_0,\delta_1,\ldots,\delta_T)$ will be referred to as a policy or control law. A policy specifies the decision rule $\delta(L_t)$ to be used by the decision maker at each $t$, $t=0,1,\ldots,T$, namely it tells the decision maker what action to choose in order to maximize profits, in any possible future state the system might be. A policy is said to be stationary if it uses the identical decision rule in each period i.e. $\delta=(\delta^0,\delta^1,\ldots,\delta^T)$.

<table>
<thead>
<tr>
<th>$\delta_i$</th>
<th>$w_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_t$</td>
<td>$d_t$</td>
</tr>
<tr>
<td>stock at $t$</td>
<td>policy decision</td>
</tr>
</tbody>
</table>

**Figure 40 Law of motion**

The employer's objective is to maximize profits for an initial fixed stock of workers over all "admissible" $\delta$'s, i.e. to specify a policy $\delta^*$ with the largest expected total reward (we will explain the word "admissible" shortly after).
This reward functional represents the starting point for all the subsequent discussion.

We summarize the main ingredients of the problem:

- discrete time
- independent random parameters \( w_t \)
- a control constraint: \( d_t(L_t) \), as we will explain the control takes values form a set \( D_d \) which is a constrained set of \( D_r \)
- additive profits of the form (88)
- optimization over policies. Optimization involves choosing \( d_t \) for each \( t \) and possible values of \( L_t \). \( d_t(L_t) \), \( \delta_t(d_t) \) and \( \delta = (\delta_{\omega}, \delta_{\nu}, \ldots, \delta_{\tau_1}) \) represent three aspects of the same problem also illustrated in Figure 41: \( d_t \) is an action to be selected at time \( t \) with knowledge of the state \( L_t \), e.g. given \( L_t \) the employer adjusts or not employment; \( \delta_t(L_t) \) is a mapping of the stock \( L_t \) into decision \( d_t \), representing a decision rule for choosing at each period \( t \) a level of employment for each possible value of stock \( L_t \) that may occur; the policy or strategy \( \delta = (\delta_{\omega}, \delta_{\nu}, \ldots, \delta_{\tau_1}) \) is a sequence of \( \delta \)'s and represents a rule for choosing \( d_t \) for each \( L_t \) under all possible circumstances i.e. possible values of \( L_t \) over the entire horizon.

The distinction between the policy and specific decisions is crucial. Whereas decisions are taken in real time, policy will generally be formulated in advance and, indeed, implies the decisions to be taken in specified circumstances when the
### Decision, Decision Rule and Policy

<table>
<thead>
<tr>
<th>decision:</th>
<th>A, NA at t given ( L_t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_t(L_t) )</td>
<td>A, NA at each t for any possible ( L_t )</td>
</tr>
<tr>
<td>decision rule:</td>
<td>A, NA at each t for any possible value</td>
</tr>
<tr>
<td>( \delta([L_t]) )</td>
<td>A, NA at each t for any possible value</td>
</tr>
<tr>
<td>policy:</td>
<td>A, NA at each t for any possible value</td>
</tr>
</tbody>
</table>

**Figure 41 Decision, decision rule and policy**

Complete information relevant to that decision becomes available.

Selecting an action at each stage implies obtaining an immediate profit in the current period and, in addition, specifying a probability distribution over the subsequent system states. In order to make the problem operational we need to define a probability measure on the state space. In other words, for each \( \{\delta_0, \delta_1, \ldots, \delta_T\} \), we have to define an underlying probability space, a set \( \Omega \), that is, a collection of events in \( \Omega \), and a probability measure on these events. Moreover, the reward must be a well-defined random variable on this space, that is, a measurable function from the probability space into the real line. We assign a space notation to each component of our model:

- \( L_t \in S_t; L_n \) the current stock of employees i.e. the state variable is an element of the state space \( S_t \) (as is \( L_{n+1} \)).
- \( d_t \in D_{L_t}; d_0 \) the decision of whether to change employment or not and how much
to change it, i.e. the control variable, is an element of the decision space $D_{t,r}$. More precisely we constrained the control to take values from a given non empty subset of $D_{o} \supseteq D_{o}$ which depends on the current state $L_{r}$.

- $w_{i} \in W_{e}$; $w_{i}$ is an element of the disturbance space $W_{e}$. We characterize $w_{i}$ by a random probability measure $P(w_{i} | L_{r},d_{i})$ that may depend explicitly on state $L_{r}$ and on control $d_{i}$ but not on history of prior disturbances, $w_{1},...,w_{e}$. $P(w_{i} | L_{r},d_{i})$ is the probability of occurrence of $w_{i}$ when the current state and control are $L_{r}$ and $d_{i}$ respectively.

- $\Pi$ is a real-valued function $\Pi:S \times D \times W \to R$, as previously defined.

We notice how the control variable $d_{i}$ exercises influence over the transition form $L_{r}$ to $L_{r+1}$ in two places: once in the system or motion equation (87) and again as a parameter in the distribution of the disturbance $w_{e}$.

**assumption 4)** We assume that the disturbance space $W_{e}$ the domain on which the probability measure is defined, is a countable set and that the expected values of all terms in (89) exist for every admissible policy $\delta$. When the set $W$ is countable (possibly finite), the probability of its events or subsets is simply given by the probability of the events consisting of single elements in $W$. Thus if $W$ is a finite set $W=\{w_{1},w_{2},...,w_{n}\}$, the probability space is specified by the probability distribution over $W$: $p_{1}, p_{2},...,p_{n}$ where $p_{i}$ denotes the probability of the event consisting of $w_{i}$. Similarly, if it is not a finite set i.e. $W=\{w_{1},w_{2},...,w_{n},...\}$, the probability space is specified by the corresponding probabilities $p_{1}, p_{2},...,p_{n},...$; the associated collection of events or subsets is the collection of all subsets of $W$ (including $W$ and the empty...
We also say that $W$ represents an algebra (or a Boolean algebra). An algebra is a collection of subsets or events characterized by the following three properties:

i) the algebra contains the space $W$;

ii) if $A$ is an event in the algebra, allowing us to talk about the probability that $A$ occurs, then its complement is also an event in the algebra so that we can talk about the probability that $A$ does not occur;

iii) if $A_1$ and $A_2$ are two events contained in the algebra, their union is also contained in the algebra. We may extend this to property iiiia) where we consider the case of $n$ disjoint events, $A_1, A_2, \ldots, A_n$. If their union remains an event in the algebra, then we have a sigma-algebra. If a probability measure is defined on an algebra then it may be extended to a sigma-algebra. (See Mood, Graybill and Boes (1974) for more details).

If a set is uncountable, for example an interval in the $\mathbb{R}^1$ Euclidean space, it is not practicable to define a measure on the class of all its subsets: the class of all subsets despite being a well-defined sigma-algebra, a collection of all open intervals, is too large to be useful. In this case we have to restrict the problem to a smaller class of subsets we want to deal with. It is common to resort to the smallest sigma-algebra containing all the open intervals known as a Borel algebra. Within this formalisation we may now take $P_i(\cdot | L_\alpha d_i)$ as probability measure defined on the disturbance space $W_r$. For simplicity, we maintain the assumption of countability noting that all the results we use may be extended to the case of uncountable sets. In Appendix 5.1, we further discuss the computational usefulness of having $P_i(\cdot | L_\alpha d_i)$ defined on a countable support.
5.1.2. The Principle of Optimality

We consider the class of policies that consists of a sequence of functions \( \delta = \{ \delta_1, \ldots, \delta_T \} \) where \( \delta \) maps states \( L_t \) into controls \( d_t = \delta(L_t) \) and is such that \( \delta_t(L_t) \in D_t \). This control law is said to be admissible. Given an initial state \( L_0 \), the problem is to find an admissible control law \( \delta = \{ \delta_0, \delta_1, \ldots, \delta_T \} \) that maximises

\[
V_0(L_0) = E \left\{ \Pi_T(L_T) + \sum_{t=0}^{T-1} \Pi_t(L_t, \delta_t(L_t), w_t) \right\}
\]

where \( t = 0, 1, \ldots, T-1 \), subject to the system equation constraint

\[
L_{t+1} = f(L_t, \delta, w_t)
\]

\( \Pi(.) \) in equation (90) may include a discount factor. For any initial state \( L_0 \), the optimal control law \( \delta^* \) is one that maximises the corresponding reward \( ^4 \)

\[
V_{\delta^*} = \max_{\delta \in \mathcal{A}} V_0(L_0) = V^*(L_0)
\]

where \( \mathcal{A} \) is the set of all admissible control laws. We view the optimal reward as a function which assigns to each initial state, \( L_0 \), the optimal profit function or optimal value function. An intuitive presentation of the Principle of Optimality is the following. Let \( \delta^* = \{ \delta^*_0, \delta^*_1, \ldots, \delta^*_T \} \) be an optimal control law. Consider the sub-problem whereby we are at stage \( L_t \) at time \( \tau \) and wish to maximize the reward from time \( \tau \) to \( T \), hence

\[ ^4 \text{We are, here, already assuming that a maximum exists and hence we are replacing the supremum over all actions by the action corresponding to a specific policy so that we may use the "max" operator.} \]
assuming that when using $\delta^*$ the state $L_\omega$ occurs with positive probability. The truncated control law $\{\delta^*, \delta_{s+1}^*, \ldots, \delta_{T-1}^*\}$ is optimal for this sub-problem. In fact, if the truncated control law were not optimal, we would be able to attain a higher reward by switching to an optimal policy for the sub-problem once we reach stage $L_\omega$. This is indeed the principal characteristic of the Bellman's method: at any time $t$ and in any state, the policy has the property that it must be optimal for the continuation of the problem, treating the current state as the starting point. Another characteristic of the Principle of Optimality is that the sequence problem (89) may be represented in terms of a valuation function $V$ which determines the optimal policy. The valuation function enables us to decentralize a complicated multiperiod problem into a sequence of two-period decision problems, providing the correct valuation of the future consequences of current actions. At each time $t$, we split the whole sequence of decisions into two parts: the immediate choice, and the remaining decisions all of whose effects are summarized in $V_{i+1}$, the continuation value. In terms of the decision tree presented in Figure 39 in the previous chapter, it summarizes the expected future pay-offs at any node of the tree, assuming that an optimal policy will be followed in the future.

We may now state the dynamic programming (DP) algorithm which rests on the Principle of Optimality (Bellman, 1957).

\[
E\left\{\Pi(L_T) + \sum_{i=t}^{T-1} \Pi_i(L_{i+1}, d_i, w_i)\right\}
\]
5.1.3. The DP algorithm

Let $V'(L_0)$ the optimal reward. Then

$$V'(L_0) = V_o(L_0)$$ (94)

where $V_o$ is given by the final step of the following algorithm, which proceeds backwards in time from period $T$ to period $0$:

$$V_t(L_t) = \Pi_t(L_t)$$

$$V_t(L_t) = \max_{d \in \mathcal{D}_t} E_{w_t}[\Pi_t(L_t, d_t, w_t) + V_{t+1}[f_t(L_t, d_t, w_t)]]$$ (95)

If $d^*_t = \delta^*_t(L_t)$ maximizes the right hand side of (95) for each $L_t$ and $t$, the control law $\delta^* = \{\delta^*_0, \delta^*_1, ..., \delta^*_T\}$ is optimal. In general, the DP algorithm will not define closed-form solutions for $V$. However, we may always evaluate the solution by numerical methods. Even if, this must be carried out for each value of the state variable and, in certain cases, it may become computationally prohibitive. As we shall see in Sections 5.2.3.-5.2.4., from a mathematical and computational point of view, the DP algorithm may be defined in terms of a contraction mapping.

5.1.4. Imperfect state information

*assumption 5*) Assume that the controller does not have perfect knowledge of all the relevant state variables but only has access to observations $z_r$. This may reflect difficulties in obtaining exact information about some state variables: prohibitive costs, inaccessibility and imperfect monitoring. Rust (1992a) assumes that the
decision maker, the employer, has complete information on the relevant state variables but the econometrician does not. For the time being we do not need to give a specific identity to the controller and simply consider the case of lack of knowledge about the state variable, as an example of observation noise mentioned in the introduction to this chapter. We assume that, at each stage, the controller receives some observations about the value of the current state, which may be contaminated by stochastic uncertainty:

\[ z_0 = h_0(L_0, e_0) \quad z_t = h_1(L_t, d_{t-1}, e_t) \]  

where \( h \) is some function and \( e \) is a random disturbance. The observation \( z_t \) belongs to a given observation space \( Z_0 \) and the random observation disturbance \( e_t \) belongs to a given space \( E_1 \) and is characterized by a probability measure

\[ P_{e_t}(\cdot | L_0, L_0, d_{t-1}, \ldots, d_{t-1}, w_{t-1}, \ldots, w_{t-1}, e_{t-1}, \ldots, e_0) \]  

which depends explicitly on the current state and past states, controls and disturbances. The initial state \( L_0 \) may also be considered as random and characterized by a probability measure \( P_{L_0} \). The probability measure \( P_w(\cdot | L_n, d_i) \) of \( w_i \) is given and may depend explicitly on \( L_n \) and \( d_i \) but not on earlier disturbances, \( w_0, \ldots, w_{i-1}, e_{i-1}, \ldots, e_0 \). The control \( d_i \) is constrained to take values from a given non empty subset of \( D_i \) of the control space, but this time, in contrast with the hypothesis in 5.1.1., we assume that this subset does not depend on \( L_0 \) which we know imperfectly. If we cannot know the relevant state variables (or some of them) with certainty, we are forced to use the information which may contribute to the most accurate prediction of them. The best candidate for this purpose is the information set. If \( I_i \) is the information set, i.e. the information available to the controller at
At time $t$ we have

$$I_0 = z_0 \quad I_t = (z_0, \ldots, z_{t-1}, d_0, \ldots, d_{t-1}) \quad (98)$$

Consider the class of policies which consists of functions $\delta = \{\delta_0, \delta_1, \ldots, \delta_t\}$ where each function $\delta_i$ maps the information vector $I_i$ into the control space $D_i$ and $\delta_i(I_i) \in A$ for all $I_i$. We call such control law admissible. The set $A$ is the set of all admissible policies. Our problem is to find an admissible control law, $\delta = \{\delta_0, \delta_1, \ldots, \delta_t, \ldots\}$, that maximizes the following objective

$$V_\delta = \mathbb{E}_{t_0, w_0, y_0}\left\{ \Pi_T(L_T) + \sum_{t=0}^{T-1} \Pi_t[L_t, \delta_t(I_t), w_t] \right\} \quad (99)$$

subject to the system equation or law of motion

$$L_{t+1} = f_t[L_t, \delta_t(I_t), w_t] \quad (100)$$

and the measurement equation

$$L_0 = h_0(L_0, e_0)$$

$$L_t = H_t[L_t, \delta_{t-1}(I_{t-1}), e_t] \quad (101)$$

The profit functions $\Pi$ are as defined before.

5.1.5. Reformulation of the problem of imperfect information into a problem with perfect information

What is the difference between this and the perfect information case? In the case of perfect information about the state we were trying to find a rule that would specify the control $d_t$ to be applied for each state $L_t$ and time $t$. With imperfect information we are looking for a rule that gives the control to be applied for every
possible state of information, that is for every sequence of observations and controls actuated up to time $t$. We may reformulate the imperfect information case into the basic problem with perfect information. Intuitively, this may be done by defining a new system the state of which consists, at time $t$, of all variables the knowledge of which may be of benefit to the controller when making the $t^{th}$ decision; in other words the information vector $I_t$. The new equations which describe the evolution of the system are now the following:

$$I_{t+1} = (I_t Z_{t+1}, d_t)$$  \hspace{1cm} (102)

which gives the new law of motion. The state of the system is $I_t$, the control is $d_t$ and $Z_{t+1}$ may be viewed as a random disturbance with probability

$$P(z_{t+1} \in Z_{t+1} | I_t d_t) = P(z_{t+1} \in Z_{t+1} | I_t d_t, z_0 z_1, \ldots, z_t)$$  \hspace{1cm} (103)

for any event $z_{t+1}$, a subset of $Z_{t+1}$, since $z_0, \ldots, z_t$ are part of the information vector $I_t$. Thus the probability measure of $z_{t+1}$ depends explicitly only on the state $I_t$ and control $d_t$ of the new system (103) and not on the prior disturbances $z_0, \ldots, z_t$. We may reformulate the problem in terms on the new variables and obtain a problem with perfect state information

$$E\{\Pi_t(L_t d_t w_t)\} = E\{E_{L_t w_t}\{\Pi_t(L_t d_t w_t) | I_t d_t\}\}$$  \hspace{1cm} (104)

The new profit function is

$$\Pi_t(I_t d_t) = E_{L_t w_t}\{\Pi_t(L_t d_t w_t) | I_t d_t\}$$  \hspace{1cm} (105)

The DP algorithm is now
\[ V_{T-1}(I_{T-1}) = \max_{d_{T-1} \in A_{T-1}} \left[ E_{I_{T-1},w_{T-1}} \left( \Pi_{T-1}(f_{T-1}(I_{T-1},d_{T-1},w_{T-1})) \right) + \Pi_{T-1}(L_{T-1},d_{T-1},w_{T-1}) | I_{T-1},d_{T-1} \right] \]  \[ (106) \]

and

\[ V_t(I_t) = \max_{d_t \in A_t} \left[ E_{I_t,w_t} \left( \Pi_t(I_t,d_t,w_t) + V_{t+1}(I_{t+1},d_{t+1}) | I_{t+1},d_{t+1} \right) \right] \]

\[ (107) \]

Equations (106) and (107) form the basic DP algorithm for the dynamic programming problem with imperfect state information. An optimal policy \( \delta_0^*, \ldots, \delta_T^* \) is obtained by first solving the maximization problem (106) for every possible value of the information vector \( I_{t-1} \) to obtain \( \delta_{t-1}^*(I_{t-1}) \). Simultaneously, \( V_{T-1}(I_{T-1}) \) is computed and used in the computation of \( V_T(I_T) \) via the maximization in (106), which is carried out for every possible value of \( I_T \). Proceeding similarly, we obtain \( V_{T-2}(I_{T-2}) \) and \( \delta_{T-2}^* \); we continue until \( V_0(I_0) = \mathbb{V}_0(z_0) \) is computed and the optimal reward is obtained as

\[ V^* = \mathbb{V}_0(\{V_0(z_0)\}) \]

\[ (108) \]

5.1.6. Sufficient statistics

The familiar problem with dynamic programming is that the algorithm (106) and (107) must be carried out over a state space of expanding dimensions as is clearly evident in the decision tree in Figure 39, in Chapter 4, well shows. In particular, in the imperfect information case, as a new measurement is added at each stage \( t \), the dimension of the state and of the information set \( I_t \) increases accordingly. If we could reduce the problem to one in which we consider only the data that are strictly necessary for the specific control purposes, and thereby "prune
the tree" we would obtain a more tractable problem. This amounts to looking for *sufficient statistics* which ideally should be of smaller dimension than $I_t$ and yet summarize all the essential content of $I_t$ so far as the control is concerned. Suppose that we can find a function $g_t(I_t)$ of the information vector such that the maximization of (106) and (107) depends on $I_t$ via $g_t(I_t)$. The maximization in the right hand side of the DP algorithm may be written in terms of some function $H_t$ as

$$\max \{H[I_t g_t(I_t)d_t]\}$$

(109)

and the function $g_t(.)$ is a sufficient statistic for the control problem. The optimal control law may be written as

$$\delta_t(I_t) = \delta^* g_t(I_t)$$

(110)

where $\delta^*$ is an appropriate function. This allows us to take advantage of the data reduction.

