

IFAU – INSTITUTE FOR LABOUR MARKET POLICY EVALUATION

Essays on the evaluation of social programmes and educational qualifications

Barbara Sianesi

DISSERTATION SERIES 2002:3

Presented at the University College London, United Kingdom

The Institute for Labour Market Policy Evaluation (IFAU) is a research institute under the Swedish Ministry of Industry, Employment and Communications, situated in Uppsala. IFAU's objective is to promote, support and carry out: evaluations of the effects of labour market policies, studies of the functioning of the labour market and evaluations of the labour market effects of measures within the educational system. Besides research, IFAU also works on: spreading knowledge about the activities of the institute through publications, seminars, courses, workshops and conferences; creating a library of Swedish evaluational studies; influencing the collection of data and making data easily available to researchers all over the country.

IFAU also provides funding for research projects within its areas of interest. There are two fixed dates for applications every year: April 1 and November 1. Since the researchers at IFAU are mainly economists, researchers from other disciplines are encouraged to apply for funding.

IFAU is run by a Director-General. The authority has a traditional board, consisting of a chairman, the Director-General and eight other members. The tasks of the board are, among other things, to make decisions about external grants and give its views on the activities at IFAU. Reference groups including representatives for employers and employees as well as the ministries and authorities concerned are also connected to the institute.

Postal address: P.O. Box 513, 751 20 Uppsala Visiting address: Kyrkogårdsgatan 6, Uppsala Phone: +46 18 471 70 70 Fax: +46 18 471 70 71 ifau@ifau.uu.se www.ifau.se

This doctoral dissertation was defended for the degree of Doctor in Philosophy at the University College London, United Kingdom, November 18, 2002. Part one of this thesis contains revised versions of research previously published by IFAU as Working Paper 2001:5 and Working Paper 2002:5.

ISSN 1651-4149

Essays on the Evaluation of Social Programmes and Educational Qualifications

Barbara Iole Marina Sianesi

September 2002

A Thesis Submitted for the Degree of Doctor of Philosophy to University College London, the University of London.

Abstract

The first part of the thesis addresses a programme evaluation problem: the estimation of the short- and long-term effects of the Swedish active labour market programmes on participants' subsequent labour market outcomes, in particular individual employment probability and collection of unemployment benefits over time.

Exploiting a unique and comprehensive new Swedish dataset with extensive information on more than 110,000 individuals followed for five years, semiparametric propensity score matching techniques are adapted to the Swedish institutional context to investigate the overall effectiveness of the 'Swedish model' of labour market policy in the context of the high unemployment atypically experienced by Sweden in the 1990s. The performance of the Swedish system is thus considered in its entirety, combining all the programmes into one and focusing on the interactions between the unemployment benefit system and the programme system. Subsequently, the relative performance of the six main types of programmes available to unemployed adults in the 1990s is analysed, both relative to one another and vis-à-vis more intense job search in open unemployment. The differential performance of labour market training, workplace introduction, work experience placement, relief work, trainee replacement and employment subsidies is investigated using a multiple-treatment extension of the propensity score matching method.

The second part of the thesis deals with the evaluation problem in the returns to education framework. Different non-experimental estimation methods to recover the effect of education on earnings – ordinary least squares, instrumental variables, control function and matching – are reviewed and contrasted in the context of alternative microeconometric models – single- and multiple-treatment, homogeneous- and heterogeneous-returns models. The methods are subsequently applied to high quality data – the British 1958 NCDS birth cohort – to estimate private returns to schooling and to illustrate the sensitivity of the different estimators to model specification and data availability.

Acknowledgements

The person who most of all has encouraged, supported and advised me in my research since its very start is my supervisor, Costas Meghir, to whom I owe a huge debt of gratitude for countless stimulating discussions, suggestions and continuous guidance. In fact, in order not to forsake his wisdom and kindness in times to come, I've promptly promoted him to Life Coach.

An intellectual debt of great magnitude is of course owed to my second supervisor, Richard Blundell, for numerous fruitful discussions, comments and interactions, especially on the second part of the thesis, as well as for constant encouragement.

Costas' initial suggestion and the IFS offer to provide me with excellent facilities, the perfect human and physical environment to carry out my research as well as a PhD scholarship – funded by the ESRC Centre for Microeconomic Analysis of Fiscal Policy – have easily tipped the scales in favour of my starting all over again with a second PhD, in a completely new but so fascinating a subject. Thanks are due also to the Economics Department of UCL for encouraging me to take a place on their doctoral programme.

A very special thanks goes to Susanne Ackum Agell for her enthusiasm in encouraging my work on the Swedish programmes all along, as well as for organising financial support through the IFAU.

Funding from the DfES's Centre for the Economics of Education (CEE) for research covering the second part of the thesis is similarly gratefully acknowledged.

Special mention as regards IFS should be given to Andrew Dilnot, for his generosity and kindness, to Lorraine Dearden, the perfect Sector Leader, a wonderful colleague and a dear friend, to Erich Battistin for countless (hopefully mutually) enriching discussions, and to Tom Clark, Emla Fitzsimons, Hide Ichimura and Elena Martinez for comments and suggestions.

Special thanks as regards Sweden, in addition to Susanne, go to Kenneth Carling, Anders Forslund, Bertil Holmlund, Laura Larsson, Katarina Richardson, seminar participants at IFAU and an anonymous referee for helpful comments. Kerstin Johansson has kindly sent me the municipality-level data, and Helge Bennmark and Altin Vejsiu and especially Anders Harkman have been extremely helpful with institutional information and data issues clarifications.

Helge has been patiently dealing with a number of technical questions about the available data, and Anders H. in particular has provided me with invaluable programme-insider information and detailed comments.

Hapless IFS visitors Bernd Fitzenberger, Kei Hirano and Jeff Smith deserve special thanks for very intense and just as beneficial discussions.

Additional thanks go to Edwin Leuven, Astrid Kunze, seminar participants at IFS, Copenhagen University, IZA and the IVA Stockholm conference, and to the referees for the *Journal of the Royal Statistical Society* for very helpful suggestions.

On a more personal front, the "IFS weekend and evening clique" (Elena, Emla, Erich, Mario, and Renata) has been providing encouragement, entertainment and emotional support. Edwin, met through serendipity over my psmatch programme, deserves additional heartfelt thanks for his constant support and sunshine late-night phone calls to a deserted dark office in the final month of the thesis.

And last, but by no means least, my parents, who have always been there, in any sense and in any way needed, and to zio Edoardo and zia Barbara, for never forgetting their niece in London.

Declaration

- 1. No part of this thesis has been presented to any other university for any degree.
- 2. Chapters 4 and 5 were undertaken as joint work with Richard Blundell and Lorraine Dearden; they are under revision for the *Journal of the Royal Statistical Society*, *Series B*.
- 3. A non-technical, policy-oriented summary of the results from Chapters 2 and 3 is forthcoming in the *Swedish Economic Policy Review* as "Swedish Active Labour Market Programmes in the 1990s: Overall Effectiveness and Differential Performance".

Barbara Iole Marina Sianesi

Table of Contents

Ab	stract		2
Ac	know]	ledgements	3
Dee	clarat	ion	5
Int	roduc	etion	12
	<u>Prog</u>	PART ONE ramme evaluation: A Swedish application	
I.	Evalı gram	uating the Swedish Active Labour Market Pro- umes: An Overview	16
II.	Over labou	all performance of the Swedish system of active Ir market programmes in the 1990s	21
II.1	Introd	uction	21
II.2	The S	wedish labour market policy	23
II.3	Data a	nd sample selection	25
II.4	The ev	valuation problem and propensity score matching	26
II.5	Matching in the Swedish institutional set-up		29
	II.5.1	Treated and non-treated in Sweden and other key choices	29
	II.5.2	Methodology	33
	II.5.3	Plausibility of the CIA: Selection into the Swedish pro- grammes and available information	38
II.6	Empirical findings		42
	II.6.1	Entitlement status as a determinant of programme participa- tion	42
	II.6.2	Outcomes over time	44
	II.6.3	Accounting for a partially unobserved outcome variable	49
	II.6.4	A summary so far	56
	II.6.5	Job accession	59
II.7	Unem	ployment-programme cycling behaviour	62
II.8	Progra	ummes for unemployed adults: The role of entitlement status	68

II.9 Discussion and conclusions	70
II.10 Appendix	73
III. Differential performance of the Swedish active labour market programmes for unemployed adults in the 1990s	75
III.1 Introduction	75
III.2 The Swedish labour market policy	78
III.3 Data, sample selection and a preliminary look at the data	82
III.4 Methodology	87
III.4.1 The evaluation problem in a multiple-treatment framework	87
III.4.2 Multiple-treatment matching in the Swedish institutional set- up	92 05
III.4.3 Plausibility of the matching approach in the Swedish context	93 98
III.5 Empirical findings	99
III.5.1 Employment probability over time	103
III.5.2 Unemployment-benefit collection probability over time	105
III.5.3 API versus ALU	105
III.5.4 The problem of the 'lost' individuals	108
III.6 Discussion and conclusions	
III.7 Appendix	
Summary and conclusions to Part One	114

PART TWO Estimating the returns to education

128

References

Intr	Introduction		
IV.	. Models and methods		
IV.1	The earnings-education relationship: A general set-up and alternative models		
	IV.1.1 The homogeneous returns model	141	
	IV.1.2 The heterogeneous returns model	143	

IV.2	The ear	rnings-edu	acation relationship: Alternative methods	145
	IV.2.1	An overv	iew	145
	IV.2.2	2 Least squares		
	IV.2.3	Instrume	ntal variable methods	151
		IV.2.3.1	IV in the homogeneous one-factor model	152
		IV.2.3.2	IV in the heterogeneous one-factor model: A special case	153
		IV.2.3.3	IV in the heterogeneous single treatment model	154
		IV.2.3.4	Some drawbacks to IV	158
	IV.2.4	Control f	unction methods	159
		IV.2.4.1	The heterogeneous single treatment model	160
		IV.2.4.2	The homogeneous returns model	161
		IV.2.4.3	The multiple treatment model	162
		IV.2.4.4	The heterogeneous one-factor model	163
		IV.2.4.5	Some drawbacks to CF	163
	IV.2.5	Matching	g methods	164
		IV.2.5.1	General matching methods	165
		IV.2.5.2	Propensity score matching	168
		IV.2.5.3	Implementing propensity score matching estimators	170
		IV.2.5.4	The multiple treatment model	172
		IV.2.5.5	Some drawbacks to matching	173
	IV.2.6	OLS, mai	ching, instrumental variables and control function	174
IV.3	Appen	dix: A bala	ancing score for sequential multiple treatments	176
	IV.3.1	Looking j	for a balancing score	179
	IV.3.2	A balance	ing score for the pairwise comparisons	179
v.	An ap	oplicatio	n to the British NCDS	181
V.1	Introdu	iction		181
V.2	Single	treatment	model: Higher education	181
V.3	The mu	ultiple trea	tment model	191
V.4	Appen	dix		198
Sum	Summary and conclusions to Part Two			200
References			203	

List of Tables

2.1	Computation of π_i to derive worst- and best-case bounds	54
2.2	A crude summary of the effect of participation <i>versus</i> wait- ing: The more permanent treatment effect on the probability of being in the various states, six months after joining a pro- gramme (% points)	58
2.3	Average treatment effects by month of placement into the programme (averaged over the 5-year horizon since the start of the programme)	59
2A.1	Matching protocol	73
2A.2	Treated and pool of potential controls by time in unemploy- ment	74
2A.3	Matching quality in terms of the propensity score	74
2A.4	Matching quality in terms of imbalance of the most impor- tant covariates between treated and matched controls: stan- dardised differences after matching (% bias)	74
3.1	Synoptic table of the main features of the programmes	80
3.2	Transitions from the first programme onwards (% of respec- tive participants)	84
3.3	Selected individual descriptive statistics, by type of exit from first unemployment spell	86
3.4	Informal summary of the various conditional average treat- ment effects on employment probability over 5-year horizon since programme start	102
3A.1	Specification chosen and indicators of resulting matching quality	114
5.1	The returns to higher education compared to less-than higher education (% wage gain) – Average treatment effect (ATE), average effect of treatment on the treated (ATT) and average effect of treatment on the non-treated (ATNT)	183
5.2	Incremental treatment effects (% wage gain) – Matching and OLS estimates	193
5A.1	Summary statistics, NCDS Men	198
5A.2	Matching quality indicators	199

List of Figures

1.1	Swedish total unemployment, broken into open unemploy- ment and programme participation rates, 1985-2000 (per- centage points)	16
2.1	Marginal effect of UI-status on the probability of joining a programme (percentage points difference in the treatment probability with respect to non-entitled with the same char- acteristics of UI individuals), by time unemployed prior to programme	44
2.2	Treated and matched controls' differential probability of (A) Programme participation and (B) Employment	45
2.3	Treatment effect (% points) on the probability of (A) Pro- gramme participation, (B) Unmployment, (C) Benefit collec- tion, (D) Employment, (E) De-registration/Employment, (F) Regular education, (G) Inactivity and (H) 'Lost'	46
2.4	Average treatment effects on employment probability (aver- aged over the 5-year horizon since start of the programme) by misclassification rate	50
2.5	Average employment effect by cut-off probability (averaged over the 5-year horizon since start of the programme)	51
2.6	Treatment effect on employment probability, using observed employment rates, imputed employment rates, worst-case and best-case bounds (% points)	55
2.7	Expected labour market status of individuals in the case of (A) waiting longer in open unemployment and (B) joining a programme, broken down into probability of being on a pro- gramme, in employment or in unemployment, the latter in turn broken down into compensated and uncompensated one. (Average over the 5-year horizon since the start of the pro- gramme)	57
2.8	Survival in unemployment for treated and controls (days)	61
2.9	Compensated cycle probability over time: (A) Treated and matched controls and (B) Treatment effect	63
2.10	Compensated cycling by month of placement: (A) Average treatment effect on compensated cycling probability and (B) Percentage of individuals becoming compensated cyclers	65
2.11	Programme and benefit collection probability over time: 14- group versus 1-12 group	66

2.12	Average effect for participants of joining any of the programmes compared to waiting longer in open unemployment on employment probability and benefit collection probability: entitled and non-entitled adults	70
3.1	Raw data: employment probability over time, by type of pro- gramme	85
3.2	Selection process into the Swedish programmes and key available regressors	96
3.3	Average effect on employment probability over time of join- ing the specified programme compared to waiting longer in open unemployment for participants in the specified pro- gramme	101
3.4	Differential performance of ALU and API	106
3A.1-5	Differential average effects on employment probability over time of the specified programme compared to the various alternatives for participants in the specified programme	115
3A.6-10	Differential average effects on compensated unemployment probability over time of the specified programme compared to the various alternatives for participants in the specified programme	118
3A.11	Average effect on employment probability over time of join- ing the specified programme compared to waiting longer in open unemployment for participants in the specified pro- gramme: estimated effect and best- and worst-case bounds	123

Introduction

The essays in this thesis address evaluation questions. Part One deals with a programme evaluation problem: the estimation of the effect of the Swedish active labour market programmes on subsequent labour market outcomes of participants. Part Two relates to the returns to education issue: the estimation of the effect of schooling on earnings.

The first chapter of Part One is an introduction which highlights a series of features that make the Swedish system of active labour market programmes particularly interesting to analyse. It also anticipates the importance of a number of issues which cannot be overlooked when performing the evaluation or interpreting the results.

Exploiting a unique and comprehensive new Swedish dataset with extensive information on the labour market history of more than 110,000 individuals followed for five years, Chapters 2 and 3 investigate how effective these programmes have been over the last decade. The effectiveness of the programmes in improving the labour market prospects of unemployed participants in the short- and long-term is assessed in terms of a number of labour market outcomes, in particular individual employment probability and collection of unemployment benefits over time. Given the richness of the data, semi-parametric matching techniques assuming selection on observables have been adapted to the Swedish institutional set-up.

Chapter 2 investigates the overall effectiveness of the 'Swedish model' of labour market policy, in particular how successful it has been in the context of the high unemployment atypically experienced by Sweden in the early 1990s. The performance of the Swedish system is thus considered in its entirety, combining all the programmes into one and focusing on the interactions between the unemployment benefit system and the programme system. In fact, a labour market programme in Sweden effectively comes as a bundle of two conflicting components: it is intended to equip job-seekers with marketable skills which should improve their opportunities on the labour market, but at the same time it allows to renew eligibility to generous unemployment compensation, thus reinforcing the work disincentive associated with the unemployment insurance system.

The Swedish active labour market programmes encompass a wide variety of different interventions, and various programmes may in fact have differential effects. It is thus interesting to evaluate the relative effectiveness of different types of programmes, ideally with a view of identifying the best performing ones. Focusing on individuals entitled to unemployment benefits – one that group whose incentives are most likely to be affected –, Chapter 3 evaluates the relative performance of the six main types of programmes available to them in the 1990s: labour market training, workplace introduction, work experience placement, relief work, trainee replacement and employment subsidies. The differential performance of the programmes is investigated both relative to one another and vis-à-vis more intense job search in open unemployment using a multiple-treatment extension of the propensity score matching method.

The overall summary and conclusions to Part One reviews and combines the results from the two main chapters and discusses their policy implications.

Part Two deals with the evaluation problem in the returns to education framework. Its first chapter reviews alternative models and estimation methods to recover the causal effect of education on earnings.

As to the specification of the model, the chapter highlights the importance of distinguishing between models which refer to the impact of a specific educational level (single treatment models) and models which allow for a number of sequential levels of schooling (multiple treatment models), an issue that the preceding two chapters have addressed in the Swedish programme evaluation context. A second crucial choice concerns the nature of the returns to education, in particular the importance of allowing the returns to vary across individuals for the same educational qualification.

As to the choice of estimation method, the matching method used in the preceding chapters is discussed both on the methodological and empirical side, adapting it to the returns to education context. The assumptions for its validity as well as the empirical issues arising in its implementation are contrasted with those for other three evaluation methods: in addition to the benchmark of ordinary least squares (linear matching), the instrumental variable method and the control function method. For the different model specifications, the chapter highlights the assumptions needed for the validity of each method and investigates the estimators' properties.

The last chapter subsequently applies these methods to high quality data – the British 1958 NCDS birth cohort – to estimate private returns to schooling and to illustrate the sensitivity of the different estimators to model specification and data availability. In particular, the different estimation approaches are applied in the single and in the multiple treatment framework. The relative magnitude of the different estimates is then compared and contrasted to see what can be learnt about the selection and outcome models.

PART ONE

Programme Evaluation: A Swedish Application

CHAPTER I

Evaluating the Swedish Active Labour Market Programmes: An Overview

Sweden occupies a special place when it comes to active labour market programmes (ALMP): a long-standing reliance on such measures has been accompanied by traditionally low unemployment rates by European standards, two features which several observers have often related to one another (e.g. Layard, Nickell and Jackman, 1991), viewing the 'Swedish model' as an example for other countries to emulate.¹

The beginning of the 1990s has however witnessed a dramatic change in the labour market situation in Sweden, hit by its most severe recession since the war: unemployment swiftly reached unprecedented levels, more than quadrupling between 1990 and 1993, and as a policy response, so too did the offer of labour market programmes (see Figure 1.1).

Figure 1.1 Swedish total unemployment, broken into open unemployment and programme participation rates, 1985-2000 (percentage points)



Sources: Open unemployment rate is from the Labour Force Survey (LFS). Programme rate is defined as the number of programme participants over the labour force; the former is from the National Labour Market Board (AMS) register, the latter from the LFS.

¹ Quite interestingly, in the UK a programme (the 'New Deal') sharing some of the features of the Swedish set-up has recently been introduced (April 1998).

This deep and sudden recession has posed new challenges to the Swedish labour market policy, when expenditure on the country's extensive offer of labour market programmes reached 3 percent of GDP. In the presence of both rising unemployment and expanding budget deficit, the economic justifiability of the expense of the Swedish programmes would in fact hinge on the assessment of their actual effectiveness. Accordingly, interest has been rising in evaluating how successful such large-scale measures have effectively been.

The issue of whether the Swedish system was in fact successful in the context of the high unemployment atypically experienced by Sweden is taken up by Chapter 2, which investigates how effective the Swedish ALMPs as a whole have been in improving the labour market opportunities of unemployed individuals over the last decade.

A second feature that makes this 'Swedish model' particularly interesting to examine is the wide array of different options among which unemployed individuals can potentially choose. Since different programmes may in fact have heterogeneous effects, a natural question in such an institutional context concerns the relative performance of the various programmes. This provides the motivation for Chapter 3, which quantifies the relative effectiveness of different types of programmes, ideally aiming at singling out the best-performing ones. General lessons as to which type of programme is more effective could in fact be shared across countries, in particular those who have recently been expanding their active labour market policies (cf. e.g. Martin and Grubb, 2001).

The interactions between the unemployment benefit system and the programme system are a distinctive Swedish feature which should not be ignored when assessing the programmes' effectiveness. In fact, while in principle capped at 60 weeks, eligibility to unemployment compensation can be renewed by participating in a programme, making it *de facto* possible to indefinitely extend the period during which unemployed individuals can receive benefits.² A labour market programme in Sweden thus effectively comes as a bundle of two conflicting components: while intended to equip job-seekers with marketable skills which should facilitate their re-employability, it allows to renew eligibility to generous unem-

² This link between the passive and active components of the Swedish labour market policy was severed in February 2001.

ployment compensation, thus reinforcing the work disincentive associated with the unemployment insurance system. The next chapter looks at these issues in considerable detail, while Chapter 3 restricts its analyses to job-seekers entitled to benefits, that group whose participation incentives are most likely to be affected by the intertwined unemployment compensation-programme system.

Further general interest in the Swedish case arises from a few features of its institutional set-up which raise several methodological and modelling issues, which while not previously addressed in the typical US programme evaluation literature are shared by several Western European labour market policies.³

The object of the evaluation is a system of ongoing programmes which are open to all registered job-seekers. Unemployed individuals in turn can – and in fact often do – register repeatedly, and they can participate in various programmes at different times during their observed unemployment history. More important still is the fact that even when focusing on individuals who have just entered unemployment, it can in general be claimed that they will join a programme at some future point, provided they remain unemployed 'long enough'; in fact, a stylised representation of the Swedish system is that if unemployed individuals are not observed to go into any programme, it is *because* they have already found a job. Since all individuals will eventually enter a programme when they have experienced a sufficiently long period of open unemployment, there is no obvious control group whose experience can be used to derive the counterfactual. These difficulties in the selection of the group of 'non-participants' are further compounded by the fact that all the programmes take place continuously over time, so that individually differing starting dates make the time before and after the programme well defined only for a given participant.

The particularly rich and highly representative administrative dataset from which this analysis of the Swedish labour market programmes benefits may finally give rise to additional general interest in the results.

Often in the literature programme effects are evaluated at a given – and arbitrary – point in time (e.g. on the last observation day, or after a year). By contrast,

³ Some Western European countries' labour market programmes share a few of the Swedish features (in particular, a variety of ongoing measures); see e.g. the recent work by Lechner (e.g. 1996, 1999a, 1999b) and Gerfin and Lechner (2000).

being able to follow up individuals for a relatively long time (five to six years) allows us to capture both short and long-term effects. In addition, not only does the data record the programme, unemployment and unemployment benefit receipt history (to the day) of all unemployed individuals registered at public employment offices, but it also includes a wide array of demographic and human capital variables, information on previous wage and working hours, as well as the caseworker's appraisal of various factors relating to the overall situation and needs of service of the job-seeker.

The end result is thus a very large (110,000 individuals), representative (over 90% of the unemployed register at employment offices) and rich dataset, which permits both the short- and long-term evaluation of the programmes and with respect to a larger number of outcomes than is generally possible.

Despite the richness and thoroughness of the administrative data just described, the unemployment register does however suffer from an attrition problem, resulting in unemployed individuals being de-registered with the reason recorded as 'contact ended'. Such 'lost' individuals are problematic, in the sense that the researcher is prevented from fully observing their true labour market status: which of these spells are in reality an employment spell the former unemployed individual failed to report to the agency, and which are by contrast still part of the preceding unemployment spell? The problem is not just how severely under-reported employment status may be in the data, but also the fact that such under-reporting may systematically differ between programme participants and comparison group members. Both of the following chapters are thus careful in checking the results on programme employment effects in terms of their robustness to these lost spells.

Before moving on to the two main empirical analyses, it may be worth reiterating the important issues that need to be taken into account when setting up the evaluation framework or interpreting the results: benefit entitlement considerations should not be overlooked when assessing the past effectiveness of the Swedish programmes; both analyses and interpretation of the results need to take account of the non-standard definition of the 'comparison group' in the Swedish context, and some robustness checks should try to take account of the 'lost' individuals. This last issue aside, the available data is very representative and includes a wide array of demographic and human capital variables, together with information on unemployment history and unemployment benefit receipt, as well as the caseworker's appraisal of various factors relating to the overall situation and needs of service of the job-seeker. This unusual informational richness by international standards has motivated the matching approach on which the analyses in the following two chapters are based.