5.1.7. Conditional probability measures

*assumption 6)* If we assume that the probability distribution of the observation disturbance $\epsilon_{t+1}$ depends explicitly only on the immediately preceding state, the control and the system disturbance $L_t$, $d_t$, $w_t$ and not on $L_{t-1}$, ..., $L_{t-n}$, $d_{t-1}$, ..., $d_{t-n}$, $w_{t-1}$, ..., $w_{t-n}$, we may show that a sufficient statistics is given by the conditional probability measure $P_{L_t|I_t}$ of the state $L_t$ given the information vector $I_t$. The conditional probability $P_{L_t|I_t}$ is generated recursively in time and may be viewed as the state of a controlled discrete-time dynamic system. By using Bayes' rule we may write for
where $\Phi$ is some function that may be determined from the data of the problem, $d_t$ is the control of the system and $z_{t+1}$ is a random disturbance, the statistics of which are known and depend explicitly on $P_{t+1}$ and $d_t$, only and not on $z_t, ..., z_{t'}$. Equation (111) is the new law of motion in the presence of imperfect state information; $P_{t+1}$ is the state variable redefined in probabilistic terms. The conditional density $p(L_{t+1} | I_t)$ is generated from $p(L_t | I_t)$, $d_t$ and $z_{t+1}$ by means of the equation

$$p(L_{t+1} | I_t, d_t) = \frac{p(L_{t+1} | I_t, d_t, z_{t+1})}{p(L_{t+1} | I_t, d_t)} = \frac{\int - \infty \int - \infty p(L_{t+1} | I_t, d_t, z_{t+1}) \, dL_{t+1} \, dz_{t+1}}{\int - \infty \int - \infty p(L_{t+1} | I_t, d_t) \, dL_{t+1} \, dz_{t+1}}$$

All the probabilities densities appearing on the right hand side may be expressed in terms of $p(L_t | I_t)$, $d_t$ and $L_{t+1}$ alone. In particular, the density $p(L_{t+1} | I_t, d_t)$ may be expressed through $p(L_t | I_t)$, $d_t$ and the system equation $L_{t+1} = f_t(L_t, d_t, w_t)$ using the given density $p(w_t | L_t, d_t)$ and the relation

$$p(w_t | I_t, d_t) = \int - \infty \int - \infty p(L_t | I_t) \, p(w_t | L_t, d_t) \, dL_t$$

Similarly the density $p(z_{t+1} | I_t, d_t, L_{t+1})$ is defined by the measurement equation $L_{t+1} = h_t(L_{t+1}, d_t, e_{t+1})$ using $p(L_t | I_t)$, $p(w_t | L_t, d_t)$ and the specified probability density $p(e_{t+1} | L_t, d_t, w_t)$. Rust (1987, 1992a) calls (106) the conditional independence hypothesis (CI). By substituting these expressions in the equation for $p(L_{t+1} | I_t)$, we obtain an equation of the form (111). In principle, the controller may calculate
the conditional choice probability measure $P_{\text{Lijt}}$ and hence be back in a situation of perfect state information. The DP algorithm may be now rewritten in terms of the sufficient statistics given by the conditional choice probability, $P_{\text{Lijt}}$ and using the system equation (111)

$$
\overline{V}_{T-1}(P_{L_{T-1}|I_{T-1}}) = \max_{d_{T-1},w_{T-1}} \left[ E_{L_{T-1},w_{T-1}} \left[ \Pi_{T-1} \left( f_{T-1}(L_{T-1},d_{T-1},w_{T-1}) \right) + \Pi_{T-1}(L_{T-1},d_{T-1},w_{T-1}) | I_{T-1},d_{T-1} \right] \right] + (114)
$$

and

$$
\overline{V}_t(P_{L_t|I_t}) = \max_{d_t} \left[ E_{L_t,w_t} \left[ \Pi_t(L_t,d_t,w_t) + \overline{V}_{t+1}(P_{L_t|I_t},d_{t+1}) \right] | I_t,d_t \right] \right] (115)
$$

The DP algorithm yield a control law of the form

$$
d_t^* = \delta_t^* (P_{L_t|I_t}) \tag{116}
$$

The optimal reward is given by

$$
V^* = E_{e_0} \left\{ \overline{V}_0(P_{L_0|I_0}) \right\} \tag{117}
$$

where $V_0$ is the last step of the DP algorithm and the probability measure of $z_0$ is obtained from the statistics of $L_0$ and $e_0$ and the measurement equation $z_0 = h_0(L_0,e_0)$.

We have now obtained an alternative reduction of the basic problem with imperfect state information to a problem with perfect state information. This involves the system (106), the state of which is $P_{\text{Lijt}}$ and an appropriately reformulated reward functional. As $L_t$, the state variable in the case of perfect information, the conditional probability $P_{\text{Lijt}}$ summarizes all the information that is necessary for control purposes at period $t$. In the absence of perfect knowledge of the state, the controller may be viewed as controlling the "probabilistic state"
P_{L_t} so as to maximize the expected reward conditional on the information I, available at time t.

5.1.8. Markov decision processes

We now give a specific structure to the discrete dynamic system such as (87): \(L_{t+1} = f(L_t, d_t, w_t)\), that is we aim to give an expression to the function \(f\). For this purpose we define a finite horizon \textit{discrete-time Markov decision process} as the collection of objects \((T, S, D, p_{t+1}(L_{t+1} | L_t, d_t))\) where

- \(T\) is a time set, \(0 \leq T < \infty\) with elements \(t \in \{0, 1, 2, \ldots, T\}\).
- \(S_t\) is the state space
- \(D_{t,\alpha}\) is the decision space
- \(p_{t+1}(L_{t+1} = j | L_t = i, d_t)\) a transition probability function. The term transition probability denotes the probability that the system is in a certain state \(j \in S_{t+1}\) if action \(d_t \in D_{t,\alpha}\) is selected when in state \(L_t\) at time \(t\).

Then the expected reward \(\Pi_t(L_t, d_t)\) in period \(t\) is given by

\[
\Pi_t(L_t = i, d_t) = \sum_{j \in S_{t+1}} \Pi_t(L_t = i, d_t) p_{t+1}(L_{t+1} = j | L_t = i, d_t)
\]

(118)

\(\Pi_t\) is the reward received in period \(t\) if the state of the system at time \(t\) is \(i = L_t\) decision \(d_t\) is selected and the system is in state \(j = L_{t+1}\) at time \(t+1\). The distinguishing feature of MDP is that the transition probability and the reward function depend only on the current state of the system and the current choice.
selected i.e. it is not necessary to keep track of the entire previous history of the system. A Markovian decision rule is a function \( \delta : S_t \rightarrow D_{L_t} \) and a sequence of decision rules \( \delta = (\delta_\tau, \ldots, \delta_1, \delta_0) \) form a Markovian policy.

**The Principle of optimality for the Markov decision problem.** The controller’s objective is to specify at decision stage \( t \) a policy \( \delta \) with the largest expected total reward. As we know, the problem may be written in terms of a value function

\[
V_t^*(L_t) = \Pi(L_t) + \sum_{i \in S_{t-1}} \left( \Pi(L_{t-1}, i) + \beta \sum_{j \in S_{t+1}} V_{t+1}^*(L_{t+1}, j) \right) p_{t+1}(L_{t+1} = j | L_t = i, D_t)
\]  

Equation (119) is the optimality equation for the Markov problem. The Markov optimal policy \( \delta_t^*(L_t) \) is given by

\[
\delta_t^*(L_t) = \text{argmax}_{i \in D(L_t)} \left( \Pi(L_t, i) + \beta \sum_{j \in S_{t+1}} V_{t+1}^*(L_{t+1}, j) \right) p_{t+1}(L_{t+1} = j | L_t = i, D_t)
\]  

(120)

where the operation argmax corresponds to choosing an action which attains the maximum on the right hand side of (118). It is not necessarily unique. Equation (120) is simply another way of representing (119).

**The firm’s problem**

We may employ the MDP structure to illustrate the firm’s employment optimal decision process. In this case the MDP represents an employment decision process that is observed at time points \( t \) (where \( t = 0, 1, \ldots, T \)) to be in one of the two possible states: adjustment or non adjustment. Hence, after observing the state of
the process the employer must decide on what action to take. For instance, the employer finds himself in period t, after having changed employment in period t-1 (observed state), and must decide whether to further vary the number of employees or whether to postpone adjustment to a subsequent period, thereby leaving current employment unchanged. The transition probability describes the evolution of this system: the probability of adjusting given that the firm adjusted last period, or its complement i.e., the probability of non adjusting given that the firm adjusted last period. Transition probabilities are functions only of the current state and action and they fully specify the adjustment process.

5.2. Infinite horizon problems

To what extent do the finite horizon results carry over to the case of an infinite horizon? The infinite horizon problems have the two following defining characteristics: the number of stages is infinite, and the system is stationary. Stationarity implies that the system equation, the per-stage profit function and the random disturbance statistics, do not change from one stage to the next. It provides a reasonable approximation to a situation in which the system parameters vary slowly with time. The assumption of stationarity substitutes for that of time homogeneity which characterized finite horizon models. In the previous pages we have provided expressions for the dynamic adjustment processes and rewards as being independent of time. In other words, the system equation and the reward functional do not involve t as an explicit argument.

In the case of an infinite horizon, a stronger property holds: that of time
invariance which demands that the whole structure of the process be invariant under a translation of the time axis. This property, which does not hold for finite horizon models, implies stationarity of the optimal policy. This represents an advantage in considering the infinite horizon case: the implementation of optimal policies is often simple and policies are typically stationary. This means that the optimal rule for applying controls does not change from one stage to the next.

Two complications arise when we deal with infinite horizon: first, we need to study the convergence properties of the DP algorithm and of the corresponding optimal policies; second, we need a rigorous consideration of the probabilistic aspects of the problem involving the possibility of uncountable disturbance spaces. This requires the use of measure-theoretic probability theory as discussed in Section 5.1.2. As in the finite-horizon models we restrict ourselves to the case where the disturbance space is a countable set.

### 5.2.1. Formulation of the infinite-horizon problem

In order to provide a classification for those infinite-horizon problems that are of major interest we first assume perfect information about the state variables and subsequently extend it to the case of imperfect information, using the methodology discussed in 5.1.6-8. Traditionally, we have three classes of infinite-horizon problems:

i) *the discounted case with bounded reward per stage* is the simplest infinite horizon problem. The functional for the reward takes the form
where \( V_\delta(L_o) \) denotes the cost associated with the initial stage \( L_o \) and a policy \( \delta = \{\delta_o, \delta_i, \ldots\} \) and \( \beta, 0 < \beta < 1 \), is a scalar or discount factor. The reward per stage is uniformly bounded from above and below. There is a contraction mapping underlying the DP iteration and effective computational methods are available for solution.

ii) the case with unbounded reward per stage has the same functional form as the discounted case with bounded reward except that \( \beta \), although positive, is not necessarily less than 1. Moreover, the reward per stage is allowed to be unbounded either from above or from below.

iii) the case of average reward per stage where, rather than (121), we maximize

\[
\lim_{T \to \infty} \frac{1}{T} E_w \left\{ \sum_{t=0}^{T-1} \Pi[L_o, \delta(L_t), w_t] \right\}
\]

(122)

This maximization is useful if it turns out that \( V_\delta(L_o) \) is infinite but the limit (122) is finite for every policy \( \delta = \{\delta_o, \delta_i, \ldots\} \) and initial state \( L_o \). If \( V_\delta(L_o) \) is finite for at least some admissible policies and some initial states then maximization may proceed as in i).

In the remainder of this work we will represent the infinite horizon problem as the discounted case reported in i) and assume therefore that \( 0 < \beta < 1 \).
5.2.2. Principle of Optimality

Consider again the stationary discrete-time dynamic system

\[ L_{t+1} = f(L_t, d^t, w_t) \]  \hspace{1cm} (123)

where as in the case of 5.1.2.:

- the state \( L_t \) is an element of the state space \( S_t \)
- the control \( d_t \) is an element of the space \( D_{L_t} \), a non empty subset of \( D_t \) which depends on the current state \( L_t \)
- the random disturbance \( w_t \) is an element of a space \( W_t \), where we assume that \( W \) is a countable set.

The random disturbances have identical statistics and are characterized by probabilities \( P(\cdot | L^t, d^t) \) defined on \( W \). The probability of \( w_t \) does not depend on values of prior disturbances \( w_{t-1}, \ldots, w_0 \).

Given an initial state \( L_0 \), the problem is to find a policy \( \delta = \{\delta_0, \delta_1, \ldots\} \), where \( w_t: S_t \to D_{L_t} \) for all \( L_t \in S_t \), \( t = 0, 1, \ldots \), which maximizes the reward

\[ V_\delta(L_0) = \lim_{T \to \infty} \mathbb{E}_{w_t} \left\{ \sum_{t=0}^{T-1} \beta^t \Pi[L_t, \delta_t(L_t), w_t] \right\} \]  \hspace{1cm} (124)

subject to the system equation (123). The reward given in (124) represents the limit of a finite horizon reward. Denote as before the set of all admissible policies \( \delta \), by \( A \). This is the set of all sequences of functions \( \delta = \{\delta_0, \delta_1, \ldots\} \) with \( \delta_t: S \to D_t \), \( \delta_t(L_t) \in D_{L_t} \). The optimal reward \( V^* \) is given by

264
A class of admissible policies is the class of stationary policies of the form \( \delta = \{ \delta^0, \delta^1, \ldots \} \) for which the rule for control selection is the same at every stage, hence

\[
V_\delta(L_t) = \lim_{T \to \infty} E_{w_t} \left\{ \sum_{t=0}^{T-1} \beta^t \Pi_{\delta_t(L_t), w_t} \right\}
\]

Similarly to \( V^* \), \( V_\delta \) is a well defined real-valued function.

5.2.3. The DP algorithm

Since in the infinite horizon problem the number of stages is not fixed, in order to show the following results, we need to consider our problem in a slightly different to that in Section 5.1.4. by changing the indexation of the stage functions.

We consider a T-stage problem obtained by truncating an infinite horizon problem. We need to find a policy \( \delta_T = \{ \delta_0, \ldots, \delta_T \} \) with \( \delta_t(L_t) \in D_{L_t} \) for all \( L_t \in S_t \) that maximises

\[
V_{\delta_T}(L_0) = E_{w_t} \left\{ \sum_{t=0}^{T-1} \beta^t \Pi_{\delta_t(L_t), w_t} \right\}
\]

subject to the system equation constraints. The optimal reward for each initial stage \( L_0 \) is \( V_0(L_0) \), where \( V_0 \) is given by the last step of the DP algorithm:

\[
R_T(L_T) = \Pi_T(L_T)
\]

\[
R_t(L_t) = \max_{\delta_t \in D_{L_t}} E_{w_t} \left\{ \beta^t \Pi_{\delta_t(L_t), w_t} + R_{t+1}(L_{t+1}, d_{t+1}, w_{t+1}) \right\}
\]

Dividing both sides of the second expression in (128) by \( \beta^t \) we obtain
\[ V_{T-t}(L_t) = \beta^{-t}R(L_t) \quad (129) \]

and we may write

\[
V_0(L) = \Pi_0(L) \\
V_{t+1}(L) = \max_{d \in D_L} E_w \{ \Pi(L,d,w) + \beta \, V_t[f(L,d,w)] \} \quad (130)
\]

The optimal reward is \( V_t(L_0) \). The algorithm (130) is equivalent to the ordinary DP algorithm. The main difference is that the indexation of the stage functions has been reversed so that, now, the algorithm proceeds from lower to higher values of the index \( t \). We may interpret \( V_t(L) \) as the maximal reward that can be obtained starting at stage \( L \) and proceeding for \( t \) (rather than \( T-t \) stages (i.e., it is the \( k \)-stage optimal reward). Since the number of stages is not fixed in an infinite horizon problem, working with \( V_t \) rather than \( R_t \) is convenient. For this reason we will use the form of the DP algorithm as in (130).

We aim to develop relations between the \( T \)-stage problem with reward (128) and its infinite horizon counterpart, but before doing so we need to introduce some notation.

For any function \( V:S \to \mathbb{R} \), \( \delta:S \to D \) with \( \delta(L) \in D_L \) for all \( L \in S \), we denote

\[
T(V)(L) = \max_{d \in D_L} E_w \{ \Pi(L,d,w) + \beta \, V[f(L,d,w)] \} \quad (131)
\]

and

\[
T_\delta(V)(L) = E_w \{ \Pi(L,d,w) + \beta \, V[f(L,d,w)] \} \quad (132)
\]

assuming that all the expected values are well defined. \( T(V)(\cdot) \) and \( T_\delta(V)(\cdot) \) are functions defined on the state space \( S \), and \( T, T_\delta \) may be viewed as mapping that transform a function \( V \) on \( S \) into another function \( [T(V) \text{ or } T_\delta(V)] \) on \( S \).
The mapping $T$ and $T_g$, also known as a contraction mappings, play an important theoretical role and provide a convenient shorthand notation for expressions that would be too complicated to write otherwise. From the definitions (131) and (132) it may be seen that $T(V)(L)$ is the optimal reward for the one-stage problem with initial state $L$, stage reward $\Pi$, and terminal reward function $\beta V$. Similarly, $T_g(V)(L)$ is the reward corresponding to policy $\{\delta^0, \delta^1, \ldots\}$ for the same problem.

Denote by $T^t$ the composition of the mapping $T$ with itself $t$ times; that is, for all $L$ and $t$

$$T^t(V)(L) = T[T^{t-1}(V)](L), \quad T^0(V) = V(L) \tag{133}$$

It is seen that $T^t(V)(L)$ is the optimal reward for the $t$-stage, $\beta$-discounted problem with the initial state $L$, reward per stage $\Pi$, and the terminal reward function $\beta V$. Similarly, $T^t_g(V)(L)$ is the cost of a policy $\{\delta^0, \delta^1, \ldots\}$ for the same problem. In terms of this notation the DP algorithm (130) which corresponds to a finite horizon problem with a terminal reward function $\Pi$, may be written

$$V_0(L) = \Pi(L)$$
$$V_t(L) = T^t(V_0)(L) \tag{135}$$

Finally consider a $t$-stage policy for a $t$-stage problem $\{\delta_0, \delta_1, \ldots, \delta_t\}$. Then $(T_{\delta_0} T_{\delta_1} \cdots T_{\delta_{t-1}})(V(L))$ is defined recursively for $i=0,\ldots,t-2$ by

$$T_{\delta_i} T_{\delta_{i+1}} \cdots T_{\delta_{t-1}}(V)(L) = T_{\delta_i} [T_{\delta_{i+1}} \cdots T_{\delta_{t-1}}(V)](L) \tag{136}$$

and represents the reward for the policy for the $t$-stage, $\beta$-discounted problem with
5.2.4. Contraction mapping

Let $B(S)$ denote the set of all bounded real-valued functions on $S$. With every function $V:S \to \mathbb{R}$ that belongs to $B(S)$ we associate the scalar

$$|V| = \max_{L \in S} |V(L)|$$

(137)

where the function $\| \cdot \|$ is a norm on the linear space $B(S)$. With this norm, $B(S)$ becomes a complete normed linear space, i.e. a Banach space. We may define a contraction mapping as the following: a mapping $T:B(S) \to B(S)$ is said to be a contraction mapping if there exists a scalar $\rho < 1$ such that

$$\|T(V) - T(V')\| \leq \rho \|V - V'\|$$

(138)

for all $V, V' \in B(S)$. We may also have a $t$-stage contraction mapping $\|T(V) - T(V')\| \leq \rho^t \|V - V'\|$ for all $V, V' \in B(S)$, where $t$ is a positive integer and $\mathbb{H}^n$ is the composition $T \circ \cdots \circ T$ of $T$ with itself $t$ times.

The **Contraction mapping fixed point theorem**: if $T$ is a contraction mapping or a $t$-stage contraction mapping, there exists a unique fixed point of $T$; that is there is a unique function $V' \in B(S)$ such that $T(V') = V'$. Furthermore, if $V$ is any function in $B(S)$ and $T^*$ is the composition of $T$ with itself $t$ times, then

$$\lim_{t \to \infty} \|T^t(V) - V^*\| = 0$$

(139)

Setting $\rho$ equal to the discount factor $\beta$ in (138) we can see the contraction property induced by the discount factor in (131) and (132).
5.2.5. Convergence of the DP algorithm, necessary and sufficient conditions for the Bellman's equation

The first important property of the DP algorithm is monotonicity. For any functions $V:S \rightarrow \mathbb{R}$, $V':S \rightarrow \mathbb{R}$, such that $V(L) \leq V'(L)$, for all $L \in S$, and for any function $\delta:S \rightarrow D$ with $\delta(L) \in D_{L}$ for all $L \in S$, we have that

\begin{align*}
T^{t}(V)(L) & \leq T^{t}(V')(L), \quad L \in S, \quad t = 1,2,\ldots \\
T_{\delta}^{t}(V)(L) & \leq T_{\delta}^{t}(V')(L), \quad L \in S, \quad t = 1,2,\ldots
\end{align*}

(140)

The second property of the DP algorithm is convergence.

For any bounded function $V:S \rightarrow \mathbb{R}$, it is the case that

\begin{equation}
V^{*}(L) = \lim_{t \rightarrow \infty} T^{t}(V)(L)
\end{equation}

(141)

for all $L \in S$. Application of this proposition leads to the following corollary: let $V_{\delta}(L)$ be the value of the reward functional (121) corresponding to a stationary policy $\{\delta,\delta,\ldots,\delta\}$ when the initial state is $L$. Then for any bounded function $V:S \rightarrow \mathbb{R}$

\begin{equation}
V_{\delta}(L) = \lim_{t \rightarrow \infty} T_{\delta}^{t}(V)(L)
\end{equation}

(142)

holds for $L \in S$.

We may now state the necessary and sufficient conditions for optimality: i) the optimal reward function $V^{*}$ satisfies

\begin{equation}
V^{*}(L) = \max_{d \in D_{L}} \mathbb{E}_{w}\left\{\Pi(L,d,w) + \beta V^{*}[f(L,d,w)]\right\}
\end{equation}

(143)

$L \in S$; or equivalently

\begin{equation}
V^{*}(L) = T(V^{*})(L)
\end{equation}

(144)
Furthermore $V'$ is the unique bounded solution of this equation;

ii) a stationary policy $\{\delta', \delta', \ldots, \delta'\}$ is optimal if and only if $\delta'(L)$ attains the maximum in (143) for all $L \in S$; that is

$$ T(V')(L) = T_{\delta'}(V')(L) $$

Equation (143) corresponds to the Bellman equation. The optimal reward function $V'$ is the unique bounded solution to (143). The Bellman equation yields an optimal stationary policy provided the maximum in its right hand side is attained.

iii) Furthermore, the DP algorithm yields, in the limit, the function $V'$ starting from an arbitrary bounded function $V$, and the rate of convergence is at least as fast as the rate of a convergent geometric progression:

$$ \max_{L \in S} |T'(V)(L) - V'(L)| \leq \beta \max_{L \in S} |V(L) - V'(L)| $$

for any bounded function $V$. (For the proofs of these propositions see Bertsekas (1987, pp. 183-186).

With the mathematical preliminaries behind us, we may now turn to the analysis of employment decisions under uncertainty in the presence of fixed costs of adjustment. This will be the topic of the next two chapters.
Appendix 5.1. Measurability problems

This appendix uses the notation and algorithm outlined in the main sections of this chapter. Consider the following two-stage problem:

\[ V_\delta(L_0) = \int \Pi[L_1, \delta_1(L_1)] p(dL_1 | L_0, \delta_0(L_0)) \quad \text{(A1)} \]

Knowing the initial state \( L_0 \in \mathbb{R} \), the real line - the decision maker selects the control \( d_0 \in \mathbb{R} \). A state \( L_1 \in \mathbb{R} \) is generated according to the probability \( p(L_1 | L_0, d_0) \) on \( \mathcal{B}_\mathbb{R} \), the Borel subsets of \( \mathbb{R} \), depending on \( L_0 \) and \( d_0 \). Knowing \( L_1 \), the decision maker select \( d_1 \in \mathbb{R} \). Given \( p(L_1 | L_0, d_0) \) for every \((L_0, d_0) \in \mathbb{R}^2\) and a function \( \Pi : \mathbb{R}^2 \to \mathbb{R} \). The problem is to find a policy \( \delta = (\delta_0, \delta_1) \) consisting of two functions \( \delta_0 : \mathbb{R} \to \mathbb{R} \) and \( \delta_1 : \mathbb{R} \to \mathbb{R} \) that maximizes (A1). Assume no profits at the first stage and that \( \Pi, \delta_0 \) and \( \delta_1 \) are such that the integral in (A1) is well-defined. The DP algorithm associated with this problem is:

\[ V_1(L_1) = \sup_{\delta_1} \Pi(L_1, d_1) \quad \text{(A2)} \]

\[ V_2(L_0) = \sup_{\delta_0} \int V_1(L_1) p(dL_1 | L_0, d_0) \quad \text{(A3)} \]

and assuming that \( V_2(L_0) < \infty \) and \( V_1(L_1) < \infty \) for all \( L_0 \in \mathbb{R} \) and \( L_1 \in \mathbb{R} \), the results one expects to be true are:

R1. The following result holds:

\[ V_2(L_0) = \sup_\delta V_\delta(L_0) \quad \forall \ L_0 \in \mathbb{R} \quad \text{(A4)} \]

R2. Given \( \epsilon > 0 \), there is an \( \epsilon \)-optimal policy, i.e. a policy \( \delta_* \), such that

\[ V_{\delta_*}(L_0) \leq \sup_\delta V_\delta(L_0) + \epsilon \quad \forall \ L_0 \in \mathbb{R} \quad \text{(A5)} \]
R3. If the supremum in (A2) and (A3) is attained for all \( L \in \mathbb{R} \) and \( l \in \mathbb{R} \), there exists a policy that is optimal for every \( L \in \mathbb{R} \).

R4. If \( \delta^*(L) \) and \( \delta_0^*(L) \), respectively, attain the supremum in (A2) and (A3) for all \( L \in \mathbb{R} \) and \( l \in \mathbb{R} \), then \( \delta^* = (\delta_0^* \circ \delta^*) \) is optimal for every \( L \in \mathbb{R} \), i.e.,

\[
V_{\delta^*}(L) = \sup_{L} V_{\delta}(L) \quad \forall L \in \mathbb{R} \tag{A6}
\]

A formal derivation of (R1) consists of the following steps:

\[
sup_{\delta} V_{\delta}(L) = \sup_{\delta_0} \sup_{\delta_0} \int \Pi[L_1, \delta_1(L_1)] p(dL_1 | L_0, \delta_0(L_0)) \tag{A7a}
\]

\[
= \sup_{\delta_0} \{ \sup_{\delta_1} \Pi(L_1, \delta_1) \} p(dL_1 | L_0, \delta_0(L_0))
\]

\[
= \sup_{\delta_0} \int V_1(L_1) p(dL_1 | L_0, \delta_0(L_0)) \tag{A7b}
\]

\[
= \sup_{\delta_0} \int V_1(L_1) p(dL_1 | L_0, 0) = V_2(L_0)
\]

Similar formal derivations can be given for (R2), (R3), and (R4). The following points need to be emphasized: in (A7a), \( \Pi \) and \( \delta_1 \) must be such that \( \Pi[L_1, \delta_1(L_1)] \) can be integrated in a well-defined manner; in (A7b) the interchange of supremization and integration must be legitimate. Furthermore, \( \Pi \) must be such that \( V_1(L_1) \) can be integrated in a well-defined manner.

If for each \( (L_0, 0) \), \( p(L_1 | L_0, 0) \) has countable support, i.e., is concentrated on a countable number of points, then integration in (A7a) and (A7b) reduces to supremum summation. Thus there is no need to impose measurability restrictions on \( \Pi \), \( \delta_1 \) and \( \delta_0 \) and the interchange of supremization and integration in (A7b) is justified in the view of the assumption.
If \( p(L_1 \mid L_{\omega, d_0}) \) does not have countable support, there are two main approaches. The first is to expand the notion of integration, and the second is to restrict \( \Pi, \delta, \) and \( \delta_0 \) to be appropriately measurable. The major alternative approach was initiated in more general form by Blackwell (1965). We assume that \( \Pi \) is Borel-measurable, and furthermore, for each \( \mathcal{B}_{B_R} (B_R \text{ is the Borel } \sigma\text{-algebra on } \mathbb{R}) \), the function \( p(D \mid L_{\omega, d_0}) \) is Borel-measurable in \( (L_{\omega, d_0}) \). (See Bertsekas and Shreve (1978) for further developments, in particular the extension to universally measurable policies).
Chapter 6. A DYNAMIC MODEL OF LABOUR DEMAND WITH FIXED ADJUSTMENT COSTS

Introduction

6.1. A structural model

6.2. Reduction of the model

6.3. S,s-type policy

Appendix 6.1. The Logit model
Introduction

In this chapter we will develop the concepts introduced in Section 4.2. and model firms' optimal employment choices as a sequential decision process, over an infinite time horizon, in an environment which displays uncertainty with regard to output prices and wages. The main ingredients of the problem are:

i) discrete time. This represents a natural assumption in modelling firms' employment decisions and is consistent with the characteristics of the data;

ii) fixed costs of adjusting labour. This implies that the firm may optimally decide not to change employment if the profits expected after adjusting do not at least outweigh fixed costs. Together with discrete time, fixed costs require the framework of discrete dynamic processes as will be illustrated later.

iii) uncertainty. We have two sources of uncertainty. The first, which drives the dynamics in the model, is given by the possibility of shocks to output prices and wages. This may entail a revision of the optimal plans after the realization of a shock. The sequential process comprises decisions taken in stages based on information as it becomes available.

The second source of uncertainty is given by the lack of perfect knowledge about the state variables relevant for the decision process. This may reflect difficulties in obtaining exact information about some state variables: prohibitive costs, inaccessibility and imperfect monitoring. This introduces an observation noise which is going to play a large role in the model which will follow. We assume that each firm knows the state variables relevant for its optimal employment plans but
the econometrician only has imperfect knowledge about these variables.

iv) discrete decision process. Discreteness in the model is not only introduced by
the assumption of discrete time but also by the fact that the decision variable is
restricted to a countable set of alternatives. In the present model there are two
alternatives: to adjust, and pay fixed costs, or not to adjust and avoid fixed costs.

The essential technical aspects of the discrete decision process were
explained in Chapter 5. In what follows we describe firm's optimal employment
policy in the presence of fixed adjustment costs within the framework of a structural
model, where rational employers have well-defined objective functions and base
their decisions on current information and their beliefs about the state of nature.
Our focus is on the dynamic implications of optimal employment rules: a decision
taken at the present time affects the evolution of the system and the employer
cannot select an action without taking into account its future consequences.

6.1. A structural model

As described in Section 5.1.1. the firm's objective is to find an optimal rule
for choosing whether to change or not the number of workers in each period t, for
each possible stock L, that may occur. In order to keep notation simple we consider
a single firm and hence omit the subscript i, i=1,...,N. We maintain:

assumption 1), discrete time;

assumption 2), additive separability over time of the reward function; and

assumption 3), sequential decision making;

as illustrated in the last chapter, Section 5.1. We also assume that the firm has
infinite time horizon and it only faces fixed adjustment costs

\[ C(L_{t-1}, L_t) = 1(L_{t-1} \neq L_t)K \]  \hspace{1cm} (147)

where \( K \) indicates fixed costs, which arise each time the company adjusts, and \( 1(L_{t-1} \neq L_t) \) is an indicator function such that

\[
1(L_{t-1} \neq L_t) = \begin{cases} 
0 & \text{if } L_{t-1} = L_t \\
1 & \text{if } L_{t-1} \neq L_t
\end{cases}
\]  \hspace{1cm} (148)

The two mutually exclusive alternatives - to adjust or not to adjust - characterizing the decision problem may be summarized by a dichotomous variable

\[ d_t = \begin{cases} 
0 & \rightarrow L_t = L_{t-1} \\
1 & \rightarrow L_t \neq L_{t-1}
\end{cases} \]  \hspace{1cm} (149)

where \( d_t \), the decision of whether or not to change employment, represents the control variable in the problem. The firm starts at period \( t \) with stock of workers \( L_{t-1} \). In each period there is a pay-off or current profit, \( \Pi_t \), associated with each choice. It chooses \( \{d_t\}_{t=0}^{\infty} \) sequentially to maximize the value of a discounted stream of profits:

\[
\max_\delta \left\{ \mathbb{E}_w \left( \sum_{t=0}^{\infty} \beta^t \Pi_t(L_t, \delta_t, \omega_t, \theta) \right) \right\} \]  \hspace{1cm} (150)

subject to the law of motion

\[ L_{t+1} = f[L_t, \delta_t(L_t, \theta), \omega_t] \]  \hspace{1cm} (151)

where \( \beta \), the discount factor, \( 0 < \beta < 1 \), is the rate at which the firm discounts pay-offs in future periods, \( f \) is a function which we will specify later and profits \( \Pi_t \) include fixed costs and are of the type defined in Chapter 3, as
\[ \Pi_t = AL_t^\alpha - 1(L_t \# L_{t-1})K - WL_t \]  

(152)

where \( L_t \) is the state variable, \( A \) is a constant incorporating technical progress and \( 0 < \alpha < 1 \).

The expectation in (150) is taken with respect to the random disturbance \( w_i \), which represents uncertainty with respect to future output prices and wages. This defines the structure of the exogenous shocks in the model. At the beginning of each period the output price and wages become known and the company modifies its entire horizon plans accordingly. It decides \( L_i=0 \) or \( L_i=1 \) and makes forecasts about future prices and wages. The existence of the stochastic term implies that the reward functional (150) should be specified probabilistically. We maintain assumption 4) from the last chapter, Section 5.1.1., which defines the disturbance term \( w_i \), on a countable set \( W_i \). This allows us to use a probability measure which maps each element of this space onto \([0,1]\).

The pay-offs and the stochastic processes influencing outcomes depend upon a vector of structural parameters, \( \theta \in \Theta \). The maximization is undertaken for all admissible policies, \( \delta(L_t,\theta) \). Optimization involves choosing \( L_i=\delta(L_t,\theta) \) for each \( t \) and possible values of \( L_t \). We maintain the same space notation for the state and the control variables as in the previous chapter i.e.

i) the state variable, the current stock of employees \( L_t \), is an element of the state space \( S_t \):

ii) the decision variable, \( L_t \), is an element of the decision space \( D_{L_t,\theta} \), a subset of \( D_t \) which depends on the current state \( L_t \). This means that the control is restricted to take values compatible with current \( L_t \). In other words, the policy \( \delta = \{\delta_0,\delta_1,\ldots\} \) where \( \delta \) maps states \( L_t \) into controls \( L_t=\delta_t(L_t,\theta) \) is admissible.