CHAPTER II

Overall performance of the Swedish System of Active Labour Market Programmes in the 1990s

II.1 Introduction

Sweden has traditionally enjoyed low unemployment rates by European standards, a successful macroeconomic performance which several observers have related to the country's extensive offer of labour market programmes. Labour market programmes are a large-scale undertaking¹, representing a sizeable investment for the Swedish government, which spends over 3% of GNP on such measures, compared to 2.1% in Germany and 0.4% in the US (Forslund and Krueger, 1995, Table 1). The unprecedently severe recession that hit the Swedish economy at the beginning of the 1990s has however posed new challenges to the Swedish labour market policy.

This chapter investigates how successful the 'Swedish model' of active labour market programmes has been in the context of the high unemployment atypically experienced by Sweden in the 1990s.

In the presence of both rising unemployment and expanding budget deficit², the economic justifiability of the expense of the Swedish programmes does in fact hinge on the assessment of their actual effectiveness.^{3 4} Secondly, it may well be that the bundle of measures that was effective in tackling a certain kind of unemployment is no longer so successful in a different environment. Since the underlying causes of

¹ In 1997, for instance, the equivalent of 4.5% of the labour force participated on average in the programmes (excluding measures for the disabled). In each month of the same year, 10% of the total yearly stock of unemployed were enrolled in the most expensive programme, labour market training.

 $^{^{2}}$ In 1993 Sweden's budget deficit at 13% of GDP is among the highest in the Western world, while the national debt has reached 80% of GDP.

³ Ideally, even if positive net (social) benefits from the programmes were found, one would have to show that this was the best outcome achievable from the resources invested.

unemployment are diverse and most likely to vary among target groups, geographical regions as well as over time (see Schmidt, 2000), the developments in the last decade seem to cast doubts as to the universal cure-all nature of the 'Swedish model'.

Throughout, special attention is devoted to the distinctive Swedish feature whereby, up to February 2001, participation in any labour market programme for five months would qualify for a renewed spell of unemployment compensation, thus likely to reinforce the work disincentives associated with the benefit system.

The unusual richness of the available data allows us to identify a larger number of destination states and outcomes than generally available. Programmes will thus be evaluated on several dimensions, in particular in terms of their effectiveness in helping participants find a job faster; in endowing them with skills and good working habits to enhance their employment prospects; in fostering the further acquisition of human capital in the regular education system; in leading to repeated participation in subsequent programmes; or in providing incentives for participants to alternate between compensated unemployment and eligibility-renewing programme spells.

The chapter proceeds as follows. After an outline of the Swedish labour market policy in Section II.2, Section II.3 describes how the data used captures such an institutional framework. Section II.4 highlights the main features of propensity score matching in relation to the general evaluation problem, whereas Section II.5 explains how these statistical matching techniques have been adapted to the specific Swedish context. Section II.6 presents the set of empirical results, including some robustness analyses aimed at addressing the problem of the unobserved outcome variable for the group of individuals 'lost' due to attrition problems in the available database. Section II.7 focuses on unemployment-programme cycling behaviour, before drawing all the findings together in the last section of this chapter.

⁴ Research has been increasingly carried out in this direction. See the extensive review by Calmfors, Forslund and Hemström (2001).

II.2 The Swedish labour market policy

The Swedish labour market policy has two components: a benefit system that supports individuals while unemployed and various active labour market programmes offered in order to improve the opportunities of unemployed workers.

Unemployment compensation is provided in two forms, the most important one being unemployment insurance (UI). UI benefits are very generous – the incomerelated daily compensation is 80% of the previous wage⁵ – and are available for a long duration – 60 calendar weeks, more than twice the maximum duration of unemployment benefits in the US. An (even part-time) unemployed person registered at a public employment office and actively searching for a job is eligible for unemployment benefits if in addition to a membership condition⁶, the work condition is satisfied: the claimant must have been working for at least five months during the twelve months preceding the current unemployment spell. Until 1996, a 5-month participation in practically any labour market programme would count as employment in allowing participants to become eligible for their first time. In addition, an offer of 'suitable' work – or of a labour market programme – must be accepted; refusal to accept a job/programme might lead to expulsion from compensation.

The second form of unemployment compensation is cash labour market assistance (KAS). This supplementary compensation system has been mainly designed for new entrants in the labour market who usually are not members of any UI fund. Daily taxable benefits are significantly lower than UI (around half) and are paid out for half the UI period (30 calendar weeks)⁷ and claimants are subject to a work condition similar

⁵ This maximum level of compensation has changed a few times during the 1990s; from 90% of the previous wage, it was reduced to 80% in July 1993, then further to 75% in January 1996, before being restored to 80% in September 1997. Note however that the system has a ceiling in terms of the amount of daily compensation.

⁶ The claimant must have paid the (almost negligible) membership fees to the UI fund for at least 12 months prior to the claim.

⁷ The maximum durations reported for both UI and KAS benefits refer to individuals below 55 years. Claimants aged over 55 are entitled to 90 weeks of UI and 60 weeks of KAS; if aged over 60, to 90 weeks of KAS.

to the one for UI, which can however be replaced by the education condition of having finished at least one year of school in excess of the nine compulsory ones.

Turning now to the labour market programmes, their stated overall purpose is to prevent long periods out of regular employment and to integrate unemployed and economically disadvantaged individuals into the labour force. There are various kinds of programmes available, ranging from labour market retraining to public sector employment such as relief work, to subsidised jobs, trainee replacement schemes, work experience schemes and job introduction projects, to programmes targeted at specific groups, such as a variety of special 'youth measures', special measures targeted at the disabled, or self-employment and relocation grants. The different types of programmes may thus variously benefit participants by either facilitating their job search, augmenting their human capital with formal teaching or by providing them with job experience, improving their working habits or offering a cheap way for employers to screen their productivity. Most programmes have a maximum duration of six months, though participants stay an average of four months.

The passive and active components of the Swedish labour market policy just outlined used to be closely linked: up to February 2001, participation in any labour market programme for five months would count as employment and thus qualify for a renewed spell of unemployment compensation. Thus despite the fact that the period during which an unemployed job-seeker can receive unemployment benefits is fixed, it used to be in fact possible to extend it indefinitely by using programme participation as a passport to renewed eligibility.

A peculiar type of 'programme' (and one which does not allow to renew eligibility) should be mentioned before concluding. Simply being registered as openly unemployed at an employment office gives access to the various employment services offered by the offices, not only in terms of the increasingly computerised job information and matching of vacancies to applicants, but also in terms of the 'job-seeker activities', which include search-skill-enhancing activities such as training courses on how to apply for a job and motivation-raising activities. In Sweden, the state to which programme participants can be compared to is thus not one of being completely abanabandoned to fend for oneself, but the baseline services offered by the employment offices. Since job services are expected to enhance job search intensity and efficiency rather than human capital or working habits, it might be possible that they are more effective than standard programmes in reducing unemployment duration, but less so in prolonging job retention. Finally, if such job services turn out to be more effective than standard programmes, they will most likely be more cost-effective as well.

II.3 Data and sample selection

The dataset that has been constructed to capture the institutional framework described in the previous section is the result of combining two main sources, which reflect the programme component (Händel) and the benefit component (Akstat) of the labour market policy.

Händel is the unemployment register, of which the various databases contain information on all unemployed individuals registered at the public employment offices. This longitudinal event history dataset, maintained by the National labour Market Board (AMS) and available from 1991, provides each individual's labour market status information over time (e.g. unemployed, on a given programme, temporarily employed), together with important personal characteristics of the job-seeker and of the occupation sought. The information regarding the reason for ending the registration spell (e.g. obtained employment, gone on regular education or left the workforce) has been used to impute the individual's labour market status in between registration periods.

Akstat, available from 1994, originates from the unemployment insurance funds and provides information on individuals entitled to UI or KAS, in particular on the amount and type of compensation paid out, on previous wage and working hours.

The end result is a very large, comprehensive and representative⁸ dataset, which in

⁸ Unemployment individuals entitled to compensation are required to be registered at a public employment office in order to collect their benefits, and only registered individuals have access to the

addition to being extremely rich in providing information (to the day) about the duration of stay in a labour market state, includes the necessary array of demographic and human capital variables together with information on unemployment benefit recipiency and type of entitlement.

From the original dataset, a subset of 116,130 individuals has been selected who became unemployed for their first⁹ time in the same calendar year 1994 (when unemployment was still at its highest – cf. Figure 1.1). In particular, given that the main purpose of the programmes is to enhance the re-employability of the unemployed, those registering as employed or directly entering as programme participants are excluded from the sample. All our individuals are in the 18-55 age group, have no occupational disabilities and are followed until the end of November 1999.

II.4 The evaluation problem and propensity score matching

The prototypical evaluation problem¹⁰ can be fruitfully framed within the potentialoutcome approach.¹¹ Object of the evaluation is the causal effect of a treatment of interest (treatment 1), relative to another treatment (treatment 0), on the outcome or response variable *Y*, which is experienced by individuals in the population of interest. Let *Y*₁ be the outcome that would result if the individual were exposed to treatment 1 and *Y*₀ the outcome that would result if the same individual received treatment 0. The binary indicator of the treatment actually received is $D \in \{0,1\}$. For a given individual

programmes. In fact, over 90% of the unemployed do register at an employment office (from a validation study by Statistics Sweden, quoted in Carling, Edin, Harkman and Holmlund, 1996, Footnote 7), making our employment register-based dataset quite representative of the population of interest.

⁹ Strictly speaking, one cannot exclude that our individuals have had contact with the unemployment office before August 1991, date when Händel starts.

¹⁰ Standard references in the evaluation literature include the comprehensive work by Heckman, LaLonde and Smith (1998), as well as Heckman and Robb (1985), Heckman, Ichimura and Todd (1997, 1998), Heckman, Ichimura, Smith and Todd (1998) and Rosenbaum and Rubin (1983, 1985).

¹¹ Though the concepts behind the framework have been around since Fisher (1935) and Neyman (1935), Rubin (1974) formally applied them to the study of causation; see however also the work by Roy (1951) and Quandt (1972).

i, the actually observed outcome is then $Y_i = Y_{0i} + D_i(Y_{1i} - Y_{0i})$.¹² Finally, let *X* be the set of attributes (i.e. characteristics not affected by the treatment, such as time-invariant or pre-exposure variables).

The parameter which receives most attention in the evaluation literature is the 'average treatment effect on the treated' $E(Y_I-Y_0/D=1)$.¹³ In our case, this amounts to addressing the question of how the post-programme outcome of unemployed workers participating in a Swedish programme compares to how they would have fared had they not taken part in the programme, on average.

For participants we do observe Y_I , the outcome after the treatment, so that the average observed outcome for participants is an unbiased estimate of the first component of the effect of treatment on the treated, $E(Y_I/D=1)$. We do not however observe participants' Y_0 , the outcome they would have experienced had they not participated. At the core of the evaluation problem is thus the attempt to estimate missing data – in this case the counterfactual $E(Y_I/D=0)$ – to overcome the 'fundamental problem of causal inference' (Holland, 1986a) that no individual can be in two different states at the same time.

A general issue is that treated individuals may not be a random sample of the population, but they may receive treatment on the basis of characteristics which also influence their outcomes. One approach to construct a suitable comparison group is based on statistical matching, whereby a comparison group from all non-participants is chosen such that the selected group is as similar as possible to the treatment group in terms of their observable characteristics.¹⁴ When the relevant differences between

¹² At this very general stage, the stable unit-treatment value assumption has to be made. SUTVA (first expressed by Rubin, 1980 and further discussed in Rubin, 1986 and Holland, 1986b) requires that an individual's potential outcomes depend only on his own participation, not on the treatment status of other individuals in the population (thus ruling out cross-effects or general equilibrium effects) and that whether an individual participates or not does not depend on the participation decisions of others (e.g. thus excluding peer effects in the participation decision).

¹³ For the average effect on the non-treated $E(Y_I - Y_0 | D=0)$, a procedure symmetric to the one discussed below applies; while the average treatment effect $E(Y_I - Y_0)$ is simply a weighted average of the other two parameters. In the special case where the effect of the treatment is homogeneous among individuals, the three effects would coincide.