278
For any initial stock of workers $L_{t-1}$, the infinite horizon problem (150) corresponds to finding an admissible control law $\delta = \{\delta_0, \delta_1, \ldots\}$ that maximizes the value function $V(L_t | L_{t-1}, \theta)$

$$V_{\delta}(L_t | L_{t-1}; \theta) = E_{\theta} \left\{ \sum_{i=1}^{\infty} \beta^i \Pi_i [L_0 \delta_i (L_0 \theta), w_0 \theta] \right\}$$

(153)

where $L_t | L_{t-1}$ indicates that the company chooses the level of employment $L_t$, given last period stock of workers, $L_{t-1}$. Obviously, $L_t = L_{t-1}$ if the employer decides to leave the stock unchanged.

From the Principle of Optimality (Section 5.1.2.) we know that, given an initial state $L_{t-1}$, the optimal control law $\delta^*$ is one that maximizes expression (153) subject to the motion equation (151). If our problem takes the form of a Markov decision process (MDP) [see Section 5.1.8.] then we may write the valuation as

$$V^*(L_t | L_{t-1}; \theta) = \max_\delta \{ V_i(L_t | L_{t-1}, \theta) \}$$

(154)

$$= \max_\delta \left\{ \Pi_i (L_t | L_{t-1}, d_{t}, \theta) + \beta \sum_{j \in \delta_{t-1}} V_{t+1}(L_{t+1} | L_0 \theta) p_{t+1} [(L_{t+1} = j | L_t = i, d_t) | \theta] \right\}$$

where $p_{t+1}$ is the Markov transition probability for next period state, $L_{t+1} = j$, given current state and decision, $L_t = i, d_t$. $L_{t+1} | L_t$ indicates that firms choose $L_{t+1}$ given $L_t$. The transition probability represents the firm's subjective beliefs about uncertain future states characterized by price and wage uncertainty i.e. the determinants of the disturbance term $w_t$. The firms starts at period $t$ with a stock of workers $L_{t-1}$ and decides whether or not to change its initial stock to obtain either $L_t \neq L_{t-1}$ or $L_t = L_{t-1}$, respectively. Given its choice in period $t$, the outcome next period, $L_{t+1}$, is determined by the transition probability conditional on the choice made at $t$ and
the state variable $L_{t}$. Equation (154) has the Markov property according to which the decision rule depends on the past history of the process only via the current state, $L_{t}$. At each time, the optimal decision rule $\{d_{t}\}_{t=0}^{\infty}$, has the property that whatever the initial state and initial decision are, the remaining choices are also optimal. In terms of the decision tree shown in Chapter 4., Section 4.2., Figure 39, each node representing the state of the system is treated as starting point; from each node the decisions of whether or not to adjust must be optimal for the continuation of the process. The decision rule satisfies the following condition

$$\delta^{*}(L_{t}, \theta) = \arg\max_{d_{t}} \{V(L_{t} | L_{t-1}, \theta)\}$$

(155)

which corresponds to maximizing (154) (see Section 5.1.8.). The structural estimation problem is to use data on the behaviour of individual firms, decisions and relative states, to infer the underlying primitives $(\Pi, \beta, p)$ - the firm's profit function, the discount factor and the firm's beliefs as summarized by the transition probability - specified up to an unknown parameter vector $\theta$, and the implied decision rule $d_{t} = \delta(\cdot)$ that best fits the data. Structural estimation implies uncovering the stochastic process generating the observed decisions in the panel $\{d_{t}, L_{t}\}$ and the primitives, $(\Pi, \beta, p)$, which determine the process. For notational simplicity, we will omit from now on the symbol $\theta$, indicating structural parameters, but it is important to remember that all the following formulations are actually conditional on it.

If we were able to fully observe firms' states and decisions, i.e. if we had full information, in each period, on all companies, and if primitives were correctly specified we would essentially be able to predict the firm's behaviour at each point in time perfectly, given forecasts of the exogenous shocks. However, in practice, no
dataset is rich enough to fully measure all the characteristics of a firm. We suppose that the firm is assumed to observe the full state vector but the econometrician only observes certain components. This is basically a way of taking into account unobserved heterogeneity among companies. For example, while firms know their patterns of productivity, the econometrician cannot control for the fact that in certain periods productivity of certain firms goes up while that of others goes down and, in other periods, the opposite will happen. We thus make assumption 5) in last chapter, Section 5.1.4. In Section 5.1.5. we have shown that a problem with imperfect knowledge on some of the state variables may be reformulated into a problem with perfect information. This involves finding a set of sufficient statistics which allows us to consider only the information strictly necessary and known for our control purposes and make the problem only dependent on those. We start gradually to reformulate the model in terms of the sufficient statistics by explicitly considering the introduction of observation noise into the model.

The model is, thus, now characterized by two sources of disturbance which we assume uncorrelated:

1) the errors in forecasting future output prices and wages, which represent the exogenous shocks to the model and are captured by the transition probabilities which embody the main dynamics in the model; and

2) an error, $\epsilon_n$, resulting from imperfect knowledge of some of the features of the firms in the sample, relevant for control purposes. Manski (1977) lists a number of reasons for this random component: it reflects unobserved characteristics, unobserved taste variation and similar imperfections which force the analyst to treat the choice process as random. In what follows, we will concentrate on this second
source of random disturbance and explain its role in the optimization problem.

We denote the distribution of \( \epsilon_t \) as

\[
g_t(\epsilon_t | I_t)
\]

We assume that this type of error is choice specific

\[
\epsilon_t = [\epsilon^A_t, \epsilon^\text{NA}_t]
\]

where \( A \) indicates the choice of adjusting and \( \text{NA} \) indicates the choice of not adjusting. The current pay-off, \( \Pi_t \), may be, thus, partitioned into two components: a systematic component, \( \Pi_t'(L_t, \theta) \) and a random component, \( \epsilon_t \).

\[
\Pi_t = \Pi_t' + \epsilon_t(d_t) = \begin{cases} 
\Pi_t^A + \epsilon_t^A & \text{if } d_t = 1 \\
\Pi_t^\text{NA} + \epsilon_t^\text{NA} & \text{if } d_t = 0
\end{cases}
\]

Using this choice specific representation of profits and extending it to the forward component \( V_{t+1} \), we may write the decision rule as a

\[
d_t = d^*_t = \begin{cases} 
0 & \text{if } \epsilon_t^\text{NA} - \epsilon_t^A > (\Pi_t^A - \Pi_t^\text{NA}) + \beta E(V_{t+1}^A - V_{t+1}^\text{NA}) \\
1 & \text{if } \epsilon_t^\text{NA} - \epsilon_t^A < (\Pi_t^A - \Pi_t^\text{NA}) + \beta E(V_{t+1}^A - V_{t+1}^\text{NA})
\end{cases}
\]

From (159) it is clear that the optimal decision rule depends on the differences in expected lifetime pay-offs associated with the two alternative choices and not their absolute levels. The choice specific term relative to the future \( V_{t+1} \) indicates that a decision today, which yields current profits \( \Pi_t^A \) or \( \Pi_t^\text{NA} \) has also consequences for tomorrow, embodied in the choice specific value functions for next period, \( V_{t+1}^A \) and \( V_{t+1}^\text{NA} \). It should be noticed that the difference between the two valuation functions at \( t+1 \) \( (V_{t+1}^A - V_{t+1}^\text{NA}) \) is relative to have taken the decision of whether or not to
adjust at time t. For example, \( V_{t+1}^A \) indicates the valuation function at \( t+1 \) after having made the choice of adjusting at \( t \), indicated by superscript \( A \). Also, the forward terms \( V_{t+1}^A \) and \( V_{t+1}^{NA} \) are, among other things, functions of \( \epsilon_{t+1} \). The decision rule takes the form of a threshold rule, where it is possible to compare the rewards relative to each decision. For instance, the firm will decide to adjust if the difference between the choice specific stochastic terms \( (\epsilon_t^A-\epsilon_t^{NA}) \) is less than the difference between profits (current and expected) from those actions, and it will decide not to adjust if the reverse inequality holds. The presence of the choice specific random component, requires the use of choice probabilities as a representation of the choice process. The problem may be stated in the following way: given the choice between \( A \) and \( NA \), what is the probability that the employer selects, for instance, to adjust: \( A \)? In a static world, and hence not considering the \( t+1 \) component in (159), we would have

\[
Pr[d = 1] = Pr[\Pi^A > \Pi^{NA}] = Pr[\Pi^A + \epsilon^A > \Pi^{NA} + \epsilon^{NA}]
\]

\[
= Pr[\epsilon^{NA} - \epsilon^A < \Pi^A - \Pi^{NA}] = F(\Pi^A - \Pi^{NA})
\]

which is the static correspondent of (159). If the error terms \( \epsilon_i \) have a Type 1 extreme value distribution then the difference \( \epsilon^{NA} - \epsilon^A \) is distributed as logistic, and we indicate the logistic distribution function by \( F(\cdot) \).

Before proceeding, it is interesting to note from condition (159) that if the error term \( \epsilon_i \) were not different for each choice, \( d_i = 0 \) and \( d_i = 1 \), we could easily get rid of it by taking the difference between the two valuation functions made conditional on each decision (the right hand side of the two inequalities). In this case we would obtain a deterministic rule as in the case of perfect information. This explains why the error term has to be choice specific.
Even if the decision rule expressed in (159) or (160) explains much about the optimal decision rule, we have to take into account that our context is dynamic and, thus, we can not neglect the forward component $t+1$. This requires adaptation of the static logit model to become dynamic. In the presence of this error term the optimal control law (155) may be rewritten as

$$\delta^* (L, e, \theta) = \arg\max_{d} \left[ \Pi_t + e_t + \beta E(V_{t+1}) | I_t \right]$$

(161)

where, given the existence of this information error, we have made the decision rule explicitly conditional on the information set. The explicit consideration of dynamics introduces a substantial difficulty into the model since we have to admit the possibility of serial correlation in the error terms, $\{e_t\}_{t=0}^\infty$. This implies that the state space at $t$ also must include all of the $e$'s that are known at $t$ and that affect the distribution of $e_{t+1}$ with the consequence that the state space approaches an infinite dimension. The following section will deal with this problem.

This complication, however, does not affect our ability to use the choice probabilities as a representation of the preference relations characterizing the problem. We have shown in last chapter, how reformulation in terms of choice probabilities greatly simplifies the model and represents the most parsimonious way of finding the optimal control law. We have also shown how the choice probability is in fact a sufficient statistic summarizing all the information relevant for control purposes. As shown in last chapter, Section 5.1.7., if we assume that the probability distribution of the observation disturbance $e_{t+1}$ depends only on the immediately preceding state, the control and the system disturbance, $L_{n}, d_{n}, w_{n}$, and not on $L_{n-1}, d_{n-1}, w_{n-1}$ and $w_{n-1}...w_{0}$ (assumption 6), then we may refer to the conditional probability, $P_{t}$, as a sufficient statistic. $P_{t}$ is a conditional probability measure of the
state \( I_t \) given information at time \( t \), \( I_t \), and it is generated recursively in time

\[
P_{t+1}(I_{t+1} | I_t) = f[P_t(I_t | I_t) d_t \epsilon_{t+1}]
\]  

Equation (163) represents the new law of motion in the presence of observation errors. It describes the evolution of the system as the difference equation (151) did, but now it allows us to obtain a substantial reduction in the information we need for the control purposes. The decision law \( d_t \) is now

\[
d_t = \delta_t[P_t(I_t)]
\]  

Assumption 6) is natural within the Markov framework where it is unnecessary to keep track of the entire history of the system, and the transition probability, the reward functional and the valuation function depend only on the current state of the system and the current choice selected.

Hence the conditional probability may be expressed as

\[
P_t(I_t) = pr\left\{ \arg\max_{d_t} \left[ \Pi_t^{*} + \epsilon_t + \beta E(V_{t+1}) | I_t \right] \right\}
\]

\[
= \sum_{i} \left\{ d_t = \delta_t^{*}(L_t \epsilon_t) \right\}
\]  

it specifies the probability of choosing an alternative \( d_t \), given that an action must be decided from the choice set, given the characteristics of the decision maker. The crucial feature of the representation of the model based on choice probabilities is that the problem is made to depend only on the measured attributes of alternatives and individual characteristics. The recursive structure inherent in the problem is maintained and hence choice probabilities may be directly used as representation of the intertemporal choice problem.

As we have already mentioned, in this dynamic context, a complication is
given by the fact that the observation error $\epsilon_t$ enters as a conditioning variable in the transition probability, $p_{t+1}$, as we may see from rewriting the optimal Markovian decision rule as

$$\delta_t^*(I_t, \epsilon_t) = \arg\max_{d_t} \left\{ \Pi_t^*(L_t | L_{t-1}) + \epsilon_t(d_t) \right\}$$

where the valuation function dated $t+1$ in the second line of (165) may also be written in terms of the conditional probability as $V_{t+1}(P_{t+1} | L_t)$.

There are, here, two possibilities:

1) non-serially correlated error terms. In this case we may rewrite the optimal decision rule as

$$\delta_t^*(I_t, \epsilon_t) = \arg\max_{d_t} \left\{ \Pi_t^*(L_t | L_{t-1}) + \epsilon_t(d_t) \right\}$$

where the transition probability is no longer conditional on the error at time $t$, $\epsilon_t$.

The state space in non-explosive and remains finite and so we may rewrite the problem as

$$V(L_t, \epsilon_t) = \max_{d_t} \left\{ v_t(L_t, d_t) + \epsilon_t(d_t) \right\}$$

where

$$v_t(L_t | L_{t-1}) = \Pi_t^*(L_t | L_{t-1}, d_t)$$

$$+ \beta \sum_j \left\{ V_{t+1}(L_{t+1} | L_t, \epsilon_{t+1}) p_{t+1}(L_{t+1}, \epsilon_{t+1} | L_t) \right\}$$

where the joint probability $p_{t+1}$ may be decomposed into $r_{t+1}(L_{t+1} | L_t, \epsilon_{t+1}) g_{t+1}(\epsilon_{t+1} | L_t)$.

Formulation (167) shows that the decision rule for the dynamic choice problem has
the same form as the decision rule for the static choice problem (158) except that
the static profit function is replaced by the pseudo-value function \( v \). We may see,
but we will examine the issue in greater detail later, that if the error terms follow
a type I extreme value distribution, the model takes the form of a logit and we
obtain a dynamic generalization of the static logit model.

2) serially correlated error terms. Observation disturbances are correlated with their
preceding values. In this case the optimal decision rule is a non-separable function
of \( \epsilon \)'s. This complication has an immediate undesirable effect: it aggravates the
problem of dimensionality of the DP algorithm by generating a state space of
expanding dimensions. In this case, to obtain computational feasibility it is necessary
to assume "conditional independence" (Rust, 1988). In Section 5.1.7. we have
shown how the stochastic evolution of \( L \) and \( \epsilon \) may be simplified by using the
definition of conditional probability. In particular under conditional independence
we may write the transition probability

\[
p_{r+1}(L_{r+1}, \epsilon_{r+1}, L_p, \epsilon_t) = g_{r+1}(\epsilon_{r+1} | L_{r+1}, L_p, \epsilon_t)r_{r+1}(L_{r+1} | L_p, \epsilon_t)
\]  

(169)

as

\[
p_{r+1}(L_{r+1}, \epsilon_{r+1}, \epsilon_t) = r_{r+1}(L_{r+1} | L_r) g_{r+1}(\epsilon_{r+1} | L_{r+1}, L_r)
\]  

(170)

where \( r(\cdot) \) is the conditional and \( g(\cdot) \) is the marginal probability. As we may notice,
as a consequence of the assumption of conditional independence, the term \( \epsilon_t \) does
not appear as a conditioning element in either \( r_{r+1}(L_{r+1} | \cdot) \) or \( g_{r+1}(\epsilon_{r+1} | \cdot) \) in (170).
Through this assumption we limits the pattern of dependence in the \( \{L, \epsilon\} \) process
by assuming that any serial dependence between \( \epsilon_t \) and \( \epsilon_{r+1} \) is transmitted entirely
through the observed state $L_{s+1}$. $L_{s+1}$ is a sufficient statistic for $\epsilon_{s+1}$. Also, the probability density for $L_{s+1}$ depends only on last period stock, $L_s$ and not on the error $\epsilon_s$. In other words, the residuals from the regression of $\epsilon_{s+1}$ on $L_{s+1}$ and $L_s$ are serially independent.

Hence, under the assumption of conditional independence we may rewrite Bellman's equation as

$$V_t(L_n, \epsilon_t) = \max_{d_t} \left[ v_t(L_n, d_t) + \epsilon_t(d_t) \right]$$ (171)

where

$$v(L_t | L_{t-1}) = \Pi_t(L_t | L_{t-1}, d_t)$$

$$+ \beta \sum_j \left[ V_{t+1}(L_{t+1} | L_n, \epsilon_{t+1}) r_{t+1}(L_{t+1} = j | L_t = i, d_t) g_{t+1}(\epsilon_{t+1} | L_{t+1}, L_t) \right]$$ (172)

where $r(.)$ is the transition probability and $v(.)$ is the unique fixed point of a contraction mapping $T(v)(L)$ (see Chapter 5., Section 5.2.4.). The decision rule is now

$$\delta^*(L_n, \epsilon_t) = \arg\max_{d_t} \left[ v_t(L_n, d_t) + \epsilon(d_t) \right]$$ (173)

As before, we obtain the same basic structure of a static discrete choice problem. The optimization problem is now defined on the space $(L_n, d_t)$ smaller than $(L_n, \epsilon_t)$ on which $V(.)$ was defined. The value function $v(.)$ is, in fact, defined on the reduced space where $L_{s+1}$, after the decision is taken, contains all the necessary information.
6.2. Reduction of the model

The function $v_i$ may be computed by solving the contraction fixed point which involves using computationally intensive backward recursion methods. Hotz and Miller (1993) provide an alternative representation which avoids having to solve explicitly for $v_i$. Their approach is based on a new representation of the valuation function which is expressed in terms of future pay-offs, choice probabilities and probability transitions of choices and outcomes that remain feasible in future periods. In this Section we use the methodology suggested in Hotz and Miller's work and the results obtained in the previous section in order to reduce the model to an estimable form.

We start by expressing the value function $v(.)$ as the combination of the conditional value functions relative to the two mutually exclusive choices, weighted by the conditional probabilities

$$ v_i = P^A_i v^A_i + (1 - P^A_i) v^{NA}_i $$

(174)

where $v^A_i$ represents the value function conditional on having optimally chosen to adjust in period $t$ and $v^{NA}_i$ is the valuation function relative to optimally choosing not to adjust in period $t$. $P^A_i$ is the probability that the firm will optimally choose to adjust conditional on information at time $t$, while its complement $(1 - P^A_i) = P^{NA}_i$ is the conditional probability of non adjusting. Simple algebra allows us to express $v_i$ as

$$ v_i = v^A_i + (1 - P^A_i)(v^{NA}_i - v^A_i) $$

(175)

At time $t$, the conditional probability is given as
\[ P^A(d_i = 1 | I_t) \]  

(176)

where \( d_i^* \) indicates that the decision of to adjust is optimal.

The conditional choice probability \( P^A \) may also be represented in terms of the decision rule (160) and the optimal control law (161) as

\[
P^A(I_t) = \int_{\epsilon_i}^{e_i} \int_{e_i + \epsilon_i}^{e_i + \epsilon_i} \int_{e_i + \epsilon_i}^{e_i + \epsilon_i} \int_{e_i + \epsilon_i}^{e_i + \epsilon_i} dg(e^A, e^{NA} | I_t)
\]

(177)

where \( g^A(e^A | I_t) = \partial g(e_i | I_t) / \partial e_i \), which is also the formulation given in McFadden (1981) where (177) is used to generate choice probabilities from parametric families of probabilities, which give the distribution of tastes of the population of individuals with certain characteristics. The integrand in the last line of (177) is the probability density function for \( \epsilon_i^A \), given history \( I_t \) and the choice \( A \), to adjust, is optimal i.e.

\[
\left( v_i^A + \epsilon_i^A \right) \succeq \left( v_i^{NA} + \epsilon_i^{NA} \right)
\]

(178)

Condition (178) corresponds to (158) in the previous section after the reduction of the problem to the smaller space \((L_n, d_i)\) implied by (171).

Let \( \Delta v_i \) denote the difference between the conditional valuation functions associated with the two alternatives choices, \( v^A \) and \( v^{NA} \). The key result we exploit from Hotz and Miller (1993) is that this difference in the conditional valuation functions may be expressed as function of the conditional choice probabilities. In order to show this rewrite condition (178) in the following way
We can always express \( P^{\Delta}(\cdot) \) as a 1:1 mapping, \( Q(\cdot) \), from \( \Delta v_t(\cdot) \) (and \( I_0 \) since \( \partial g(\epsilon_t | I_t) / \partial \epsilon_t \) varies with \( I_t \)) defined from \( P \in (0,1) \) to \( (-\infty, \infty) \). Let \( G(\Delta \epsilon) \) the distribution function of \( \Delta \epsilon \) then

\[
\epsilon_t^{\Delta} - \epsilon_t^{\Delta} > \left( v_t^{\Delta} - v_t^{\Delta} \right)
\]

or

\[
\Delta \epsilon_t > -\Delta v_t
\]

Equation (180) may be illustrated by the following graphical representation:

\[
P^{\Delta} = 1 - G(-\Delta v) = Q(\Delta v)
\]

Equation (180) may be illustrated by the following graphical representation:

![Figure 42 Equation (180)](image)

\( Q(\Delta \epsilon) \) corresponds to the shaded area, defined by \(-\Delta v\), in the distribution \( G(\Delta \epsilon) \).

\( Q(\Delta \epsilon) \) is a continuous monotonic function of \( \Delta v \) so long as the density \( g(x) = [Dg(x)/dx] > 0 \) is non-zero over the entire support of \( G(\cdot) \). This rules out the possibility of points with zero density. In this case \( Q(\cdot) \) is invertible and we obtain Proposition 1 in Hotz and Miller, i.e.
Equation (181) allows us to express the difference in the two conditional valuation functions as function of the conditional choice probabilities.

A further simplification is obtained if we assume that \( \{e_i\} \) process, driven by idiosyncratic errors, follows an IID extreme value distribution. In this case, Inversion (181) reduces to a log-odds transformation which we indicate with \( q \). The conditional choice probability formula, thus, becomes

\[
Q_t^{-1} = \Delta v_t = (v_{t}^{NA} - v_{t}^{A})_t = \log \left[ \frac{1 - P_t^A}{P_t^A} \right] = q_t
\]  

(182)

We may express the difference \( v_{t}^{NA} - v_{t}^{A} \) by using the fixed point condition (172)

\[
v_{t}^{NA} - v_{t}^{A} = \{(\Pi_t^{NA} - \Pi_t^{A}) + E \beta [V_{t}^{NA} - V_{t}^{A}]_{k+1} | d_t=1, I_t \}\]

(183)

where the conditional valuation functions \( V_t \) are \( V_t[P_s(I,s)] \). Under conditional independence and additive separability, we may characterize the sequence of choices and state variables which would be feasible for the employer in future periods if she were to choose to adjust in period \( t \) by adopting the same formulation as in (182) and expressing the conditional valuation function in term of the future probabilities.

It follows from (175), (182) and (183) that equation (172) may be rewritten as
where the product between the conditional choice probability and the transition probability \( r_{t+1} \) gives the probability of \( L_4 \) conditional on \( I_4 \).

This representation may be viewed as the dynamic generalization of the static logit model (McFadden 1973) where the difference between the pay-offs associated with two choices is given by the log-odds ratio. By substituting a consistent estimate of the log-odds ratio we may compute \( v(.) \) avoiding the cost of recursively calculating the valuation function many times. Using this new expression of the valuation function we may now give a clearer representation to the firm’s optimal decision process. We write the conditional valuation function associated with optimally choosing to adjust in period \( t \) explicitly as the sum of the future pay-offs:

\[
v^A_t = \left( \Pi_t^A + E_2 \beta \left[ V^A_{t+1} + P^A_{t+1} \left( V^{NA}_{t+1} - V^{A}_{t+1} \right) \right] | L_{t+1} \right)_{t+1}^{d_t^* = 1}
\]

where \( A | A \) indicates that the company adjusts after having adjusted the period before and \( NA | A \) indicates that a non-adjustment period follows an adjustment. \( d_t^* = 1 \) indicates that the choice at \( t \) "adjust" was made optimally. We analyze equation (184) by considering its elements separately. First of all we know from (173) and (182) that the last component of the conditional valuation function may be written as
while \( v_t^{A|A} \) may be developed in the following way:

\[
V_{t+1}^{A|A} = \left\{ \beta \Pi_{t+1}^{A|A} (L_{t+1} \mid L_t) + E \beta^2 V_{t+2} (L_{t+2} \mid L_{t+1}) + \ldots \mid I_{t+1}, d_t^* = 1 \right\} 
\]  

which appears as a convenient formulation since the pay-off for periods greater than \( t+1 \) is no longer a function of \( L_t \) and cancels from the first order conditions, as shown below. Equation (184) thus may be rewritten as

\[
v_t^{A} (L_t \mid L_{t-1}) = \left\{ \Pi_t^{A} + E \left[ \beta \Pi_{t+1}^{A|A} (V_{t+1} + P_{t+1}^{NA|A} (V_{t+1}^{NA} - V_{t+1}^{A|A})) \right] \mid I_t, d_t^* = 1 \right\} 
\]

We may apply the same procedure to study the sequence of decisions starting with non adjusting as the optimal choice

\[
v_t^{NA} (L_t \mid L_{t-1}) = \left\{ \Pi_t^{NA} (L_t \mid L_{t-1}) + E \left[ \beta \left[ V_{t+1}^{NA} + P_{t+1}^{NA} (V_{t+1}^{NA} - V_{t+1}^{NA}) \right] \right] \mid I_t, d_t^* = 0 \right\}  
\]

Provided we can obtain consistent estimates of the future conditional choice probabilities we may estimate the model without having to compute the valuation function using backward recursion. This representation of conditional valuation functions has two important features: 1) it allows the estimation of the structural parameters of the model, through the estimation the conditional choice probabilities (see Hotz and Miller (1993) and Hotz et al. (1993) and 2) it allows us to form orthogonality conditions which can be directly exploited in the estimation of the conditional valuation function, as we will show in the next Chapter.
6.3. S,s-type policies

The two conditional valuation functions, $v_t^A$ and $v_t^{NA}$, represented in (187) and (188) help us to further study the optimal employment decision rule. Figure 43. shows these two functions drawn with respect to $L_{t-1}$. The value function relative to optimally choosing to adjust in period $t$ is a horizontal line, since it is no longer a function of $L_{t-1}$. Indeed, the firm decides to move from level $L_{t-1}$ to $L_t$, hence $v_t^A$ is only a function of the latter. The value function relative to optimally choosing not to vary employment in period $t$ is, instead, a function of $L_{t-1}$ since the firms decides to remain at the previous period's level of employment. We may draw $v_t^{NA}$ as a concave function, since DP implies that an optimal employment policy exists and it is the solution of the maximization problem. As we may see in the picture, the two functions intersect twice giving rise to two "reservation" values, $l^*$ and $l$, where
the firm is indifferent to whether or not to adjust.

\[ v_t^A = v_t^{NA} \rightarrow l^+, l^- \quad (190) \]

For values of \( L_{-1} \) less than \( l^* \) the firm will move on to \( v^* \), when \( L_{-1} \) is included between the two reservation values it will switch on to \( v^{NA} \) and after \( l^* \) the company will be again on the value function relative to adjusting. We may also represent the valuation function \( v_n \) which is given by the envelope as the thick line in the figure shows. Hence the optimal employment policy which results from this problem is the following threshold decision rule:

\[
L_t = \begin{cases} 
L^+ & \text{if } L_{t-1} \leq l^* \\
L_{t-1} & \text{if } l^* > L_{t-1} \geq l^- \\
L^- & \text{if } L_{t-1} \geq l^-
\end{cases} \quad (191)
\]

The optimal decision rule implies that if the stock of workers is below (above) the critical level \( l^* \) (\( l^- \)) the firm brings it up (down) to the desired level \( L^* \) (\( L^- \)); otherwise the firm will choose inaction. The effect of fixed costs, \( K \), is to induce the firm to hire or fire workers less frequently, but to hire or fire rather more when the firm does. Their effect is, thus, to create lumpy adjustments. \( L^* \) and \( L^- \), represent the target levels of employment when the firm decides that it is optimal to adjust respectively up or down. The reservation values, \( l^* \) and \( l^- \), arise through the existence of fixed costs and they are the cause of the discontinuities in employment changes generating a zone, \( l^* > L_{t} > l^- \), of no adjustment. Thus the firm’s optimal choice problem is to decide whether to adjust employment and bear the fixed costs at period \( t \) or whether to send the adjustment to the following period. If, for simplicity, we do not distinguish between the direction of employment changes, then the optimal policy for a \( t+s \) period problem is given by the sequence of critical
levels \((l_0, L_i), (l_{i+1}, L_{i+1}), \ldots, (l_{n+n}, L_{n+n})\).

Notice how this optimal employment decision resembles the inventory strategy called the "S,s" rule, also known as "two bin policy", according to which every time inventories fall below some minimum level, s, a purchase restores them to their maximum level S. Fixed costs bring about this trigger point s. In this case, it pays for a company to place fewer orders, make each order larger, and store more inventories than it would if fixed costs did not exist. Scarf (1960) first showed the optimality of multiperiod (S,s) policies. In our case the policy becomes slightly more complicated owing to the existence of two critical values and two directions of changes as shown in (186).
Appendix 6.1. The Logit model

Let the probability that an individual will make a certain choice, given knowledge of attribute $X_j$, be $P_i$

$$P_i = F(\beta' X_i) = \frac{1}{1 + e^{-\beta' X_i}}$$  \hspace{1cm} (A1)

where $F(.)$ takes the form of a logistic. We may multiply both side of the equation by $(1 + e^{\beta' X_i})$ to get

$$\left(1 + e^{-\beta' X_i}\right) P_i = 1$$  \hspace{1cm} (A2)

Dividing by $P_i$ and subtracting 1 leads to

$$e^{-\beta' X_i} = \frac{1}{P_i} - 1 = \frac{1 - P_i}{P_i}$$  \hspace{1cm} (A3)

or

$$e^{\beta' X_i} = \frac{P_i}{1 - P_i}$$  \hspace{1cm} (A4)

Taking the natural logarithm on both sides we have

$$\beta' X_i = \ln\left(\frac{P_i}{1 - P_i}\right)$$  \hspace{1cm} (A5)

The dependent variable in the regression is simply the logarithm of the odds that a particular choice will be made.
Chapter 7. FIXED AND CONVEX ADJUSTMENT COSTS: CAN WE RECOVER THE EULER EQUATIONS?

Introduction

7.1. First order conditions

7.2. Stochastic specification and the choice of estimation method

7.3. Estimation of the conditional choice probabilities

7.4. Estimation of the conditional Euler equation

Appendix 7.1. Euler equation

Appendix 7.2. Logit estimates

Appendix 7.3. Calculation of short and long run elasticities
Introduction

In Chapter 3, we considered a model of dynamic labour demand based on the assumption of strictly convex adjustment costs and in Chapter 6, we analyzed the case in which the costs of changing employment are fixed. In this chapter we include both structures of adjustment costs and attempt to derive a more general model of dynamic labour demand which encompasses the Euler equation and the discrete decision process generated by the presence of fixed costs. We proceed by including variable adjustment costs in the form of a quadratic function in the employment change, in the theoretical framework of the discrete decision process analyzed in last chapter. We maintain the structure of a DDP since, even with quadratic costs, the presence of fixed adjustment costs imposes a discrete choice: the company must decide whether or not it is profitable to start to adjust and bear the fixed costs. This decision of whether or not to pay the fixed costs has to be taken every time the firm considers the possibility of changing employment. The presence of quadratic costs implies that once the employer decides to vary the labour input, adjustment takes place in a smooth way as predicted by the Euler equation. The adjustment path generated by choosing to adjust at time \( t \) or to postpone the adjustment may be, thus, summarized in the following way:

i) because of fixed costs, the employer optimally chooses not to change employment at \( t \) and postpone the decision of whether or not to adjust to subsequent periods.

ii) the employer optimally decides to vary the labour input in period \( t \) and to pay fixed costs. Adjustment takes place, and conditionally on having chosen to change
employment, follows an adjustment path described by the Euler equation.

In both cases, a decision of whether or not to adjust and pay fixed costs in any other subsequent period must be taken optimally. As we will illustrate in more detail in the next pages, the interesting implication of the assumption of both fixed and quadratic adjustment costs is that, in principle, it allows us to recover the Euler first order conditions. Nevertheless, given the hypotheses which characterize the discrete decision process described in last chapter and which we will consider later, the model which results is more restrictive than the Euler equation generated in the case in which there are only quadratic costs.

7.1. First order conditions

In the presence of both quadratic and fixed adjustment costs we have the following cost structure

\[ C(L_{t-1}, L_t) = \gamma \left( \Delta L_t \right)^2 + \begin{cases} 0 & \text{if } L_{t-1} = L_t \\ K & \text{if } L_{t-1} \neq L_t \end{cases} \]  

The first component of (192) represents quadratic costs, \( \frac{1}{2} \gamma (\Delta L_t)^2 \), while the second part relates to fixed costs, equal to K if adjustment occurs and equal to zero otherwise. The decision problem is the same as in the previous section: at each period the firm decides whether or not to adjust and pay the fixed costs. As before, the presence of fixed costs implies that, in certain periods, the company may not consider it worth while to adjust. This gives rise to the dichotomous structure we have already illustrated in past sections.

Assume that the optimal decision at time \( t \) is to adjust. Hence, conditional
on this initial choice, we may write the (conditional) valuation function as (188) in Section 6.2.,

\[
\nu_t^A(L_t | L_{t-1}) = \left\{ \prod_t^A(L_t | L_{t-1}) + E \beta \left[ V_{t+1}^A(L_{t+1} | L_t) + \frac{P_{t+1}^{IA}(L_{t+1}^o \cdot \cdot \cdot) q_t(L_o L_{t+1} \cdot \cdot \cdot) }{I_o d_t^* = 1} \right] \right\} | I_o d_t^* = 1
\]

\[
= \left\{ \prod_t^A(L_t | L_{t-1}) + E \beta \prod_t^A(L_t | L_{t-1}) + \beta^2 V_{t+2}(L_{t+2} | L_{t+1}) + \beta \left( P_{t+1}^{IA}(L_{t+1}^o \cdot \cdot \cdot) q_t(L_o L_{t+1} \cdot \cdot \cdot) + \cdots \right) | I_o d_t^* = 1 \right\}
\]

(193)

It is important to notice for the subsequent results that both the conditional choice probability \( P_{t+1}^{IA} \) and the log-odds ratio are functions of \( L_o \) among other variables, since the optimal choice at time \( t \) implies that the firm chooses to adjust from \( L_{o.t-1} \) to \( L_o \). Equation (193) represents the left-hand side of the decision tree in Section 4.2. drawn in the following figure.
In period $t$ the employer chooses to adjust, but because of fixed costs, in subsequent periods he will face the decision of whether to continue the adjustment or to stop. As previously, the conditional choice probability, $p_{NA/A}^{NA}$ - or equivalently $(1-p_{A/A})$ - summarizes the expected future value of a chosen action.

If we differentiate the conditional value function (193) with respect to $L$, we obtain the Euler representation for this optimization problem. In fact, by conditioning on the optimal choice $A$, i.e. to adjust in period $t$, and given the
presence of quadratic costs in the cost function (192), we have made expression (193) continuous and differentiable with respect to \( L \). However, because of fixed costs which must be paid every time the firm decides to adjust, discontinuities may occur in the subsequent stages.