¹⁴ Simple matching estimators, for instance, pair each treated unit to an observably similar non-treated unit; smoothed versions associate to the outcome Y_i of treated unit *i* a 'matched' outcome given by a weighted average of the outcome of more (possibly all) non-treated units, where the weight given to

any two individuals are captured in their observable attributes (where 'relevance' is in terms of their potential outcomes), matching methods can yield an unbiased estimate of the treatment impact. This underlying identifying assumption, henceforth CIA ('conditional independence assumption', also called 'ignorable treatment assignment' or 'selection on observables') requires the evaluator to have access to a set of conditioning observed attributes *X* such that, for a given value of the *X* vector, the distribution of the (counterfactual) outcome *Y*₀ in the treated group is the same as the (observed) distribution of *Y*₀ in the non-treated group, in symbols: $Y_0 \perp D \mid X$.¹⁵

Formally, the required counterfactual is identified under CIA as follows:

$$E(Y_0 | D = 1) = E_X [E(Y_0 | X, D = 1) | D = 1]^{ClA} = E_X [E(Y_0 | X, D = 0) | D = 1]$$

= $E_X [E(Y | X, D = 0) | D = 1]$

For the matching procedure to have empirical content, it is also required that P(D=1|X)<1 over the set of X values where we seek to make a comparison, which guarantees that all treated individuals have a counterpart in the non-treated population for each X for which we seek to make a comparison.¹⁶

A practical problem in implementation arises if the X's are highly dimensional, and especially if they contain continuous variables (which would have to be discretised). A very useful variable to avoid the curse of dimensionality is the 'propensity score' $e(x) \equiv P(D=1 | X=x)$ – the conditional probability of participation given the value of a vector of observed characteristics *x*. As Rosenbaum and Rubin (1983) show, by definition treatment and non-treatment observations with the same value of the propensity score have the same distribution of the full vector of regressors *X* (the balancing property of the propensity score: $X \perp D \mid e(X)$). It is thus sufficient to match exactly on the propensity score – a single variable on the unit interval – to obtain the

non-treated unit j is in proportion to the closeness of the observables of i and j (see Heckman, Ichimura and Todd, 1997 and 1998 and Heckman, Ichimura, Smith and Todd, 1998).

¹⁵ For the average effect of treatment on the treated, the weaker version of the CIA in terms of conditional mean independence suffices: $E(Y_0 | X, D=1) = E(Y_0 | X, D=0)$.

¹⁶ If there are regions where the support of X does not overlap for the treated and non-treated groups, matching has to be performed over the common support region; the estimated treatment effect has then to be redefined as the mean treatment effect for those treated falling within the common support.

same probability distribution of *X* for treated and non-treated individuals in matched samples. Consequently, if CIA holds conditional on *X*, it will also hold conditional on the propensity score: $Y_0 \perp D \mid e(X)$.¹⁷

In a model-free environment, matching methods are general enough to allow for heterogeneous individual treatment effects without suffering from two of the three selection-bias sources identified by Heckman, Ichimura, Smith and Todd (1998): the bias due to the difference in the supports of X in the treated and control groups (failure of the common support condition) and the bias due to the difference between the two groups of the distribution of X over its common support. The remaining source of bias is the one due to selection on unobservables. Arguing its importance amounts to arguing the inadequacy of CIA for a particular problem; this should be done on a case-by-case basis, with account being taken of the informational richness of the available dataset (i.e. X) and the institutional set-up where selection into the treatment takes place. A discussion of how likely the basic matching assumption is to be fulfilled in our case is postponed to Section II.5.3.

II.5 Matching in the Swedish institutional set-up

II.5.1 Treated and non-treated in Sweden and other key choices

As mentioned in the Introduction to Part One, the Swedish institutional set-up poses a few interesting methodological issues which have to be resolved before applying matching techniques. Object of our evaluation is a system with a wide array of different ongoing programmes, which take place continuously over time and are open to all registered job-seekers; unemployed individuals in turn can – and in fact often do – register repeatedly, and they can be treated at different times during their observed unemployment history. In such a context, several choices have to be taken, including

¹⁷ This result holds for a more general 'balancing score' b(X), a function of X such that $X \perp D \mid b(X)$. The finest and most trivial balancing score is X itself, while the propensity score is the coarsest one.

how to actually define the treatments, as well as the treated and the non-treated individuals.

Since object of the analysis are those individuals registering as unemployed for their first time, the focus is on the *first* treatment individuals may receive within their first unemployment experience, so that any subsequent programme participation is viewed as an outcome of that first treatment. Furthermore, the Swedish active labour market policy is considered in its totality: the aim of the present study is to shed some light on the *overall* effectiveness of the Swedish unemployment system, a system comprising both a collection of different programmes and a closely intertwined unemployment benefit component. All the various types of programmes are consequently aggregated into one 'programme', so that treatment 1 is any programme in which a first-time unemployed can participate. The next chapter will go a step further and disaggregate treatment 1 into its main components in order to look at the differential effectiveness of the various programmes.

The definition of treatment 0 is however considerably less straightforward.¹⁸ The usual specification of the evaluation problem envisages a situation where individuals are either treated – participate in a programme – or not treated – do not participate in any. In the Swedish institutional context, however, an unemployed individual will – in principle – be treated eventually, if he is unemployed long enough. In fact, bringing this reasoning to its limit, one could argue that the reason an unemployed individual will – ual has not been observed to go on a programme is *because* he has found a job (before).

When considering individuals entering unemployment, it is likely that unemployed job-seekers (and programme officers) will take their decisions sequentially over time in unemployment, so that the choice open to them is not whether to participate or not to participate *at all*, but whether to participate now or not to participate *for now*. An unemployed individual may decide to not to join any programme for now, in the knowledge that the opportunity to participate is potentially only postponed, not for-

¹⁸ The discussion of an absent 'no-treatment' group was initiated by Carling and Larsson (2000a, b).

saken for all times. For those finding a job before being treated, the waiting choice has proved successful.

The solution to the absence of an obvious control group is based on noting that the relevant decision problem faced by all those who are currently unemployed is not whether to join a programme, but when to join: now or – *provided still unemployed* – later. When looking at individuals flowing into unemployment, the effect one can thus try to assess concerns the possible benefits from joining a programme at a given time compared to not joining any programme at least up to then, or in other words compared to waiting in open unemployment at least some time longer, knowing that one will always be able to join a programme later on.

While different from the conventional evaluation question, formulating the problem this way offers several advantages. From a practical point of view, it allows us to address an effect which is at least in principle identifiable. From a conceptual point of view, it lets the evaluation question mirror the relevant decision problem faced by the target group. Finally, it is informative about the usual question; if it were to be found that those who join a programme now do worse than those who wait (and possibly join later), the implication is that it is better to wait. In the limit, it is better to wait forever, which is equivalent to say that joining a programme has no positive effect.

The average effect of joining now rather than later or never will be evaluated with respect to a series of outcomes, ranging from the duration of the first unemployment experience (Section II.6.5), to the probability of being in a number of labour market states over time (unemployment, in particular of the benefit-compensated type, employment, subsequent programme participation, inactivity and further studies in the regular education system – Sections II.6.2-6.4; and cycling between unemployment compensation and programmes – Section II.7). When interpreting this results it is always important to keep in mind that the chosen comparison group does not reflect a no-programme state, but rather a possibly postponed participation. Note also that as discussed at the end of Section II.2, an essential ingredient of this kind of treatment 0 is the baseline set of services offered by the employment offices to individuals searching for a job as openly unemployed.

A final difficulty arises from the fact that we are considering a system of ongoing programmes, taking place continuously over time and with individually differing starting dates (both in calendar terms and in relative terms, i.e. in terms of time since registration at the office). This implies that potentially important variables related to the distance in time to the beginning of the programme are only defined for individuals observed to join a programme. Probably the most crucial of these variables is the unemployment duration prior to entering a programme.

Since one needs to be (still) unemployed to be enrolled into a programme, all potential comparisons to a treated individual should have at least reached a stage where they could have joined the programme. Secondly, elapsed unemployment duration is likely to capture some important unobservables; since average unobserved 'ability' (standing for any unobserved skill linked to an individual's ability to exit an unemployment spell) declines over unemployment duration, a participant's observed time spent in (open) unemployment before entering a programme reveals an upper bound on his expected unobserved ability. To avoid a clear source of bias, then, a given treated should not be paired to a non-treated individual who has remained openly unemployed for a shorter duration than has the participant at entry into the programme. Thirdly, there are some (albeit loose) regulations, as well as some particular incentives which link elapsed unemployment duration to programme participation¹⁹; unemployment duration is thus an important X variable for explaining not only subsequent outcomes as discussed above, but also the participation decision (D=1). Finally, in the presence of individually differing starting dates one needs to set a relative time scale to begin measuring the effect of the programme. Matched non-participants need to be assigned a hypothetical unemployment duration prior to the programme, which splits their observed unemployment duration into a pre-programme attribute and an outcome sub-spell. The interrupted unemployment duration of their respective matched participants (i.e. their time to the programme) would seem an obvious candidate.

Ideally, we would then match on T^{l} , the duration of the first unemployment spell interrupted by the programme. The obvious problem is that T^{l} for the non-treated is an unobserved counterfactual (i.e. their waiting time before the start of a programme had they entered one) and thus cannot be included in the estimation of the propensity score. Still, for the reasons expounded above, it is essential to impose that all comparison individuals eligible to be matched to a given treated unemployed for T^{l} before entering the programme have reached an unemployment duration of at least T^{l} : this ensures that both treated and control have a 'complete'²⁰ unemployment duration T^{0} at least as long as T^{l} .

II.5.2 Methodology

The requirement in terms of unemployment duration can be fulfilled in quite a natural way when implementing our version of the average effect of treatment on the treated, i.e. the average effect, for those observed to join a programme, of joining when they did compared to waiting longer than they have.

The causal inference problem to be addressed here can be formalised as follows. Let $D \in \{J, W\}$ denote the treatment indicator, where *J* stands for joining and *W* for waiting. The response variable is Y(i, D, t), the labour market status of individual *i* at time *t* if individual *i* were exposed to treatment *D*. To simplify the notation, in the following the fact that the response is a function of time rather than simply a scalar is ignored and the short-hand notation Y_J and Y_W is used to denote the potential outcomes corresponding to joining and waiting, respectively. The mean effect of treatment on the treated we aim to estimate is thus $\tau \equiv E(Y_J - Y_W | D=J)$.

¹⁹ Some programmes for instance formally require 4 months of open unemployment prior to enrolment, while approaching unemployment benefit exhaustion may make individuals more likely to enter a programme.

²⁰ 'Complete' refers to the unemployment duration an individual would experience were he not to join a programme. T^0 is thus observed for individuals whose unemployment spell ends 'naturally', i.e. in employment or exit from the labour force (nobody is censored for the full 6-year observation period), while T^l is observed for individuals joining a programme; for the latter, we know that $T^0 \ge T^l$.

This complex problem is subdivided into a sequence of M simpler problems by discretising unemployment duration and rewriting τ as:

$$\tau = \sum_{m=1}^{M} E(Y_J - Y_W | D = J, T^1 = m) \ Pr(T^1 = m | D = J)$$

In concrete terms, the sample is stratified by (discretised) unemployment duration. For each m = 1, ..., M we then calculate τ^m , the effect of treating an unemployed individual in his m^{th} month of unemployment compared to not treating him at least until the end of his m^{th} month (in other words, compared to waiting at least m months):

$$\tau^m \equiv \mathrm{E}(Y_{J(m)} - Y_{W(m)} \mid D^m = 1)$$

where

- { $Y_{J(m)}$, $Y_{W(m)}$ } is subsequent labour market status, respectively conditional on $J_{(m)}$ joining a programme in one's m^{th} month and on $W_{(m)}$ waiting longer than m months; and
- $D^m=1$ if D=J and $T^l=m$, i.e. for those joining a programme in their m^{th} month of unemployment.

Note that the treated group has in fact been divided into M exhaustive and mutually exclusive sub-groups (defined in terms of pre-programme unemployment duration): $\{D=J\} = \bigcup_{m=1}^{M} \{D^{m}=1\}$, so that the various τ^{m} 's can be viewed as the treatment effects for the different subgroups.

The form of CIA we require to identify τ^m is:

$$Y_{W(m)} \perp D^m \mid X = x, T^0 \ge m$$
 (*m*=1, ..., *M*).