As we may notice from (193) the pay-offs relating to periods beyond \( t+1 \) are no longer a function of \( L \). To show this we rewrite the part of the conditional valuation function relative to \( t+1 \)

\[
\nu_{t+1} = \left[ \beta \Pi_{t+1}^{\lambda | A}(L_{t+1} | L_t) + E \beta^2 V_{t+2}(L_{t+2} | L_{t+1}) + \beta P_{t+1}^{\lambda | A} q_{t+1} \right] | I_{t+1} = 1 \] (194)

where \( q_t = \ln[(1-P_t)/P_t^{\lambda}] \) is the log-odd ratio discussed in detail in the last chapter. Clearly the component dated \( t+2 \) disappears when we take the first order conditions with respect to \( L \). By contrast, profits at \( t+1 \) are still a function of \( L \) given the quadratic adjustment cost term which implies that \( C_{t+1} = \frac{1}{2} \gamma (L_{t+1} - L_t)^2 \).

The first order conditions are given in the following expression

\[
\frac{\partial \nu_t^A}{\partial L_t} = \frac{\partial \Pi_t^A}{\partial L_t} + \beta E \left[ \frac{\partial \Pi_{t+1}^{\lambda | A}}{\partial L_t} + \frac{\partial P_{t+1}^{\lambda | A}}{\partial L_t} q_{t+1} + P_{t+1}^{\lambda | A} \frac{\partial q_{t+1}}{\partial L_t} \right] = 0 \] (195)

Equation (195) gives the (conditional) Euler equation for the maximization problem, in terms of the current and future pay-offs and of the future conditional choice probabilities. We may, thus, use this new representation of the Euler equation to provide orthogonality conditions which we may directly exploit in the estimation of the conditional labour demand function. We maintain the same profit function as in Section 3.1., i.e. a Cobb-Douglas production function in terms of labour. We rewrite the current profit function including fixed costs \( K \).
\[ \Pi_t = AL_t^* - \frac{\gamma}{2} (\Delta L_t)^2 - K - WL_t \] 

(196)

where W is the wage. By adopting the same method as in Chapter 3, we may obtain the following orthogonality conditions (see Appendix 7.1. for the algebra)

\[ \gamma L_{t+1} - \gamma (1 + \beta) L_t + \beta (pq)_{t+1} = 0 \] 

(192)

If we solve the newly represented Euler equation with respect to \( L_t \), we obtain the following estimable form which is analogous to (26) in Section 3.1 except for the term \( pq \)

\[ L_t = \psi_0 + \psi_1 L_{t+1} + \psi_2 L_{t-1} + \psi_3 \left( \frac{Q}{L} \right)_t + \psi_4 W_t + \psi_5 (pq)_{t+1} + \nu_{t+1} \] 

(198)

where \( \nu_{t+1} \) is a realization error and

\[
(pq)_{t+1} = \left[ \frac{\partial q_{t+1}}{\partial L_t} a_{t+1} + \frac{p_{t+1}}{\partial L_t} \right]
\]

(199)

The structural parameters are

\[
\begin{align*}
\psi_1 &= \frac{\beta}{1 + \beta} \\
\psi_2 &= \frac{1}{1 + \beta} \\
\psi_3 &= \frac{\alpha}{\gamma (1 + \beta)} \\
\psi_4 &= -\frac{1}{\gamma (1 + \beta)} \\
\psi_5 &= \frac{\beta}{\gamma (1 + \beta)}
\end{align*}
\]

(200)
The theory predicts that $0 < \psi_1, 0 < \psi_2, \psi_3 > 0, \psi_4 < 0$ and $\psi_5 > 0$. \footnote{See the discussion on the restrictiveness of this parameter formulation and on the possibility of a more general adjustment costs structure in Section 3.1, Chapter 3.}

The term $(pq)_{t+1}$ captures the alternatives of adjustment or non-adjustment that firms face in the future, due to the presence of fixed adjustment costs, as illustrated in Figure 44. Indeed, even if we condition on optimally adjusting at time $t$, and we are able to obtain a new representation of the Euler first order conditions, the presence of fixed costs imposes a choice in each period between these two alternatives. A comparison with the model presented in Chapter 3, using Figure 44 helps to illustrate the role of the term $(pq)_{t+1}$. With only quadratic adjustment costs the problem would be represented by the extreme left-hand side of the decision tree, i.e. the path only formed by the segments labelled A. This was formally shown in Chapter 3, where the firm decided to adjust in a continuous and gradual manner. The term $(pq)_{t+1}$ captures the fact that in the presence of fixed costs the firm faces the alternative choice of whether or not to adjust. This term takes into account all the branches of the decision tree (in solid lines), namely all the future optimal choices the company faces after adjusting at $t$.

7.2. Stochastic specification and the choice of estimation method

The theory characterizing the discrete decision process provides a clear definition of the sources of stochastic disturbance in the model. First of all, the error term in the Euler equation with fixed and quadratic costs, $\nu_{t+1}$, reflects a realization error due to the use of actual values of $L_{t+1}$ as an approximation to
E(L_{t+1}). The treatment of this disturbance term as purely innovational implies that this error is only correlated with the variables dated t+1 or later and, if this were the only source of disturbance in the model, would suggest that we use as instruments all the available history relating to period t and earlier in the levels equation.

In Chapter 6, we also considered the possibility of observation errors arising from unobserved heterogeneity. In this formulation, errors due to imperfect knowledge on some state variables on the part of the econometrician, enter the model only through the term \((pq)_{t+1}\). We have assumed that this error is distributed according to a type I extreme value distribution. Hence, at least in principle, this rules out the possibility of fixed effects in the empirical equation. In the previous chapter we also discussed the problem of serial correlation of this observation error. This entailed two possibilities: i) non serially correlated error terms, in which case the error term at t, \(\epsilon_t\), does not enter the conditional probability \(p_{t+1}\), and which allowed us to separate the valuation function into two components: the immediate choice and the continuation of the decision process; ii) serially correlated error terms, in which case the state space at t must also include all the \(\epsilon\)'s that are known at t and that affect the distribution of \(\epsilon_{t+1}\), with the consequence that the state space approaches infinite dimension. In the last chapter, we limited the pattern of dependence in the \(\{L_t, \epsilon_t\}\) process to deal with this problem by assuming that any serial dependence between \(\epsilon_t\) and \(\epsilon_{t+1}\) is transmitted entirely through the observed state \(L_{t+1}\) which acts as a sufficient statistic for \(\epsilon_{t+1}\). We called this assumption 'conditional independence'. As a consequence, the probability density function of \(L_{t+1}\) depends only on last period's stock \(L_t\) and not on the error \(\epsilon_t\).
If this assumption is valid, it will follow that there are no fixed effects and that the composite error term ($\eta_t$ in Chapter 3., Section 3.2, equation (35)) is also serially independent and that instruments dated $t$ and earlier remain valid. However, if conditional independence is not a valid assumption, estimation using this set of instruments will be inconsistent, and it is therefore important to test for instrument validity relative to an equation which will be consistently estimated under a wider range of circumstances (even if the coefficients of this equation lack structural interpretation).

The empirical equation we propose to estimate is therefore

$$L_u = \psi_0 + \psi_1 L_{u+1} + \psi_2 L_{u-1} + \psi_3 \left( \frac{Q}{L} \right)_u + \psi_4 W_u + \psi_5 (pq)_{u+1} + u_u + v_{u+1} \quad (201)$$

We directly exploit the new Euler equation (201) to provide orthogonality conditions and apply the Generalized Method of Moments Estimator.

There are important differences with respect to the structure of the Euler equation we obtained under the hypothesis of only quadratic adjustment costs in Chapter 3: i) we have conditioned on companies choosing optimally to adjust at time $t$, considered their conditional valuation function and obtained the "conditional Euler equation". As discussed earlier, this is only a part of a more complicated intertemporal decision problem which includes also the possibility of optimally deciding not to adjust at $t$; ii) equations (198) and (201) show the presence of the term $(pq)_{u+1}$ which is the result of taking into account the company's employment policy over time in the presence of fixed costs. As already mentioned, with both fixed and quadratic costs
the employer faces the alternative of whether to adjust or not in each period. The term \((pq)_{t+1}\) captures the fact that due to fixed costs some of the companies which optimally decided to adjust at time \(t\) may optimally choose not to vary employment in some of the subsequent periods. Estimation of the new Euler equation requires an initial estimate of this term which will be the objective of next section;

iii) the model is more restrictive than the Euler equation derived in Chapter 3. In principle, it also rules out fixed effects, implying that unobserved heterogeneity comes through the \((pq)_{t+1}\) term. It rules out serial correlation in the error term through the hypothesis of conditional independence. [see Chapter 6, Section 6.1.].

7.3. Estimation of the conditional choice probabilities

In order to estimate the term \((pq)_{t+1}\), we maintain the parametric assumption which has driven the theoretical model: that the disturbance term, \(\epsilon_n\), follows a type I extreme value distribution. As shown in the previous chapter this assumption allows us to express the discrete choice problem in terms of the difference between the pay-offs associated with the two alternatives, to adjust or not to adjust, in our problem. This leads to a logit specification of the choice problem in order to construct the selectivity term \((pq)_{t+1}\). Earlier authors have estimated the conditional choice probabilities non-parametrically (Altug and Miller, 1990; Hotz and Miller, 1993; Hotz, Miller, Sanders and Smith, 1994).

To estimate conditional choice probabilities we set the dependent variable equal to 1 if there is either positive or negative adjustment and equal to 0 if no adjustment occurs. Regressors are given by the set of instruments which will be
employed in the GMM estimation of the Euler equation: employment, L, real sales, Q, wage-bill, WB, investment in machinery, INV, profitability, P and size, S. They all enter the logit lagged one period. Logits are run for each year in order to maintain the time series variability in the construction of the (pq) term. Specifically, we run for each year the following logit

\[ Y_t = \frac{e^{X_t'\delta}}{1 + e^{X_t'\delta}} \]  

(202)

where \( Y_t \) is equal to 1 if the firm changes employment either upward or downward and it is equal to zero otherwise. The vector of regressors \( X \) is given by

\[ X_{t-1} = [L_{t-1}, Q_{t-1}, WB_{t-1}, INV_{t-1}, P_{t-1}, S_{t-1}] \]  

(203)

is the vector of regressors. The results from logits are shown in Appendix 7.2. 47

From each of these year-by-year logits, estimated over the entire set of companies, we may construct the probability \( p^{maj} \), the log-odds ratio \( q \), and their derivatives with respect to \( L_{t-1} \). We have that,

---

47 We also ran logits adding the squares and cross-products of the \( X \) variables. The pq ratio determined on the basis of these logits did not show relevant differences from the one calculated on the basis of the logits reported in the appendix.
\[
\hat{p}_u = \frac{e^{x_{u-1}^t \delta_t}}{1 + e^{x_{u-1}^t \delta_t}}
\]
\[
\hat{q}_u = \ln \left[ \frac{\hat{p}_u}{1 - \hat{p}_u} \right] = x_{u-1}^t \delta_t
\]
\[
\frac{\partial \hat{p}_u}{\partial L_{t-1}} = \delta_t^L (1 - \hat{p}_u) \hat{p}_u
\]
\[
\frac{\partial \hat{q}_u}{\partial L_{t-1}} = \delta_t^L
\]

where \( \delta_t^L \) is the coefficient of the employment variable, \( L_{t-1} \), in the logit. \( \delta \) has the subscript \( t \) since it varies over years. As already mentioned, this results from running yearly logits, obtaining the expressions for \( p \) and \( q \) and their derivatives and reorganizing them into a time-series and cross-section. If we take all the four components in (204) one period ahead we are able to calculate the term \((pq)_{t+1}\) as
it appears in the new empirical formulation (198).  

\[(pq)_{t+1} = \left[ \frac{\partial \hat{q}_{t+1}}{\partial L_t} \hat{q}_{t+1} + \hat{p}_{t+1} \frac{\partial \hat{q}_{t+1}}{\partial L_t} \right] \quad (205)\]

In the following figures we have plotted \((pq)_{t}\) and \(p^\alpha\), the probability of adjusting averaged across companies, over the sample period calculated from the logits shown above.

\[\text{\textsuperscript{48}}\text{ The pq ratio is constructed in the following way: from (199) we have that} \]

\[\hat{p}_t = \frac{e^{X_{it}^t \delta_t}}{1 + e^{X_{it}^t \delta_t}} = \frac{1}{1 + e^{-X_{it}^t \delta_t}} \]

\[\text{hence} \]

\[\frac{\partial \hat{p}}{\partial L_t} = \frac{\delta_t^L e^{-X_{it}^t \delta_t^L}}{(1 + e^{-X_{it}^t \delta_t^L})^2} = \delta_t^L \hat{p}(1 - \hat{p})\]

Again from (199) we have:

\[\frac{\partial \hat{q}}{\partial L_t} = \delta_t\]

Hence pq is given by

\[\delta_t^L (1 - \hat{p}_t) \hat{p}_t \hat{q}_t + \hat{p}_t \hat{q}_t \hat{p}_t \]

\[312\]
Figure 44 Plot of pq
In the next section we include pq term as regressor in the Euler equation where it represents the dynamic selection which arises from the presence of fixed costs in the dynamic labour demand model.

7.4. Estimation of the conditional Euler equation

We may now estimate the labour demand equation determined by the Euler
first order conditions. We adopt the same method of selection presented in Chapter 3., Section 3.6.: we keep or drop the observation on \( L_t \) according to whether it is different from or equal to \( L_{t-1} \). This allows us to keep only the companies with consecutive adjustment periods. This selection procedure is explained in detail in Section 3.6. The parameters of equation (201) are assumed weakly exogenous with respect to \( L_t \) and \( L_{t-1} \). In this case, selecting firms which have adjusted should not affect the error term in the levels equation (201). In other words, on the null hypothesis of a correctly specified model, the selection process is not endogenous and we do not need to add any correction for selectivity.

As in the case of only quadratic adjustment costs we use as instruments: employment, \( L_t \), output \( Q \), the wage-bill \( WB \), investment in machinery \( INV \), profitability \( P \) and firm's size \( S \) which are also the determinants of the new term, \( pq_{t+1} \). This term enters as an instrument at \( t+1 \). Since \( pq_{t+1} \) is a function of variables dated \( t \), its validity as an instrument relies on the absence of additional sources of disturbance in the model.

We estimate two relationships:

i) model (201), estimated in levels, which supposes the absence of fixed effects, since unobserved heterogeneity is captured by the term \( pq \); and the absence of endogenous selection, since model (201) is conditioned on \( L_{t-3} \), the initial stock of workers. In fact, we estimate the levels equation (201) with two different sets of instruments: first using instruments dated \( t,(t-1),(t-2) \) and \( (t-3) \) which assumes that

\[ \text{We employ the program discussed in Chapter 3. and Appendix 3.5, specifically written in order to take into account this selection procedure in the GMM estimation.} \]
the error term is purely innovational; and second using only instruments dated (t-1), (t-2) and (t-3) as instruments, which allows for the possibility of errors dated t [see Section 7.2.]. Note that in this case the variable \( pq_{t+1} \) is no longer available as an instrument.

ii) in order to test for fixed effects and selection bias, which should be absent if the model (201) is correctly specified, we consider the following generalization

\[
L_u = \psi_0 + \psi_1 L_{u+1} + \psi_2 L_{u-1} + \psi_3 \left( \frac{Q}{L_{u}} \right) + \psi_4 W_u + \psi_5 (pq)_{t+1} + \psi_6 \lambda_u + f_t + u_t + v_{t+1} \tag{206}
\]

where \( \lambda_u \) is the inverse Mills' ratio calculated according to the timing of the instruments as explained in Chapter 3., Section 3.5, and \( f_t \) denotes fixed effect. Even if, in principle, fixed effects do not enter (201) because unobserved heterogeneity comes only through the term \( pq \), they could in fact be generated by errors in the measurement of wages. This implies that we should allow for the possibility of fixed effects by estimating (206) in differences using variables dated (t-2) and (t-3) as instruments. In fact, when we estimate the model in first differences the error terms become \( \Delta v_{t+1} = v_{t+1} - v_t \) and \( \Delta u_t = u_t - u_{t-1} \) and this suggests that we use as instruments all the available history starting from t-2 backwards for the difference equation. Hence model (206) represents the alternative hypothesis where we suppose that there might be fixed effects and selection bias.

Results using generalized method of moments estimation are shown in the following tables.
Table XXXIII Euler equation with quadratic and fixed costs

<table>
<thead>
<tr>
<th>Estimation in:</th>
<th>LEVELS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timing of the instruments:</td>
<td></td>
</tr>
<tr>
<td>t(t-1)(t-2)(t-3)</td>
<td>(t-1)(t-2)(t-3)</td>
</tr>
<tr>
<td>const.</td>
<td>0.139</td>
</tr>
<tr>
<td></td>
<td>(0.063)</td>
</tr>
<tr>
<td></td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
</tr>
<tr>
<td>L_{t+1}</td>
<td>0.887</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
</tr>
<tr>
<td></td>
<td>0.553</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
</tr>
<tr>
<td>L_{t-1}</td>
<td>0.120</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
</tr>
<tr>
<td></td>
<td>0.443</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
</tr>
<tr>
<td>(Q/L)_t</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
</tr>
<tr>
<td></td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
</tr>
<tr>
<td>W_t</td>
<td>-0.163</td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
</tr>
<tr>
<td></td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
</tr>
<tr>
<td>(pq)_{t+1}</td>
<td>-0.013</td>
</tr>
<tr>
<td></td>
<td>(0.063)</td>
</tr>
<tr>
<td></td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>(0.060)</td>
</tr>
<tr>
<td>no. obs.</td>
<td>15,186</td>
</tr>
<tr>
<td>std. error of est.</td>
<td>0.180</td>
</tr>
<tr>
<td>first order serial corr.</td>
<td>-0.076</td>
</tr>
<tr>
<td>Sargan test</td>
<td>2089.8</td>
</tr>
<tr>
<td>Probab. value</td>
<td>0.000</td>
</tr>
</tbody>
</table>

1. Heteroscedastic consistent standard errors in parentheses.
2. The Sargan tests are tests for the validity of instruments with degrees of freedom given in brackets.
Table XXXIII shows the estimates of the Euler equation, and the \((pq)_{t+1}\) ratio, which takes into account the existence of fixed costs and of choice process implied by these costs. As already described, the two equations reported differ with respect to timing of the instruments. In the first equation, the set of instrumental variables start at \(t\) and exploits the all the available history up to \(t-3\) (recall that \((pq)_{t+1}\) is a function of variables dated \(t\)). In the second equation, the timing of the instruments, which now exclude \((pq)_{t+1}\), starts at \(t-1\) and goes back to \(t-3\). This last equation has the same structure as the model in levels labelled \((t-1)(t-2)(t-3)\) in the estimation of the Euler equation in Chapter 3, Table XXVII.

In the first model, characterized by purely innovational errors, absence of selection bias and fixed effects, the estimated coefficient on the wage variable appears well-determined and is highly significant. However, the coefficients on both productivity variable and \(pq\) are small and insignificant. By contrast, the equation estimated with instruments dated \(t-1\) and earlier is very close to the Euler equation reported in Table XXVIII, Chapter 3. In particular, the coefficient on the wage variable is positive but insignificant.

Table XXXIV reports the characteristic roots of the two models [see equation (25) in Chapter 3, Section 3.1]. Although, the \((t-1)(t-2)(t-3)\) equation has roots either side of unity implying that a forward stable optimal solution is obtained, the \(t(t-1)(t-2)(t-3)\) model has a near

<table>
<thead>
<tr>
<th>Table XXXIV Characteristic roots</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Levels</strong></td>
</tr>
<tr>
<td>(t(t-1)(t-2)(t-3))    ( (t-1)(t-2)(t-3) )</td>
</tr>
<tr>
<td>(\mu_1)   0.991     1.032</td>
</tr>
<tr>
<td>(\mu_2)   0.136     0.776</td>
</tr>
</tbody>
</table>
unit root together with a much smaller root. This suggests that these first set of estimates should be interpreted in terms of a purely backward looking relationship in \( \Delta L_{t+1} \), rather than the Euler equation (193).

The Sargan test of the overidentifying restrictions rejects the validity of the instruments for these equations.

The estimates of the alternative differenced model, given by (206), are reported in the following table.
Table XXXV Alternative model

Dependent variable: $L_t$

<table>
<thead>
<tr>
<th>Estimation in:</th>
<th>DIFFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timing of the instruments:</td>
<td>(a)</td>
</tr>
<tr>
<td>(t-2)(t-3)</td>
<td>(t-2)(t-3)</td>
</tr>
<tr>
<td>$L_{t+1}$</td>
<td>0.557</td>
</tr>
<tr>
<td>(0.193)</td>
<td>(0.193)</td>
</tr>
<tr>
<td>$L_{t+1}$</td>
<td>0.419</td>
</tr>
<tr>
<td>(0.161)</td>
<td>(0.173)</td>
</tr>
<tr>
<td>$(Q/L)_t$</td>
<td>0.036</td>
</tr>
<tr>
<td>(0.066)</td>
<td>(0.074)</td>
</tr>
<tr>
<td>$W_t$</td>
<td>-0.014</td>
</tr>
<tr>
<td>(0.203)</td>
<td>(0.224)</td>
</tr>
<tr>
<td>$(pq)_{t+1}$</td>
<td>-0.342</td>
</tr>
<tr>
<td>(1.374)</td>
<td>(1.684)</td>
</tr>
<tr>
<td>$\lambda_t$</td>
<td>0.017</td>
</tr>
<tr>
<td>(0.014)</td>
<td></td>
</tr>
<tr>
<td>no. obs.</td>
<td>8,319</td>
</tr>
<tr>
<td>std. error of est.</td>
<td>0.270</td>
</tr>
<tr>
<td>first order serial corr.</td>
<td>-0.667</td>
</tr>
<tr>
<td>Sargan test</td>
<td>53.05</td>
</tr>
<tr>
<td>Probab. value</td>
<td>0.784</td>
</tr>
</tbody>
</table>

1. Heteroscedastic consistent standard errors in parentheses.
2. The Sargan tests are tests for the validity of instruments with degrees of freedom given in brackets.
Model (a) includes the inverse Mills' ratio, \( \lambda \), (also as an instrument dated \( t \)) which is not significant, confirming the absence of endogenous selection, as implied by equation (201). In model (b) we have omitted the inverse Mills' ratio.

Again, the results are very similar to the Euler model estimated in Section 3.6. and the same conclusions may be drawn in this context. The coefficient on wages is insignificant in each set of estimates. The coefficient on \( pq \) is only correctly signed in the presence of the inverse Mills' ratio and we therefore take that set of estimates as being in greater conformity with the theoretical model despite the statistical insignificance of the coefficient on the inverse Mills' ratio.

In both specifications the dominant characteristic of the model is that the forward and the lagged employment terms, \( L_{t+1} \) and \( L_{t} \), exhibit high and significant coefficients while, on the contrary, the coefficients associated with real wages and productivity, \( W \), and \( (Q/L) \), are small and insignificant. The effect of these two last variables is clearly dominated by the dynamics. Employers respond to shocks on wages and productivity by varying the number of workers only if these shocks are very large. Because of high adjustment costs employers may be cautious in adjusting employment unless shocks to wages and productivity are very large. If we look at the coefficients of wages and productivity in (200) we may see that if \( 0 < \alpha \beta < 1 \), and \( \gamma \), the parameter in the adjustment costs specification, is very high then both \( \psi \), and \( \psi_{w} \), the parameters associated with productivity and wages respectively, will be very small. In the extreme case where \( \gamma \) approaches infinity, \( \psi \), and \( \psi_{w} \), will be close to zero. Hence, we either obtain no adjustment as the result of the size of adjustment costs and the lack of large shocks to wages and productivity over the period, or a very slow adjustment as the consequence of the cautiousness of
employers in varying the number of workers.

In both the differenced equations the Sargan test for the overidentifying restrictions, given at the foot of the table, fails to reject the null hypothesis of the orthogonality of the instruments. [The formula for the Sargan test is given by (61) in Chapter 3., Section 3.6.] This confirms the need to take into account the presence of fixed effects.

In both the differenced equations (a) and (b), the second root falls inside the unit circle confirming that a forward stable optimal solution is obtained in both cases.

The conditional independence assumption can be tested by the Hausman test. In fact, we consider two nested Hausman tests. The first takes as the alternative hypothesis the set of estimates reported as (a) in Table XXXV (i.e. including both the inverse Mills’ ratio and the \( p_{i,t+1} \) term), and compares that with a levels equation containing these variables but estimated using instruments variables dated \( t-1 \) and earlier. This equation corresponds to the equation reported in the second column of Table XXXIV, but includes the inverse Mills’ ratio and is estimated over the smaller sample of firms employed in the Table XXXV estimates.\(^9\) This test, which may be interpreted as a test for the presence of fixed effects (see Chapter 3., Section

---

\(^9\) Recall that in the levels equation we exclude observations for which \( L_{t+1}=L_t \) while in the differenced equation we exclude observation for which \( L_{t+2}=L_{t+1} \) or \( L_t=L_{t+1} \).
3.2), gives a value of $\chi^2 = 53.8$. This rejects the null hypothesis of no fixed effects, but the test statistic is not very large in relation to the sample size.

The second Hausman test tests the consistency of the estimates reported as model $t(t-1)(t-2)(t-3)$ in Table XXXIV against the model $(t-1)(t-2)(t-3)$. On the null hypothesis of conditional independence, period $t$ variables are available as instruments and so, the estimates of the model $t(t-1)(t-2)(t-3)$ would be consistent and more efficient than those of the model $(t-1)(t-2)(t-3)$. However, we encountered a computational problem in calculating this Hausman test.

Denote the estimated coefficient vector from the model $t(t-1)(t-2)(t-3)$ by $b_0$ with variance $V_o$, and that from the model $(t-1)(t-2)(t-3)$ by $b_1$ with variance $V_1$. Write $\Delta V = V_1 - V_o$. The Hausman test is

$$H = (b_1 - b_0)'(\Delta V)^{-1}(b_1 - b_0) - \chi^2_k$$

where $k$ is the number of coefficients. We find $V_o$ to be very close to $V_1$ with the consequence that four of the six eigenvalues of $\Delta V$ are very small and negative. Write $\Delta V = P'A'P$, where $A$ is the diagonal matrix of eigenvalues of $\Delta$ and $P$ is the associated matrix of eigenvectors, and partition $A$ as

$$A = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}$$

where $A_1$ corresponds to the $k_1$ negative eigenvalues and $A_2$ corresponds to the $k_2$ positive eigenvalues. This allows us to calculate a modified Hausman statistic

$$H^* = (b_1 - b_0)'P\begin{bmatrix} 0 & 0 \\ 0 & A^{-1}_2 \end{bmatrix}P(b_1 - b_0) - \chi^2_{k_2}$$

Equation (209) gave a test statistic of $\chi^2_{k_2} = 630.3$ implying a clear rejection of the
conditional independence assumption. As a check on this procedure, we calculated a lower bound on the statistic by an alternative procedure. Since

\[ \Delta V = V_1 - V_0 < V_1 \]  \hspace{1cm} (210)

it follows that

\[ H > (b_1 - b_0) / V_1^{-1} (b_1 - b_0) \]  \hspace{1cm} (211)

Equation (211) puts a lower bound on the test statistic of 522.3, which again implies a clear rejection of the null hypothesis. This rejection is not surprising in view of the fact that model t(t-1)(t-2)(t-3) lacks the forward interpretation available for the model (t-1)(t-2)(t-3).

We have calculated the point elasticities with respect to the wage rate for the models "t(t-1)(t-2)(t-3)" in levels shown in Table XXXIII and (a) "(t-2)(t-3)" in Table XXV, conditioning on productivity. We have already noted that the levels model estimated with instruments dated t and earlier has two roots within the unit circle and therefore is to be interpreted as a backward representation. This implies that \( L_{t+1} \) should be seen as the dependent variable and on this basis we calculate the elasticity from the partial derivative \( \partial L_{t+1} / \partial W_t \). The near unit root prevents our calculating a long run elasticity. In the differenced equations, the short run elasticity is based on the total derivative \( dL_t / dW_t \), which depends on both the partial derivative \( \partial L_{t+1} / \partial W_t \) and \( \partial L_t / \partial W_t \) - see Appendix 7.3. Because the firm is forward looking the long run elasticity is only a little greater than the short run elasticity.
Table XXXVII Short and long run elasticities

<table>
<thead>
<tr>
<th></th>
<th>Short-term elasticities</th>
<th>Long-term elasticities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Levels: t(t-1)(t-2)(t-3)</td>
<td>Differences: (a):(t-2)(t-3)</td>
<td>Levels: t(t-1)(t-2)(t-3)</td>
</tr>
<tr>
<td></td>
<td>-0.18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.20</td>
<td></td>
</tr>
</tbody>
</table>

In Chapter 6, we showed how, in the presence of observation disturbances correlated with their preceding values, the optimal decision rule is a non-separable function of ε's. This aggravates the problem of dimensionality of the DP algorithm by generating a state space of expanding dimensions. To obtain computational feasibility it is necessary to assume "conditional independence", and limit the pattern of dependence in the \{L_t, \epsilon_t\} process by assuming that any serial dependence between \epsilon_t and \epsilon_{t+1} is transmitted entirely through the observed state L_{t+1}. Also, the probability density for L_{t+1} depends only on the previous period's stock, L_t and not on the error \epsilon_t.

Conditional independence is therefore crucial if we are to implement DDP models on empirical data. At the same time, conditional independence is highly restrictive, and, one the data we have analyzed the assumption is clearly rejected. The evidence from the estimated levels equations clearly rejects the requirement that the disturbance \epsilon_{t+1} is independent of variables dated t, and furthermore, that assumption leads to an estimated equation which cannot be interpreted as an Euler
equation. The strong implication is that the conditional independence assumption must be relaxed if these data are to be analyzed within the DDP framework.

Rust (1991), has noted the potential restrictiveness of the conditional independence assumption. He states (page 33) "It would be desirable to abandon CI in favor of a more "realistic" specification for unobservables that allows a more flexible pattern of serial correlation. However, at present there appears to be no general computationally feasible alternative". We are forced to conclude that empirical progress in DDP modelling requires resolution of this difficulty.

---

51 See Berckovec and Stern (1990) for an attempt at relaxation of the conditional independence assumption.
Appendix 7.1. Euler equation

The profit function at time $t$ and $t+1$ is respectively

$$\Pi_t = AL_t^\alpha - \frac{Y}{2}(\Delta L_t)^2 - K - W_t L_t$$

and

$$\Pi_{t+1} = \beta \left[ AL_{t+1}^\alpha - \frac{Y}{2}(\Delta L_{t+1})^2 - K - W_{t+1} L_{t+1} \right]$$

The derivative of $\nu^\lambda$ with respect to $L_t$ is formed by the following three parts

$$\frac{\partial \Pi_t}{\partial L_t} = \alpha \left( \frac{Q}{L} \right)_t - \gamma L_t + \gamma L_{t-1} - W_t = 0$$

and

$$\frac{\partial \Pi_{t+1}}{\partial L_t} = \beta \gamma L_{t+1} - \beta \gamma L_t$$

and

$$\frac{\partial(pq)_{t+1}}{\partial L_t} = \left[ \frac{\partial p_{t+1}^{NA|A}}{\partial L_t} q_{t+1} + p_{t+1}^{NA|A} \frac{\partial q_{t+1}}{\partial L_t} \right]$$

The Euler first order conditions are

$$\beta \gamma L_{t+1} - \gamma (1 + \beta) L_t + \gamma L_{t-1} + \alpha \left( \frac{Q}{L} \right)_t - W_t + \beta (pq)_{t+1} = 0$$
Appendix 7.2. Logit estimates

Model ",(t-1)(t-2)(t-3)"

Table AI

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_{t-1}$</td>
<td>0.690</td>
<td>0.122</td>
<td>0.331</td>
<td>0.392</td>
<td>0.314</td>
</tr>
<tr>
<td></td>
<td>(0.155)</td>
<td>(0.253)</td>
<td>(0.221)</td>
<td>(0.241)</td>
<td>(0.244)</td>
</tr>
<tr>
<td>$Q_{t-1}$</td>
<td>-0.068</td>
<td>-0.001</td>
<td>-0.123</td>
<td>0.063</td>
<td>-0.031</td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td>(0.088)</td>
<td>(0.0519)</td>
<td>(0.080)</td>
<td>(0.065)</td>
</tr>
<tr>
<td>$WB_{t-1}$</td>
<td>0.075</td>
<td>0.707</td>
<td>0.246</td>
<td>0.210</td>
<td>0.243</td>
</tr>
<tr>
<td></td>
<td>(0.163)</td>
<td>(0.293)</td>
<td>(0.238)</td>
<td>(0.250)</td>
<td>(0.250)</td>
</tr>
<tr>
<td>$INV_{t-1}$</td>
<td>0.164</td>
<td>0.057</td>
<td>0.080</td>
<td>0.033</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>(0.049)</td>
<td>(0.052)</td>
<td>(0.048)</td>
<td>(0.047)</td>
</tr>
<tr>
<td>$P_{t-1}$</td>
<td>-0.087</td>
<td>-0.028</td>
<td>0.035</td>
<td>-0.053</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>(0.312)</td>
<td>(0.059)</td>
<td>(0.059)</td>
<td>(0.060)</td>
<td>(0.070)</td>
</tr>
<tr>
<td>$S_{t-1}$</td>
<td>-0.008</td>
<td>-0.106</td>
<td>-0.053</td>
<td>-0.159</td>
<td>-0.108</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.033)</td>
<td>(0.025)</td>
<td>(0.041)</td>
<td>(0.035)</td>
</tr>
<tr>
<td>const</td>
<td>0.449</td>
<td>1.003</td>
<td>0.878</td>
<td>0.901</td>
<td>1.026</td>
</tr>
<tr>
<td></td>
<td>(0.073)</td>
<td>(0.096)</td>
<td>(0.081)</td>
<td>(0.100)</td>
<td>(0.094)</td>
</tr>
</tbody>
</table>

log likelihood

-1839.4 | -1515.13 | -1644.55 | -1636.79 | -1566.97

no. obs.

3247 3247 3247 3247 3247

$\chi^2$

175.36 184.78 129.24 189.83 139.37

d.f.

(6) (6) (6) (6) (6)

prob>$\chi^2$

0.000 0.000 0.000 0.000 0.000

The dependent variable is 1 if the firm adjusts and 0 otherwise: The regressors are: one period lagged values, in levels. Variables are: employment (number of employees), L, real sales, Q, the real wage-bill, WB, investment in machinery (in real terms), INV, profitability (in real terms), P, and firm size, S measured as the reciprocal of employment, 1/L.

$\chi^2$, with the degrees of freedom in parentheses, is a test of the joint significance off all regressors except the constant.
<table>
<thead>
<tr>
<th></th>
<th>1988</th>
<th>1989</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_{t-1}$</td>
<td>0.170</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>(0.215)</td>
<td>(0.226)</td>
</tr>
<tr>
<td>$Q_{t-1}$</td>
<td>-0.027</td>
<td>-0.056</td>
</tr>
<tr>
<td></td>
<td>(0.062)</td>
<td>(0.063)</td>
</tr>
<tr>
<td>$WB_{t-1}$</td>
<td>0.226</td>
<td>0.406</td>
</tr>
<tr>
<td></td>
<td>(0.204)</td>
<td>(0.215)</td>
</tr>
<tr>
<td>$INV_{t-1}$</td>
<td>0.054</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>(0.038)</td>
</tr>
<tr>
<td>$P_{t-1}$</td>
<td>-0.069</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>(0.041)</td>
<td>(0.063)</td>
</tr>
<tr>
<td>$S_{t-1}$</td>
<td>-0.197</td>
<td>-0.110</td>
</tr>
<tr>
<td></td>
<td>(0.045)</td>
<td>(0.032)</td>
</tr>
<tr>
<td>const</td>
<td>1.295</td>
<td>1.257</td>
</tr>
<tr>
<td></td>
<td>(0.100)</td>
<td>(0.090)</td>
</tr>
</tbody>
</table>

log likelihood

-1554.83  -1467.64

no. obs.    3247    3247

$\chi^2$  130.49  115.43
d.f.       (6)     (6)
prob > $\chi^2$  0.000  0.000
Appendix 7.3. Calculation of short and long run elasticities

Long run elasticities with respect to wage rate are given by the following formula

\[ \varepsilon_w = \frac{\psi_4}{(1 - \psi_1 - \psi_2)} \frac{\bar{W}}{L} \quad \text{(A4)} \]

where \( \psi_1 \), \( \psi_2 \), and \( \psi_4 \) are the estimated coefficients of \( L_{t+1} \), \( L_t \), and \( W_t \) respectively.