In words, given a set of observed characteristics x and having reached at least the same duration in unemployment²¹, the distribution of $Y_{W(m)}$ for individuals joining a programme in their m^{th} month is the same as the one for individuals deciding to wait longer. Say two individuals are observed, with the same characteristics and who have

²¹ Note that we have in fact succeeded in ensuring that for *both* a given treated and his matched non-treated, the complete unemployment duration T^0 is larger than $T^1(=m)$.
reached the same unemployment duration, one of which joining a programme now, while the other deciding to wait longer. For the control, the waiting decision may prove successful, in the sense that he will subsequently find a job or de-register for other reasons. Alternatively, he may later decide – or be forced – to join a programme. What the CIA requires in such a context is that the probability distribution of such outcomes is the same for the observably-similar treated individual had he then decided to wait longer as well.

The parameter of interest τ can finally be calculated as a weighted average of the τ^{m} 's, weighted according to the observed month of placement (T^{l}) distribution of the treated:

$$\tau = \sum_{m=1}^{M} E(Y_{J(m)} - Y_{W(m)} | D^{m} = 1) Pr(D^{m} = 1 | D = J) = \sum_{m=1}^{M} Pr(D^{m} = 1 | D = J) \tau^{m}$$

In implementation, we set M=18, so that what we will be looking at is the impact of entering a programme within one and a half year of first registration; 94% of all treated are however observed to enter a programme within such a time span.²²

This methodology has the rather peculiar feature that in the calculation of the overall effect a given treated individual (joining a programme in his m^{th} month) may also act as control (being a potential control for (m-1) groups of treated). In fact, the *relative* time scale, which is implicit in the set-up of the problem and central in the way it has been addressed, makes it quite intuitive to think of him as different persons, whose contributions start at different origins (i.e. at different T^{l} 's).²³

Following this procedure also allows us to assess whether there is a differential programme impact according to T^{l} , i.e. to the time the individual has spent in unem-

²² See Appendix B for the sample sizes of the two sub-groups by unemployment duration and Appendix A for a detailed description of the matching protocol used.

²³ In particular, a given treated individual counts as one treated person, whose contribution starts being evaluated at the moment he enters the programme (i.e. from *his* T^{l}) and may count as control person for 'otherwise similar' treated individuals who have joined a programme before him. In this latter case, his outcome represents the waiting counterfactual outcome for his matched treated individuals, where evaluation begins when the matched treated start their respective programme (i.e. from his matched treated individuals' T^{l} 's).

ployment before joining the programme. A very interesting group in this respect is the one observed to enter a programme exactly at benefit exhaustion.²⁴

Additional methodological considerations

In nearest-neighbour matching with replacement (and within caliper), each participant is matched with replacement to that non-participant with the closest propensity score; treated individuals for whom a match close 'enough' has not been found are discarded. This type of matching focuses on minimising bias alone (the large samples available and the size of the effects uncovered have essentially dictated this focus). One-to-one matching typically involves an efficiency loss, since only the participant and his closest neighbour are used, instead of a larger number of close neighbours. By contrast, when more than one comparison unit is assigned to any treated (e.g. in kernel-based matching), the variance is reduced but at the possible cost of an increased bias. Similarly, using the same non-treated individual more than once can improve matching quality²⁵, but it increases the variance, which has to be adjusted accordingly: the more times a non-treated observation is used, the larger the related standard error of the estimated effect. The standard errors should also adjust for the additional sources of variability introduced by the estimation of the propensity score as well as by the matching process itself; while analytical results have been derived for kernel matching by Heckman, Ichimura and Todd (1998), the common solution for one-toone matching is to present bootstrapped confidence intervals.

A final methodological consideration concerns the choice of the origin on the relative time scale. Having matched on the precondition that the comparison individuals have to have been unemployed for at least the number of months it took their unemployed treated counterpart to get into the programme, it is quite natural (cf. the dis-

²⁴ Note that this amounts to assessing if the programme effect for those individuals who join a programme after *m* months in unemployment is better or worse than the programme effect for the k^{th} month joiners; not whether joining a programme after *m* months leads these participants to experience better or worse outcomes than if they had joined after *k* months.

²⁵ In their application, Dehejia and Wahba (1998, 1999) do in fact find the performance of simple matching with replacement very satisfactory compared to more complex extensions or methods with-out replacement.

cussion in II.5.1) to set the origin t=0 for matched comparisons when their respective treated begin their programme. For the treated, however, the choice affects what treatment is and when its effect begins. Should one start to evaluate the performance of the treated when they enter or when they exit the programme? Since the observed programme duration should be viewed as endogenous²⁶, and in the light of the series of timing sub-questions asked (what is the effect of entering a programme *now*), it seems natural to set t=0 for the treated too at the time when they *join* their programme. In the following analyses, then, treatment 1 is 'starting/being assigned to a programme', also commonly referred to in the literature as the 'intention to treat'. The causal effect of a programme starts to work with the beginning of the programme, so that any lock-in effect of the programme is viewed as a constituent part of the effect.

Estimation of the propensity score

A series of M=18 probits has been estimated, each one modelling $e(X; m) \equiv Pr\{D^m=1|X, T^0 \ge m\}$, the probability of joining a programme *conditional* on having reached an unemployment duration of $m \in \{1, 2, ..., 18\}$ months.²⁷ All the variables are measured at first entry into unemployment, except those relating to the present unemployment spell, which are calculated at 'entry' into the programme (where for the non-treated, 'entry' refers to the month of unemployment by which they have been stratified). A rather flexible specification has been allowed for by including all relevant interaction terms and a quadratic for age.

Overall, matching on the estimated propensity score balances the relevant variables between the treated and the selected controls quite well.²⁸

²⁶ Participation in some programmes requires that the individuals continue job-searching activities while participating; the offices too continue to search for them, since participants are still registered and requested to be 'at the labour market disposal'; and individuals are required to drop out of a programme if a 'suitable' job is found for them.

 $^{^{27}}$ This approach is equivalent to a discrete hazard model, with all the estimated parameters allowed to be duration-specific.

²⁸ Indicators suitable to assess the remaining 'distance' of the marginal distributions of the relevant characteristics in both groups such as the median and mean absolute standardised bias (median and

II.5.3 Plausibility of the CIA: Selection into the Swedish programmes and available information

The version of the CIA the following analyses rely upon requires to control for all those variables *X* that, conditional on having spent a given amount of time in unemployment, influence both the participation decision as well as the potential outcomes that would occur where such decision to be postponed. What is required is thus that conditional on all the information observed, the fact that an unemployed individual goes into a programme this month while another waits longer is not correlated with the outcome the individual joining now would have experienced had he not entered the programme when he did, in particular in terms of employment prospects, but also of future programme participation.

From work by Harkman (2000, as reported in Carling and Richardson, 2001) it appears that an unemployed individual's decision to participate in any programme or not to participate may depend on the individual's subjective likelihood of employment. Although unobserved to us, we are however able to control for several factors which may be highly correlated with it. In particular, several pieces of information are used to capture and characterise the employment history of the individuals under examination, the variable identified to be the most important one (even more than earnings dynamics) by Heckman, Ichimura and Todd (1997). All our individuals register at the unemployment office for their first time²⁹, so their only (recent) unemployment experience relates to the present unemployment spell. Conditional on observables, elapsed unemployment duration provides an upper bound to average unobserved ability and other important unobservables (e.g. perceived deterioration of human capital, stigma effect, loss of hope or motivation). In addition, controlling for time spent in unemployment in conjunction with information regarding the entitlement status of an individual is crucial, in that approaching benefit exhaustion would make an individual more likely to join a programme or, if having to wait longer, more

mean taken with respect to all the variables) are as low as 0.7 and 0.9% respectively (cf. Rosenbaum and Rubin, 1985). For more details, see Appendix C and D.

²⁹ At least since the beginning of the Händel dataset, in August 1991.

likely to enter a programme later on or to intensify his job search (or lower his reservation wage). Entitlement status also controls for the degree of labour market attachment due to the work requirement UI-recipients have to fulfil. Similarly, part-time unemployment spells denote individuals who are still maintaining contact with the regular labour market and are probably both subject to less human capital depreciation and in a better position to look for a (full-time) job, by exploiting their bargaining position, additional contacts and references. A subjective indicator of experience for the profession sought (none, some, good) is another interesting piece of information provided in the present dataset.³⁰ Difference in prior work experience is important since it results from both observed and unobserved differences between characteristics of the treated and non treated (cf. Ham and LaLonde, 1996). This indicator can be viewed as a summary statistics of the amount – as well as effectiveness, transferability and obsolescence – of previous human capital accumulation, on-the-job training and learning-by-doing.

Similarly, we have controlled for factors relating both to employment prospects and either to potential returns from programme participation or affecting the opportunity cost or psychological cost of participation (age, gender, previous stock of human capital in terms of both specific and general education and job-specific experience, occupation being sought, citizenship, part-time unemployment status).

The Swedish unemployment offices are characterised by decentralisation, which gives job officers quite a large degree of freedom. Also, although in general an unemployed job-seeker must be willing to participate, this may not always apply to individuals receiving unemployment compensation; for them, the proposal of a programme can be used as a 'work test', the turning down of which may entail suspension from benefits. To capture this selection by caseworkers, entitlement status is thus again of critical importance, but additional useful information has also been included which relates to an overall evaluation by the officer of the situation and character of the unemployed job-seeker – if already part-time employed, if looking for a part-time job, if willing to move to another locality, if judged to be able to take a job immedi-

³⁰ Looking at the incidence of the three levels by age reveals that it is in fact quite a reliable indicator.

ately, or to be in need of guidance, or to be difficult to place. Such individual traits are potential indicators of unobserved heterogeneity and are quite likely to affect the joining decision as well as the counterfactual outcomes in terms of subsequent participation or employment probability. It is interesting to note that the caseworker may update this subjective judgement during his client's unemployment spell, and that this time variation is captured and exploited in estimation.

A final issue relates to the gradual shift towards more decentralised decisionmaking as to labour market programmes that has taken place in Sweden in the second half of the 1990s and to the concomitant emergence of new financial incentives (cf. Lundin and Skedinger, 2000): municipal budgets may be favourably affected by moving unemployed individuals from social assistance (funded by the local authorities) to programmes (financed by the central government); some programmes (e.g. relief work) may subsidise labour in the services typically provided by the local authorities; and programmes may serve as a means of maintaining the local municipal tax base, by reducing migration among the unemployed. Thus besides the local variable denoting the county, we have constructed two indicators at the individual's municipality / employment office level³¹ over time to further control for the possibility that individual and/or case-worker joining criteria may be based on local unobserved characteristics in turn correlated with individuals' potential labour market performance. The local 'programme-rate' is given by the number of *participants* in all programmes as a proportion of all individuals registered as openly unemployed or as programme participants at the individual's municipality the month he enters the programme. This information relates to the local programme capacity, e.g. in terms of slots available. By contrast the local 'offer-rate', representing the proportion of unemployed workers who have been offered a programme out of all openly unemployed who are registered at that municipality in that month, gives an idea of the degree of utilisation of the programme capacity, e.g. in terms of a waiting list of job-seekers already been offered to participate and against whom a potential candidate is competing.

Before concluding the discussion on the plausibility of the basic matching assump-

tion in our context, it is worth considering two potential sources of its violation: the possibility of anticipatory effects in terms of future programme participation and in terms of future employment.

'Waiting' outcome as subsequent programme participation. The institutional nature of the programme system (a seemingly continuous flows of different programmes often on an individual, ad hoc basis) should make it less likely for an unemployed job-seeker to have to turn down a programme offer perceived as secondbest in order to wait for a free slot on his first-choice programme (this would also reduce the likelihood of an 'Ashenfelter dip' problem in terms of reduced job search prior to participation). Even if he did wait, though, he would not enter his first-best programme with probability one, but would still be exposed to the possibility of finding a job or deciding (or be forced) to join another programme in the meantime. An interesting piece of information in the Swedish dataset is an unemployment spell characterised by having been offered a labour market programme. Having gone through the selection process and having been offered a place makes it more likely for the individual to join a programme rather than waiting; had he not joined now, he would be more likely to join later on or to decrease his job search in anticipation of joining. In addition, the possibility of a supply constraint is directly controlled for by the municipality indicators, which should reflect local conditions and capture the possibility of a limited supply of programmes for which registered job-seekers may be competing. Simultaneously conditioning on individual programme offer, local programme rate and local offer rate prevailing at the employment office and time of an individual's joining-waiting decision should go a long way towards controlling for anticipatory effects in terms of subsequent programme participation.