Short run elasticities with respect to wage rate are given by

\[ \varepsilon'_w = \frac{\psi_4 \mu_1}{(\mu_1 - 1)(1 - \psi_1 \mu_2)} \frac{\bar{W}}{L} \quad \text{(A5)} \]

\( \psi_1 \) and \( \psi_4 \) are estimated coefficients of the \( L_{t+1} \), and \( W_t \) respectively and \( \mu_1 \) and \( \mu_2 \) are the characteristic roots as calculated in Tables XXXIV and XXVI. In order to obtain (A5) we write, conditional on productivity \( Q/L \):

\[ L_t = \psi_1 L_{t+1} + \psi_2 L_{t-1} + \psi_4 W_t + \xi_t \quad \text{(A6)} \]

First order conditions with respect to \( w_t \) are:

\[ \frac{\partial L_t}{\partial w_t} = \psi_1 \frac{\partial L_{t+1}}{\partial W_t} + \psi_4 \quad \text{(A7)} \]

where

\[ \frac{\partial L_{t+1}}{\partial W_t} = \psi_1 \frac{\partial L_{t+2}}{\partial W_t} + \psi_2 \frac{\partial L_t}{\partial W_t} + \psi_4 \quad \text{(A8)} \]

Equation (A8) represents a second order differential equation.

Define:
\[ B_0 = \frac{\partial L}{\partial W_t} = \psi_1 \frac{\partial L_{t+1}}{\partial W_t} + \psi_4 = \psi_1 B_1 + \psi_4 \]  
(A9)

\[ B_1 = \frac{\partial L_{t+1}}{\partial W_t} = \psi_1 B_2 + \psi_2 B_0 + \psi_4 \]

In general we may define

\[ B_i = \psi_1 B_{i+1} + \psi_2 B_{i-1} + \psi_4 \quad i > 0 \]  
(A10)

Thus

\[ B_i = \mu_2 B_{i-1} + \frac{\psi_4}{\psi_1 (\mu_1 - 1)} \]  
(A11)

\[ B_1 = \mu_2 B_0 + \frac{\psi_4}{\psi_1 (\mu_1 - 1)} \]

\[ B_0 = \psi_1 + \psi_4 \]

where \( \mu_1 \) and \( \mu_2 \) are the roots associated with (A10). Using the lag operator \( L \) we obtain the following expressions:

\[ \left( \psi_1 - L - \psi_2 L^2 \right) L^{-1} B_i = -\psi_4 \]

\[ \left( \psi_1 - L - \psi_2 L^2 \right) B_{i+1} = -\psi_4 \]

\[ \left( 1 - \frac{1}{\psi_1} L + \frac{\psi_2 L^2}{\psi_1} \right) B_{i+1} = -\frac{\psi_4}{\psi_1} \]  
(A12)

\[ (1 - \mu_1 L)(1 - \mu_2 L)B_{i+1} = -\frac{\psi_4}{\psi_1} \]

Moreover

\[ \mu_1 \mu_2 = \frac{\psi_2}{\psi_1} \quad \mu_1 + \mu_2 = \frac{1}{\psi_1} \]  
(A13)

where \( \mu_1 > 1 > \mu_2 > 0 \).
The last line of (A12) may be rewritten and developed as

\[
\mu_1 L \left( 1 - \frac{1}{\mu_1} L^{-1} \right) (1 - \mu_2 L) B_{i+1} = \frac{\psi_4}{\psi_1}
\]

\[
(1 - \mu_2 L) B_i = \frac{\psi_4}{\psi_1 \mu_1} \left( \frac{1}{1 - \frac{\psi_4 L^{-1}}{\mu_1}} \right)
\]

\[
B_i - \mu_2 B_{i-1} = \frac{\psi_4}{\psi_1 \mu_1} \left( \frac{1}{1 - \frac{1}{\mu_1}} \right) = \frac{\psi_4}{\psi_1 (\mu_1 - 1)}
\]

From (A9) we have that

\[
B_1 = \mu_2 B_0 + \frac{\psi_4}{\psi_1 (\mu_1 - 1)}
\]

\[
B_0 = \psi_1 B_1 + \psi_4 = \psi_1 \mu_2 B_0 + \frac{\psi_1 \psi_4}{\psi_1 (\mu_1 - 1)} + \psi_4
\]

Hence

\[
(1 - \psi_1 \mu_2) B_0 = \psi_4 \left[ \frac{1}{\mu_1 - 1} + 1 \right] = \frac{\psi_4 \mu_1}{\mu_1 - 1}
\]

\[
B_0 = \frac{\psi_4 \mu_1}{(\mu_1 - 1)(1 - \psi_1 \mu_2)}
\]

which allows us to calculate the short run elasticity in (A5)

\[
E_w = B_0 \frac{\bar{W}}{L} = \frac{\psi_4 \mu_1}{(\mu_1 - 1)(1 - \psi_1 \mu_2)} \frac{\bar{W}}{L}
\]
CONCLUSIONS AND TOPICS FOR FURTHER RESEARCH

In this thesis we have considered several issues on the dynamics of employment decisions.

The first concerns the existence of fixities in the structure of labour adjustment costs. We have considered the way that many technologically and legally imposed labour costs are fixed in the sense of being independent of the size of the change in the firm's employment level. We have also emphasized that this has an important implication for the behaviour of firms and that it leads to an adjustment process very different from that described by the traditional economic literature on dynamic labour demand, based on the assumption of quadratic adjustment costs. In fact, in the presence of fixed costs, the employment adjustment path is characterized by large and infrequent changes in employment in the face of shocks in contrast to the continuous and smooth path implied by quadratic costs. Hence, it is crucial to look at the actual form of the constraints imposed by the regulations on hiring and firing decisions, if we are to understand optimal decision making by firms in the labour market.

We have analyzed the Italian case, which provide a clear example of how institutions may influence companies' employment strategies. We have emphasized
how the interactions between rules and practices, the written versus the living law, appear to have played an important role in Italian industrial relations. For this reason, an appreciation of the way in which practice departs from the written law has become essential to an understanding of how employers formulate plans. We have also identified an extra element of cost to add to what is an already complex categorization, namely the cost implied by the attempt to circumvent constraints imposed by legislation. This type of cost represents an important source of fixity which should be added to the pre-existing category of fixed costs.

The existence of non-convex adjustment costs is also indicated by the descriptive analysis using a panel data set on 3247 companies, over the period 1982-89, relating to Lombardia: an industrialized region, representative of Northern Italy. Changes in employment show a recurring pattern throughout the sample period: a very high spike at zero reveals a considerable stickiness in employment. Throughout the period, the rate of change in employment was zero for more than 20 per cent of the firms, on average. We pointed out how, at that preliminary descriptive stage, fixed adjustment costs could only represent one possible explanation for that empirical evidence. We also considered another possible interpretation: namely, the possibility that the zero changes may be due to measurement errors, such as recording errors, misreporting, etc.

Whatever the source of the inaction, we need to take into account this censoring when we estimate a standard model of quadratic adjustment costs. This leads to a selection model where we observe adjustment in the standard case of firms facing quadratic adjustment costs, and where we have censoring arising from one or both of the reasons discussed above.
The second important issue analyzed in the thesis therefore concerns the estimation of an Euler equation by the generalized method of moments procedure in the presence of selection bias. In order to estimate the model, we adopt a two-stage procedure. First, we study the selection process by estimating the discrete choice of whether or not to adjust, using a maximum likelihood probit estimation procedure and thereby obtain the vector of inverse Mills' ratios which is included, at the second stage, in the Euler equation as the selection correction term. An important feature of our estimation procedure is that the selection process is specified in terms of the instruments which define the orthogonality conditions in the GMM estimation of the adjusted Euler equation, using the non-censored sample. We consider four different specifications of the Euler equation in order to take into account fixed effects and to experiment with different timing of the instruments. Each model includes the appropriate selectivity term calculated with regard to the timing of the instruments.

Our approach provides a general procedure for taking selectivity bias into account in panel data estimation. The use of the instrumental variables employed in the GMM estimation procedure makes the probits for selection correction consistent with the specification of the Euler equation.

The estimation reveals that employers respond to shocks on wages and productivity by varying the number of workers only if these shocks are very large. Our explanation relies on the existence of high adjustment costs which make employers very cautious in adjusting employment unless shocks to wages and productivity are perceived as substantial. Hence, we either obtain no adjustment as the result of the size of adjustment costs and the lack of large shocks to wages and
productivity over the period, or a very slow adjustment as the consequence of the cautiousness of employers in varying the number of workers. The estimates do not show significant endogenous selection bias and this leads us to consider more explicitly the way in which fixed costs should enter the labour adjustment model.

A third main issue in the thesis is methodological: given the existence of fixed adjustment costs, are we able to maintain the framework of the Euler-GMM approach? The explicit consideration of fixed costs requires a more complicated set of tools which allows us to consider the existence of a finite, countable number of alternatives - in our case, to adjust or not to adjust - from which the firm may optimally choose. We have therefore turned to the literature of discrete decision processes. This allowed us to obtain a dynamic model of employment decisions under fixed costs where the firms sequentially chooses between variation and non-variation of its labour input. If we condition on companies' decisions to optimally adjust in period t and if we suppose the existence of both fixed and quadratic adjustment costs we may obtain the Euler first order conditions and exploit them as orthogonality conditions as implied in the standard Euler-GMM approach.

There are however important differences between the Euler equation formulation we obtain in the presence with both fixed and quadratic adjustment costs and the model with only quadratic adjustment costs. With both fixed and quadratic adjustment costs we condition on companies choosing optimally to adjust at time t, and thereby obtain an Euler equation "conditional " on this choice. Indeed, this is only a part of a more complicated intertemporal decision problem which includes also the possibility of optimally deciding not to adjust at t and implying a discontinuity at this time. The Euler equation we obtain, under fixed and
quadratic costs, shows the presence of a forward looking term, which we denoted as \((pq)_{t+1}\), which is the result of taking into account that, with both fixed and quadratic costs, the employer faces the alternative of whether or not to adjust in each subsequent period. The term \(pq^+i\) captures the fact that due to fixed costs some of the companies which optimally decided to adjust at time \(t\) may optimally choose not to vary employment in some of the subsequent periods. The estimation of the new Euler equation requires an initial estimate of this term.

As a number of previous authors have shown, this model can only be estimated under the simplifying assumption of conditional independence of the disturbances. This assumption requires that the disturbances are serially independent conditional on the stock of employment, and thereby not only rules out serial dependence but also fixed effects. Unobserved heterogeneity can therefore affect current decisions only through the stock of employment. In the absence of this assumption it is not easily possible to obtain an empirically useful simplification of the firm's decision problem. This assumption is clearly very restrictive, and we found that it failed to work well in relation to our data. Specifically, we estimated a model which conditioned on the decision to adjust in period \(t+1\) on period \(t\) variables, which would be valid on the conditional independence assumption. However, the estimated equation was not interpretable as an Euler equation (both roots were within the unit circle), and which was rejected using the Hausman test. The implication is that conditional independence is not sustainable in relation to our data.

Throughout, we have considered cost structures which relate to net changes in employment, an approach taken in some but not all the literature, focusing on

337
decisions on turnover of jobs rather than of workers. We are, nevertheless, aware of the importance of gross adjustment costs and the importance of comparing the net and gross costs in analyzing dynamic factor demand. In particular, macroeconomic fluctuations may have substantial effects beyond those indicated by net employment changes at firms level. Clearly, when we consider net changes in employment we refer to a highly aggregative measure, which does not allow us to distinguish between net changes in employment, job creation, job destruction and labour turnover, concepts which imply different decisions and constraints. Firms react to shocks by activating one channel or another according to the constrains they face. The perception of the shock as transitory or permanent also determines the size of the adjustment. The information we may find on the processes which generate adjustment costs increases with the degree of disaggregation of the flows of entry and exit. Unfortunately, it is very hard to find such detailed information and the panel of Italian company data which we analyzed does not permit discussion of these important issues. This may qualify the conclusions we draw.

In relation to the same argument, we discussed in Section 2.4.2. the issue of the appropriate level of aggregation and what can be said about the structure of adjustment costs from data which are only available at a high level of temporal aggregation. Errors may arise from using data at survey intervals instead of measuring the state and control variables at the time interval that is relevant for the firm. There are in fact two related issues here:

i) spatial aggregation: Is the firm the relevant decision-making unit?

ii) temporal aggregation: Does the time interval relevant to firm decisions on whether to alter employment correspond to the calendar year?
Even if we have shown that annual data at the firm level provide some information on the structure of costs, it would be very difficult to reach any definitive conclusion on the appropriate frequency and level of spatial disaggregation in the data which would yield the best evidence on the structure of adjustment costs.

Among neglected issues, three may prove very important for this analysis. The first concerns heterogeneity of the labour input. The costs of hiring and firing may differ for production and non-production workers generating different adjustment paths for these two categories over the business cycle. Moreover, hiring and firing entail costs which have different origins and may result in differing patterns of adjustment. Although our models of dynamic labour demand with fixed costs and with both fixed and quadratic adjustment costs do not exclude the possibility of asymmetries in the adjustment (for instance, the (S,s)-type of model in Section 6.3. considers adjusting upwards, downwards and non adjusting), we have not explicitly analyzed the implications of asymmetries of this sort.

The second problem relates to the lack of information in the data on adjustment of hours. This may be an important issue, since it is likely that, in the presence of high adjustment costs, companies resort quite generally to vary hours rather than employees over the cycle. In other words, hours adjustment may interact with adjustment of the number of workers, and by acting as a buffer with respect to cyclical shocks, affect the frequency of adjustment decisions. For this reason, hours should explicitly appear in the firm’s objective function. Again, our data did not permit us to extend the empirical analysis in this direction.

A complete account of different forms of adjustment costs should also consider linear costs. In fact, linearity could represent a very important component
of adjustment costs. As we pointed out in the introduction, a number of authors have devoted attention to linear costs. However, as already appears from the brief and mostly intuitive discussion in the introductory chapter, linear costs may not be easily distinguishable at an empirical level from fixed costs. In fact, both structures entail zones of non-adjustment, which although they arise out of very different cost structures in theory, are not easily identifiable at the empirical level. In the presence of linear costs, firms find it optimal to adjust instantaneously to the desired level of employment. However, because of these costs, their adjustment will stop to adjust somewhere beneath the level implied by the absence of costs. This creates a zone of non adjustment, but adjustment does not need to be necessarily infrequent as in the case of fixed costs. Fixed costs, instead, imply infrequent adjustment (and rather large when it occurs): the firm may decide not to adjust, entailing an inaction zone which does not occur around the non-adjustment-costs-equilibrium level of employment. The issue of how to distinguish empirically between these very different theoretical structures forms part of the agenda for future research.

The final point concerns the extension of this work. Our analysis concluded with the study of the Euler equation in the presence of both fixed and quadratic adjustment costs, i.e. the branch of the decision tree starting with optimally choosing to adjust. We did not consider the other part of the tree relating to the optimal choice of not to adjust. The complete structural model, in the presence of only fixed costs, was discussed in Chapter 6, Section 6.2., equation (184). That equation was formed by combining the transition probabilities with the weighted combination of the two conditional valuation functions relating to the decisions of
whether optimally to adjust or not to adjust at time t, where weights were given by the conditional choice probabilities. That formulation allows the estimation of the structural parameters of the model. Hotz, Miller, Sanders and Smith (1994) suggest a "Conditional Choice Simulation" (CCS) estimator. This implies to consider expected profits associated with a path of simulated future choices obtained by exploiting consistent estimates of the conditional choice probabilities and the transition probabilities governing outcomes. Estimated equations for the structural parameters of the model may be formed using the utilities associated with the simulated paths to form valuation functions. The CCS estimator requires only unrestricted estimates of the conditional choice and transition probabilities associated with the nodes of a hypothetical firm's simulated future adjustment path.

Two principal conclusions may be drawn from this thesis. The analysis of the legislation of Italian labour market strongly indicates that fixed components in the structure of labour adjustment costs exist and may be relevant to employers’ optimal employment decisions. Hence, they should be explicitly considered when modelling firms’ employment strategies through time and when considering the effects of policies aimed at reducing employment fluctuations over the cycle. However, the explicit consideration of fixed costs requires more sophisticated tools, and use of these tools implies that the mathematical advantages which derive from the traditional formulations based on strict convexity are lost. In particular, the treatment of dynamics requires the use of discrete decision processes which, in order to be estimable, crucially rely upon the assumption of conditional independence. We have tested this assumption and found that in our model it is
decisively rejected. Clearly, this does not exhaust the scope for the analysis of fixed costs in this context. Some of the data limitations described earlier may affect the results we have obtained; moreover the direct estimation of a structural model entails further tests of the conditional independence hypothesis. It is also important in future work to use non-parametric methods to estimate the conditional choice probabilities which give the forward looking (pq) correction to the Euler equation. But more important, in the end, would be development of models which relax the conditional independence assumption.
REFERENCES


ASPO (1991), "Rapporto sulla Lombardia".


BANCA D'ITALIA, (1990), "Relazione Annuale".


DEL BOCA, A. and ROTA. P. (1989a), "Wage Compensation and Employment:


EUROSTAT (1989), "Labour Force Survey".


ICHINO, P. and VIOLI, M. (1988), "Neutralizzare lo handicap", Cendom-Irer,


PINDYCK, R. and ROTEMBERG, J. (1983b), "Dynamic Factor Demands and the


SHAPIRO, M. (1986), "


STOKEY, N. and LUCAS, R. Jr. (1989), "Recursive Methods in Economic


WELLS, B. (1993), "Does the Structure of Employment Legislation Affect the Structure of Employment and Unemployment?", *manuscript*.