'Waiting' outcome as subsequent employment. An issue that needs to be considered is the possibility for unemployed individuals to be re-hired by their former employer. If they *know* that their employer is going to call them back (e.g. they are seasonal workers, or have a credible agreement with their employer allowing the temporarily dismissed employee to collect unemployment benefits), they are likely to have

³¹ Basically most "kommun" have one employment office, though the largest cities might have several.

no (or less) incentives to participate in programmes at any given month in unemployment; at the same time, they are observed to find employment. In such a scenario, the CIA would be violated: programme participation would be (negatively) correlated with a subsequent outcome (job accession), which would bias our estimates of programme effects (downwardly). Other observables included to control for potential anticipatory effects of this kind include the occupation/skill type of the job-seeker, as well as the month of registration, which should help capture seasonal unemployment, as well as account for seasonal patterns³² in programme entry.³³

More generally, however, the CIA would be violated in the presence of hidden job offers, that is if an individual waiting longer has decided to do so *because* he *knows* that he will be hired shortly.³⁴ How serious this issue is going to be in our case thus largely depends on the typical time span between job offer and job commencement (and whether or not an individual who is going to start a job typically remains/is allowed to remain registered at the unemployment office in the meantime).

II.6 Empirical findings

II.6.1 Entitlement status as a determinant of programme participation

Although purely instrumental in obtaining a properly balanced matched comparison sub-sample, the estimation of the propensity scores – the various conditional probabilities of joining a programme in one's m^{th} month of unemployment – provides in-

³² Participation in the various types of programmes exhibits strong seasonal fluctuations, with significant drops in summer (August) and to a lesser extent in January.

³³ There is in fact explicit information in our data recording if an unemployed individual de-registers because he has obtained further employment by his former employer. Although qualitative analyses based on this information would indicate that recalls are not likely to be too serious a problem, the quality of this code has been found to be low: interviews with unemployed individuals who had found a job show that recalls are much more common than recorded in the data (Jansson, 1999).

³⁴ It has to be said, though, that our by-month version of CIA is less likely to be violated due to this kind of anticipatory effects than the traditional 'overall' CIA defining as non-treated those individuals whose unemployment spell is not interrupted by any programme, but directly ends in de-registration.

teresting information as to what factors significantly affect the treatment probability depending on how long the individuals have already been unemployed.

The changing impact of entitlement status over the unemployment duration prior to programme entry is undoubtedly the most interesting finding among the determinants of programme participation. Figure 2.1 shows the evolution of the marginal effect of UI status over unemployment duration prior to joining the programme. The plotted marginal effects represent the percentage points difference in the probability of entering a programme in that month of individuals entitled to UI vis-à-vis nonentitled individuals, who however have the same observed characteristics of UIindividuals. For almost the entire first year in unemployment, roughly up to the 10th month, receiving benefits discourages programme participation. The effect then becomes insignificant, while just at benefit exhaustion, i.e. after 14 months, the effect becomes significant and its sign reversed: being entitled to UI increases the likelihood of now joining a programme, with the impact reaching its peak just after benefit exhaustion – in the 15th month of unemployment. Being entitled to UI confers to individuals reaching this unemployment duration an 11 percentage points higher likelihood of joining a programme than observably identical non-entitled individuals. Carling, Edin, Harkman and Holmlund (1996) too had found that UI-entitled individuals close to benefit exhaustion are significantly more likely to join a programme than those without unemployment compensation.

An individual who in month *m* is offered to start a job in month k (k>m) represents a violation of the 'traditional' CIA, but of only (*m*-*k*) of our CIA's.

Figure 2.1 Marginal effect of UI-status on the probability of joining a programme (percentage points difference in the treatment probability with respect to non-entitled with the same characteristics of UI individuals), by time unemployed prior to programme



II.6.2 Outcomes over time

This section looks at various measures of outcomes over a five-year period to investigate how unemployed individuals who join a programme perform, on average, compared to a situation where they would have searched further in open unemployment. The assessment begins with the dynamic effect of joining a programme on the probability of programme participation, summarising both the (endogenous) duration of the programme joined as well as possible repeated participation in subsequent programmes.³⁵ Had they postponed their waiting decision, participants would have remained openly unemployed; an interesting outcome to assess is thus their unemployment rate over time. We finally focus on the treatment effect on the various routes out of the unemployment system: employment, regular education and inactivity.

Figure 2.2A depicts the probability of programme participation for treated and matched controls over time³⁶, while Figure 2.3A the corresponding treatment effect,

³⁵ It may be worth reminding that object of the evaluation is the effect of the treatment represented by *joining* a programme; anything happening after that, including subsequent time on this first programme as well as possible subsequent programme participation, has to be viewed as an outcome of that first joining decision.

³⁶ To fix ideas, the figure is to be interpreted as follows. By definition, at time 0 the probability of the treated to be on a programme is one, while for the controls it is zero. One month later, at t=1, a participant has, on average, an 89% chance of still being on the programme. Had he instead not gone on a

which starts at entry into the programme. The graph shows a relatively large and persistent treatment effect: for up to two and a half years from joining, participants are significantly more likely to be on a programme than if they had further postponed their initial participation decision.

A serious indication about the influence of programmes on subsequent labour market status is given by the unemployment probability, and in particular by the probability of being on unemployment benefits over time. While Figure 2.3B shows absolutely no treatment effect on the probability of being unemployed after the typical programme duration, Figure 2.3C indicates that *as soon as* the programme typically ends (i.e. after about 4 months), the negative effect (by construction, compensation while on programmes is not counted as unemployment benefits) abruptly turns into a large positive one. Over our horizon, participants remain significantly more likely to be drawing benefits up to four years after having joined the programme.



Figure 2.2 Treated and matched controls' differential probability of

programme when he did but had he waited longer, he would be expected, on average, to be on a programme one month later with a 12% probability.



Figure 2.3 Treatment effect (% points) on the probability of



Notes: Time in month, with *t*=0 at programme entry. 95 percent bias-corrected percentile bootstrapped confidence intervals (205 repetitions).

A particularly important outcome is the probability of being employed over time. Do individuals who have joined a programme in a given month spend more time in regular employment, overall, than if they had further postponed their participation decision? How does programme participation affect participants' short and long-term employment probabilities? The graphs in Figures 2.2B and 2.3D show that although joining a programme has a negative effect in the very short term, the programmes' impact on participants' more long-run employment probability is positive and significant. In particular, joining a programme is initially expected to reduce the chance of finding employment by up to 4 percentage points: compared to open unemployment, job search whilst on a programme is clearly reduced because less time is left due to participation itself (the 'lock-in effect'). Nevertheless, after a programme typically ends, participants do appear to perform significantly better than if they had waited longer in open unemployment, with their participation decision paying off in terms of an average of 5 percentage points higher long-term employment probability (for at least up to five years). Joining a programme 'now' thus seems to actually reduce the expected overall time out of regular employment, on average.

This conclusion is not however supported when assessing a different type of outcome: the probability of being out of the unemployment system. This outcome encompasses the previous one – the probability of being employed – but it also considers as a 'success' the other reasons for being de-registered: having gone on regular education, having left the labour force or having been de-registered because of 'contact ended'. What we know about people being de-registered is that they are somewhere 'out there', in the labour market, out of the labour force, on other education – in any case out of the official unemployment system and certainly not claiming benefits. When considering this type of outcome, programmes do not seem to be beneficial. Much to the contrary, the initial sizeable negative lock-in effect is gradually reduced in size, still the negative programme effect persists up to the end of the 3^{rd} year since joining the programme (Figure 2.3E).

In order to shed more light on these last two contradictory results, we next look at the programme effect on the various labour market states that make up the 'out of the unemployment system' one. If programmes enhance participants' human capital, they may find it easier to accumulate further human capital and may decide to deepen or specialise the acquired knowledge in the regular education system. Figure 2.3F however shows that beyond the initial negative lock-in impact, participants are no more likely to invest in further education than comparable individuals who have postponed their participation decision. By contrast, joining seems to have a significantly negative effect on inactivity rates³⁷, which persists up to 5 years after the joining decision (Figure 2.3G).

This is however a small treatment effect, so that the suspicion arises that the divergent impact on employment rates and on de-registration/employment rates may in fact be due to a negative programme impact on the last type of de-registration, the 'lost' status. In the following, 'lost' refers to an individual spell following deregistration, the reason of which has been recorded as 'contact ended'. This happens when a registered unemployed individual, having first missed an appointment at the official employment office, subsequently fails to contact the agency within a week. In fact, the negative programme effect on 'lost' rates is decidedly large (Figure 2.3H).

³⁷ Note that inactivity includes education in the regular system.

The problem of the 'lost' individuals is a serious ones; in fact, it prevents us from fully observing the outcome of interest, that is the true labour market status these individuals find themselves in. We do not know which of these spells is in reality an employment spell the former unemployed did not bother to report to the agency, and which is by contrast still part of the preceding unemployment spell. Bring and Carling (2000), who have tried to trace back a sample of 'lost' individuals, have found that around half of them had in fact found a job, which highlights how severely underreported employment status is in the available data. The large negative programme effect on 'lost' rates would thus turn out to be in part a large negative effect on employment rates.

In conclusion, the robustness of the above evidence of a positive programme impact on employment probability needs to be carefully checked against these lost spells. The next section presents the results of various sensitivity analyses, bounds and imputation techniques performed in this direction.

II.6.3 Accounting for a partially unobserved outcome variable

In the following, define Y_i to be a dummy variable equal to 1 if the individual is in employment (at a given time) and 0 otherwise. For simplicity of exposition, let us abstract from time and from the two groups (the calculations have obviously been performed separately for each group and for each point in time). Let *L* be a dummy variable indicating the 'lost' state, and *D* the usual treatment indicator.

A very simple sensitivity analysis without any additional external information looks at the estimated effects on employment rates under various assumptions about the percentage of 'lost' individuals who have in reality found a job. A 'misclassification' rate of 0% would thus mean that the observed employment rates (thus the effect on employment probability in Figure 2.3D) are the true ones, while at the other extreme a 100% misclassification rate would imply that it is the sum of the observed employment rates and lost rates that represents the true employment rate. Note that this analysis assumes that the probability of being misclassified is the same for lost treated and lost controls, that is that outcome data Y are missing completely at random:

$$P(Y=1 | L=1, D=1) = P(Y=1 | L=1, D=0).$$

Figure 2.4 confirms that the observed average employment effect (4.8%) would in fact decline with more lost individuals having in reality found a job. With the almost 50% misclassification rate found in the survey by Bring and Carling (2000), it would be more than halved. Still, to have the effect disappear or change sign, one would need to assume that 80% or more of the lost individuals had in reality found a job.

A second step makes use of external information provided by the follow-up survey by Bring and Carling (2000) to impute to each 'lost' individual spell the probability of it in reality being an employment spell. Unfortunately, the X's used by these authors do not include previous programme participation.³⁸

Figure 2.4 Average treatment effects on employment probability (averaged over the 5-year horizon since start of the programme) by misclassification rate



We thus need to assume that the misclassification probability is independent of treatment status, this time however given observables X, i.e. that the outcome data Y is missing at random:

P(*Y*=1 | *X*=*x*, *L*=1, *D*=1) = *P*(*Y*=1 | X=*x*, *L*=1, *D*=0)

³⁸ Regressors include age group, gender, foreign status, human capital indicators (work experience, education), city region, and a few age-human capital interaction terms. Implicitly, we are also conditioning on non-entitlement: being registered is a prerequisite for drawing benefits, and in fact none of the lost spells in our data is characterised by unexpired eligibility.

Using Bring and Carling (2000, Table 4) $\hat{\beta}$ coefficient estimates, the conditional probability of misclassification of a given lost individual with observed characteristics *X* is estimated by:

$$\hat{p}_i^Y \equiv \hat{P}(Y_i = 1 | L_i = 1, X = x_i) = (1 + e^{-\hat{\beta}' x_i})^{-1}$$

Two alternative strategies are then pursued.

We decide that a given lost individual has in reality found a job if his misclassification probability is larger than a given cutoff τ , that is if $\hat{p}_i^{\gamma} > \tau$ we consider that lost spell as an employment spell. The analysis of the treatment effect on employment probability is then performed as in Section II.6.2 for various cutoffs τ .³⁹ Figure 2.5 – strikingly similar to Figure 2.4 – summarises the corresponding average employment effects; a positive effect does in fact persist up to a cutoff as low as 30%.



Figure 2.5 Average employment effect by cut-off probability (averaged over the 5-year horizon since start of the programme)

An alternative possibility is to count a lost individual with an (estimated) misclassification probability \hat{p}_i^Y as a $(\hat{p}_i^Y)^{th}$ of an employed individual. In fact, simple calcu-

³⁹ Note that a cutoff of 0 corresponds to a 100% misclassification rate, while a cutoff of 1 to a 0% misclassification rate.

lations show that we can estimate the employment rate (separately for the treated and the control group and at a given time period) as⁴⁰:

$$\hat{P}(Y=1) = \frac{1}{N} \left(\sum_{i \in \{L=0\}} Y_i + \sum_{i \in \{L=1\}} \hat{p}_i^Y \right)$$

where N is the total number of individuals (in the group and time period under consideration). The resulting dynamic treatment effect on employment is plotted in Figure 2.6 below. Even though visibly reduced from the observed one, joining a programme seems to still have a long-lasting positive impact on employment rates over time, compared to similar individuals who have decided to wait longer.

In these last two types of analyses, we have used the imputed misclassification probability to estimate the employment probability of a lost individual irrespective of his treatment status – a regressor not included in the estimation by Bring and Carling (2000).⁴¹ This amounts to assuming that for a given set of the *X*'s, the distribution of the probability that a lost individual has in reality found a job is the same in the treated and non-treated groups. In our case, treated individuals are those observed to enter a programme, while all we know about non-treated individuals is that they not necessarily do so, making it not easy to argue if such an assumption is likely to be systematically violated, and if yes, in which direction. Still, since we are looking at outcome measures (probabilities or rates) which are bounded, we can apply the core idea of the literature on non-parametric bounds in the presence of missing data (see

⁴⁰ The object of interest is the employment probability (or employment rate) for a given group at a given time, P(Y=1), which can be written as: $P(Y=1) = \sum_{x} P(Y=1|X=x)P(X=x)$. P(X=x) can be estimated by $\#\{X=x\}/N$, where $\#\{A\}$ denotes the number of elements in set A and N is the total number of individuals in the group being considered. We thus focus on $P(Y=1 \mid X=x)$, which can be decomposed as: $P(Y=1 \mid X=x, L=0) P(L=0 \mid X=x) + P(Y=1 \mid X=x, L=1) P(L=1 \mid X=x)$. In our data we observe all terms except $P(Y=1 \mid X=x, L=1)$, for which we use \hat{p}_i^Y , the estimated probability that a 'lost' individual with characteristics X has in reality found a job. $P(L=l \mid X=x)$ is estimated by $\#\{X=x, L=l\}/\#\{X=x\}$ for l=0,1; and $P(Y=1 \mid X=x, L=0)$ by $\sum_{i \in \{X=x, L=0\}} Y_i / \#\{X=x, L=0\}$. Simplifying and integrating out the X's

finally yields the formula in the main text.

⁴¹ Other important information regarding the unemployment spell prior to attrition is similarly missing, in particular the duration of the unemployment spell or the number of times the individual has registered as unemployed before.

e.g. Manski, 1990) and exploit the additional information from the survey to derive worst- and best-case bounds for the treatment effect of employment rates. We start by writing the conditional misclassification probability of lost individuals with characteristics *X*, P(Y=1|X=x, L=1), as:

$$P(Y=1|X=x,L=1,D=1) P(D=1|X=x,L=1)+P(Y=1|X=x,L=1,D=0) [1-P(D=1|X=x,L=1)].$$

For each lost individual, we know his treatment status *D*; his treatment probability given the lost status, $P(D=1|X_i,L_i=1) \equiv e_i^{42}$; and his misclassification probability $P(Y_i=1|X_i,L_i=1) \equiv \hat{p}_i^Y$.

The procedure to derive worst- and best-case bounds (where worst or best are from the point of view of programme effectiveness) consists in letting a lost individual *i* of treatment status d_i count as a π_i -th of an employed individual, with $\pi_i \equiv P(Y_i = 1 | X_i, L_i = 1, D = d_i)$ obtained by setting $\overline{\pi}_i \equiv P(Y_i = 1 | X_i, L_i = 1, D = 1 - d_i)$ to its maximum or minimum, compatible with the given \hat{p}_i^Y and e_i , as well as with all probabilities $P(\bullet) \in [0,1]$. Table 2.1 displays the setting of $\overline{\pi}_i$ and the corresponding computation of π_i for the various cases, while the resulting bounds on the treatment effect on employment rates over time are shown in Figure 2.6.

⁴² Due to the absence of a 'standard' D=0 control group, the probability that a lost spell with characteristics X belongs to a treated as opposed to a 'non-treated' individual has been estimated separately by month of entry. In particular, for a given treated *i*, e_i is the estimated probability that a lost spell with characteristics X_i belongs to a treated individual as opposed to an individual who was still unemployed when treated *i* joined the programme. An individual *j* who is used (possibly repeatedly) as control for a treated entering in month m_1 starts being evaluated from m_1 and if he has lost spells, the corresponding employment probability bounds are calculated using the probability that a lost spell with his characteristics X_j belongs to an individual treated in month m_1 as opposed to an individual who was still unemployed after m_1 months. If this same individuals *j* also acts as control for another treated entering in month m_2 , he counts as another person whose outcome starts being evaluated from m_2 and whose lost spells are evaluated using bounds on the probability that a lost spell with his characteristics X_j belongs to an individual treated in month m_2 as opposed to an individual who was still unemployed after m_2 months.

Worst-Case Scenario			Best-Case Scenario				
as con	Treated assign the highest possible $\overline{\pi}_i$ compatible with \hat{p}_i^Y , e_i and $\pi_i \ge 0$			Treated assign the lowest possible $\overline{\pi}_i$ compatible with \hat{p}_i^Y , e_i and $\pi_i \le 1$			
If	Set $\overline{\pi}_i =$	Thus $\pi_i =$	If	Set $\overline{\pi}_i =$	Thus $\pi_i =$		
$\hat{p}_i^Y \leq 1 - e_i$	$\hat{p}_{i}^{Y}/(1-e_{i})$	0	$\hat{p}_i^Y \ge \mathbf{e_i}$	$(\hat{p}_{i}^{Y}-e_{i})/(1-e_{i})$	1		
$\hat{p}_{i}^{Y} > 1 - e_{i}$	1	$(\hat{p}_{i}^{Y}+e_{i}-1)/e_{i}$	$\hat{p}_i^{\scriptscriptstyle Y} < \mathbf{e}_i$	0	$\hat{p}_i^{\scriptscriptstyle Y}/{ m e_i}$		
as	Controls assign the lowest possible $\overline{\pi}_i$ compatible with \hat{p}_i^Y , e_i and $\pi_i \le 1$			Controls assign the highest possible $\overline{\pi}_i$ compatible with \hat{p}_i^Y , e_i and $\pi_i \ge 0$			
If $\hat{p}_i^Y \ge 1-\mathbf{e}_i$ $\hat{p}_i^Y < 1-\mathbf{e}_i$	Set $\overline{\pi}_i =$ ($\hat{p}_i^Y + e_i - 1$)/ e_i 0	Thus $\pi_i =$ 1 $\hat{p}_i^Y/(1-e_i)$	If $\hat{p}_i^Y \leq \mathbf{e}_i$ $\hat{p}_i^Y > \mathbf{e}_i$	Set $\overline{\pi}_i = \hat{p}_i^Y / \mathbf{e_i}$ 1	Thus $\pi_i =$ 0 ($\hat{p}_i^Y - e_i$)/(1- e_i)		

Table 2.1 Computation of π_i to derive worst- and best-case bounds

Figure 2.6 Treatment effect on employment probability, using observed employment rates, imputed employment rates, worst-case and best-case bounds (% points)



Notes: time in month, with *t*=0 at programme entry.

As expected, the treatment effect under the best-case scenario far surmounts the observed one, with the joining decision paying off in terms of a sizeable and increasing extra chance of being in employment over time. In fact, while the observed treatment effect soon stabilises at around 6%, the favourable bound keeps rising, reaching double a level (12%) five years after programme start. Quite interestingly, the upper bound on the employment effect is in fact always larger than the observed one in *absolute* size, entailing a larger lock-in effect during the first five months. Similarly, the figure confirms the expectation of a dynamic worst-case-bound treatment effect considerably lower than the observed one, with the former ranging between -3 and 0 percentage points after the lock-in phase. Overall, the impression from the graph is that one may need to invoke assumptions particularly unfavourable to the treatment in order to have the treatment effect disappear or reverse sign.

All the analyses in this section were meant to offer some qualitative evidence as to the robustness of the uncovered positive employment effect with respect to the problem of the lost individuals. Overall, these findings would seem to indicate that the effect of participating in a programme compared to postponing such a decision may remain positive under a variety of assumptions.

II.6.4 A summary so far

Despite the importance of the true nature of the 'lost' state to truthfully reflect the treatment impact on *actual* employment (as opposed to as observed by the employment offices), some conclusions can be drawn as to the treatment effect on unemployment. A large positive effect on compensated unemployment, together with the absence of an effect on (registered) unemployment, would entail a negative effect on uncompensated unemployment, which is likely to be further reinforced by the observed negative effect on 'lost' rates (since 'lost' individuals who are in reality unemployed cannot be drawing compensation). Focusing on registered unemployment, what programmes seem to affect is thus merely the *type* of the unemployment spell experienced, in particular the treatment 'swaps' uncompensated unemployment for compensated one. The treated also have a sizeably larger long-term probability of being on programmes than comparable individuals who have postponed their participation decision, so that the treated remain significantly more likely to be registered at an unemployment office over time.

To gain some insight as to the treatment effect on the overall labour market situation of participants, let us assume that, conditional on the *X*'s, the employment status of the lost individuals is missing at random for those who join and those who wait longer, so that we can impute their employment probability. Let us further abstract from the inactivity state (possibly a rather strong assumption for the lost individuals). These two assumptions allow us to completely 'attribute' the observed treatment effect on lost rates partly to the observed effect on employment and partly on unemployment rates. Joining a programme compared to waiting longer would then 'swap' uncompensated, unregistered unemployment not only for compensated unemployment as well as for further programme participation, but also for employment in the regular labour market. This is visualised in Figure 2.7, based on the 5-year average of the probability of being in various labour market states with and without the treatment. **Figure 2.7** Expected labour market status of individuals in the case of (A) waiting longer in open unemployment and (B) joining a programme, broken down into probability of being on a programme, in employment or in unemployment, the latter in turn broken down into compensated and uncompensated one.

(Average over the 5-year horizon since the start of the programme)



To roughly fix ideas, further focus on that treatment effect which after 4-5 months seems more stable and permanent in nature (thus ignoring the initial lock-in effect and the endogenous duration of the treatment). The figures resulting from this series of simplifying assumptions are presented in Table 2.2, where '*de facto*' (as opposed to 'observed') highlights that we are trying to account for the lost spells. Joining a programme would thus make participants on average 3 percentage points more likely to be in employment over time, compared to similar individuals who have postponed their participation decision. However, participants would also be 3 percentage points more likely to be in compensated unemployment and another 3 percentage points more likely to be on programmes over time. Overall, participants would thus have on average a 9% lower probability of being in an unemployment spell not supported by benefits and not registered either.

Table 2.2 A crude summary of the effect of participation *versus* waiting: The more permanent treatment effect on the probability of being in the various states, six months after joining a programme (% points)

<i>de facto</i> unemployed compensated ^a non-compensated ^c		on programmes ^a	<i>de facto</i> employed ^b	
+ 3	- 9	+ 3	+ 3	

Notes: Out of the labour force state is ignored. ^a The more long-term effect is informally gauged from Figures 2.3(A) and (C). ^b Employment probability is imputed for the lost individuals (i.e. assuming missing at random); the more long-term effect is then informally gauged from Figure 2.6. ^c As implied by the other calculations.

Treatment effects by month of placement

Further interesting insights are gleaned when looking at the various outcomes for different sub-groups of the treated based on the time they have spent in unemployment before being placed on a programme. Table 2.3 reveals that for those individuals joining a programme immediately (within their first month) or very late (in their 18th month) as well as around the time benefits expire (in their 14th month) the various treatment effects are considerably worse than those for individuals entering a programme in intermediate periods (3rd-6th months). While the differential effect for 'immediate' joiners may be explained by these individuals being possibly rushing the choice of the appropriate type of programme as well as locking themselves too soon, thus foregoing initial job offers, selection issues characterising individuals joining at different months are likely to play a considerable role in these findings (see footnote 28); in particular, for individuals entering a programme exactly at the time of benefit exhaustion, the main incentive may not be in leaving the unemployment system soon, an issue that will be the focus of Section II.7.

Rates/Probabilities	Placement in x th month:						
(% points)	1 to 18	1	3	6	14	18 *	
Employment:							
- observed	4.8	2.3	4.4	6.4	3.6	3.1	
- imputed	2.5	0.3	1.9	4.1	1.6	1.1	
- lower; upper bounds	-2.0; 8.0	-3.8; 6.2	-2.8; 8.3	0.1; 9.0	-0.6; 5.0	-0.8; 4.2	
Deregistered empl.	-2.7	-1.4	-3.2	-1.5	-5.6	-7.3	
Lost	-5.8	-4.6	-7.0	-5.5	-4.8	-6.1	
Inactivity	-1.7	0.9	-0.7	-2.4	-4.4	-4.3	
Education	0.0	0.7	0.2	-0.3	-0.3	0.5	
On Programmes	7.4	6.6	7.2	7.7	8.2	12.2	
Unemployment	-4.7	-5.2	-4.0	-6.1	-2.6	-4.9	
Benefit receipt	1.9	1.3	1.3	1.0	4.8	3.3	

Table 2.3 Average treatment effects by month of placement into the programme (averaged over the 5-year horizon since the start of the programme)

Notes: * averaged over 4.5 years.

II.6.5 Job accession

When focusing on employment, we have thus far been considering employment rates over time. The probability of being employed at a given point in time summarises potential effects on the initial job finding rate (including the lock-in effect) and on subsequent overall job stability, i.e. over the entire observation period and irrespective of any interruption. This section tries to shed some more light on the former effect, i.e. the treatment effect on the probability of finding a job and thus exiting that (first) unemployment spell.

Remaining unemployed for a long time can potentially be harmful in several ways; in addition to the sizeable social outlay for benefits needed to finance prolonged unemployment spells, the unemployed individuals' human capital may quickly deteriorate, so that the jobs they may remain suited for will become less and less skilled and qualified; they may gradually lose hope and motivation, reducing their search intensity and possibly suffering increasing psychological costs. Finally, there may be a stigma effect associated to unemployment – as well as, possibly, to lengthy programme participation –, so that it may be important to find a job rather quickly. Such

a stigma effect from both unemployment and programme participation has in fact been documented for Sweden by Agell and Lundborg (1999). Their 1998 survey of employers shows that 27% and 21% of the interviewed managers regard prolonged unemployment duration and programme participation, respectively, as a *strong* signal of low productivity.⁴³ Even participation in the most sought-after programme and the one most likely to raise human capital and thus productivity, labour market training, is viewed by 14% of employers as a strong indicator of low productivity.

A very relevant question in such a context is thus whether individuals who have joined a programme find a job sooner than if they had waited longer in open unemployment.

As in the previous analysis, unemployment duration for the treated is measured from their entry into the programme, for the controls from the moment their respective matched participants joined the programme. In the following, unemployment duration refers to an uninterrupted spell, possibly made up of various sub-spells where the individuals are either registered as unemployed or taking part in a programme.⁴⁴ Such a wider 'unemployment' spell thus ends either in employment, or in deregistration for other reasons⁴⁵, or else is censored on the last observation day.

The Kaplan-Meier survival functions for the two groups plotted in Figure 2.8 are significantly different, with participants expected to remain unemployed for 2 months or 15% longer than if they had postponed their participation decision. This finding is consistent with the existence of a considerable lock-in effect due to a reduced job search following entry into a programme, as well as being in line with the presumption, put forward in Section II.2, that it may well be the benchmark treatment in terms of job *search* assistance and enhancement offered by the employment offices to the openly unemployed that could be more effective in getting unemployed individuals back into work faster.

⁴³ On a 1 to 9 scale, strong means a score of a least 7.

⁴⁴ As before, subsequent programme participation – including the possibility of starting a cycle between unemployment and programmes – is regarded as an outcome of the treatment, defined as the *first* programme joining decision.

⁴⁵ In this paper, this includes 'lost'.





Notes: Test for equality of survivor functions: Pr>chi2=0.0000. Expected unemployment duration: treated – controls = 468 – 406 = 62 days.

Looking at heterogeneous treatment effects on subsequent unemployment duration by month of placement into the programme, reveals some interesting patterns. First of all, the longer an individual has already been unemployed, the longer his remaining expected unemployment duration is going to be, and this irrespective of treatment status. Treated individuals, though, seem to remain unemployed for longer than comparable non-treated. What's more, the magnitude of the impact varies greatly according to how long the participants have been unemployed before joining the programme. Roughly speaking, for individuals joining programmes earlier on the unfavourable treatment effect tends to be smaller in magnitude. For instance, those participants joining a programme in their first month subsequently remain unemployed for roughly one month longer than if they had waited longer, on average; by contrast, for those joining the programme in their 14th month in unemployment the extra unemployment duration is over 3 months.

Given the wider definition of 'unemployment' used in this analysis, these latter findings reinforce previous results concerning the possibility that joining a programme may prolong subsequent permanence in the unemployment system by providing incentives to switch between compensated unemployment and eligibilityrenewing programme participation, an issue explicitly explored in the next section.

II.7 Unemployment-programme cycling behaviour

The quite distinctive feature of the Swedish institutional system – whereby participation in labour market programmes qualifies for new periods of unemployment compensation – could end up strengthening the work disincentive associated with the generous unemployment insurance system. Previous Swedish evidence on the importance of issues relating to unemployment benefits, work disincentive effects and cycling behaviour would in fact seem to overall support such a conjecture.⁴⁶ In fact, significant positive treatment effects on both compensated unemployment probability and subsequent programme rates have been uncovered in Section II.6.2, while Section II.6.5 has documented how joining a programme, and especially around benefit exhaustion, is found to prolong subsequent permanence in the unemployment system.

To investigate cycling behaviour in our data explicitly, we propose a working definition of a *cycle* as follows. An individual is allowed to register (for his first time or anew) as unemployed (U^{new}), to interrupt this spell by joining a programme (\hat{P}) and to then resume it. However, if he then enters a new programme, this is considered his first spell in a cycle. A (general) cycle is then defined as a chain of alternating subsequent programme (P) and unemployment (U) spells starting after the recommencement (following the programme) of the registered unemployment spell, in symbols $U^{new} \hat{P}$ U-P(..UP..), where the spells in bold denote the cycle. A *compensated*

⁴⁶ Agell and Lundborg (1999) describe a micro-simulation based on the benefit and tax regulations (including housing allowances, social assistance and child care fees) in 1998, according to which 4% of the unemployed would actually gain nothing from finding employment, 38% would have a disposable income between 90 and 99% of their income if employed, and 36% would have a disposable income between 80 to 89% of their income in employment. Regnér (1997) provides some evidence that job-seekers may often have entered labour market training just to renew benefits; results by Carling, Edin, Harkman and Holmlund (1996) based on a competing risks model show that UI-entitled individuals close to benefit exhaustion are significantly more likely to exit their unemployment spell to a programme than those without unemployment compensation (cf. their Figure 3). They also uncover a UI work disincentive effect, though small in size. The small effect could partly be accounted for by their type of data, which do not record actual compensation received by the unemployed. In a subsequent study, Carling, Holmlund and Vejsiu (1999) do in fact find a significant and large negative UI effect on job finding rates. Ackum Agell, Björklund and Harkman (1995) find that prolonged spells of benefit-programme periods are indeed common in Sweden, while Hägglund (2000) detects a very interesting sensitivity of employment duration as well as time spent on a programme to changes in the UI work requirement.

cycle is defined as $U^{\text{new}} \hat{P} U^{c}$ -**P**(..U^c**P**..), i.e. as a cycle where in *each* unemployment spell, as well as in the one preceding the start of the cycle, the individual draws UI or KAS compensation (U^c).

A first idea of the quantitative importance of the phenomenon for the individuals in our sample can be obtained both in terms of the individuals involved – almost one in two (42%) of our 32,000 treated start cycling (from the treatment) and for more than half of them this is a compensated cycle –, as well as in terms of cycle length – over one third of the compensated cycles has four or more switches.⁴⁷

Over our horizon, results in Section II.6.2 have shown that participants remain significantly more likely to be drawing benefits up to four years after having joined the programme than if they had waited longer in open unemployment. In fact, as highlighted by Figure 2.9A, they are likely to be drawing benefits on an unemployment-programme cycling basis; overall, participants have a considerably high probability of being in the midst of a compensated cycle over time, which peaks around 15% and still remains at 5% after five years from the start of the programme.



Notes: time in month, with *t*=0 at programme entry.

Moving beyond these descriptive figures by looking at the causal effect of joining a programme (*versus* waiting longer) on the compensated cycle probability over time

⁴⁷ I.e. have an event history of the type $U^{new} \hat{P} U^c - PU^c PU^c$ or longer (with one quarter of them being censored).

reveals quite a sizeable dynamic effect, persisting well up to 50 months after entry into the programme (Figure 2.9B).

Cycling itself may be considered a worrying phenomenon for a number of reasons; the fact that treated individuals keep going on various programmes withouth exiting unemployment is clear evidence of a failure of the programme system itself, while the importance of compensated cycling behaviour points to a likely failure in the way incentives are taken into account by the intertwined unemployment benefitprogramme institutional system. In the following, the linkages between cycling behaviour and entitlement to unemployment compensation are investigated under different angles.

When looking at the determinants of programme participation by prior unemployment duration, it was all too clear that entitled individuals have (an over 10 percentage points) preferential access to programmes at and just after their benefit exhaustion (cf. Figure 2.1). Although this may be taken as an indication that joining a programme may often be done purely in order to escape benefit exhaustion (in fact, during the 1990s those at risk of benefit exhaustion were guaranteed a place in a programme), on its own this would not account for the absence of positive programme effects. It may be that programmes do manage to teach individuals some new skills and good working habits, even though participants may initially have joined them purely out of the need to renew their benefits.

This possibility does not however seem to be supported by further evidence presented in previous sections: individuals entering a programme in their 14th month of unemployment (the month coinciding with benefit exhaustion) have been found to be that sub-group of the treated for whom the various programme effects have consistently been among the worst. The disappointing results in terms of the probability of employment, de-registration, studies and especially benefit receipt over time (cf. Table 2.3), as well as in terms of survival in unemployment could partly be explained by cycling behaviour. In particular, by far the worst treatment effect in terms of compensated cycling probability is again displayed by those joining a programme in their 14th month of unemployment: for this '14-group', the average effect is an 8 percentage points higher probability of being in the midst of a compensated cycle over time, against 1.8% for individuals joining a programme in their first month of unemployment and 3.5% for the whole group of treated (Figure 2.10A).

Figure 2.10 Compensated cycling by month of placement

(A) Average treatment effect on (B) Percentage of indiv





(B) Percentage of individuals becoming compensated cyclers



When looking at sub-groups of participants based on their month of placement (Figure 2.10B), the 14-group again stands out as the one with the highest probability of becoming a cycler: 44% of the individuals joining a programme in their 14th month of unemployment are observed in our data to start a compensated cycle (60% a general cycle), compared to around 15% (32% for a general cycle) of those entering a programme early (within 3 months of open unemployment).

Figure 2.11 contrasts the behaviour of the 14-group to that of individuals who are treated within a year of registration. The graphs plot the two groups' probabilities of being on a programme over time, as well as the probability of being collecting unemployment benefits over time. Measurement starts at the beginning of the first programme (i.e. the treatment).

Figure 2.11 Programme and benefit collection probability over time: 14-group versus 1-12 group



Notes: time in month, with *t*=0 at programme entry.

As expected, programme rates steadily decline over time. However, while for the 1-12 group they keep declining up to 5 years after first receiving treatment, they visibly peak again after 20 months for the 14-group. This is very revealing: to renew benefits, participation needs to last at least 5 months, to this add another 14 months as the maximum period of compensated unemployment, after which we are witnessing these individuals going back into programmes. More than one in four (26%) of the individuals in the 14-group is again on a programme in their 21st month, compared to 15% of the 1-12 group. The story is however not over yet; just add another 5 months spent on the programme starting from their 19th month, another 14 months of compensated unemployment, and it should come to no surprise that the 14-group exhibits a further peak in participation probability at around the 38th month.

That the spells of unemployment are likely to be supported by benefits is shown by the right-hand side graph. After the initial rise due to the exit from the programme back to unemployment, the probability of benefit collection steadily declines for the 1-12 group. For the 14-group the pattern of this indicator is quite different: after the end of the programme (i.e. after around 5 months) it skyrockets reaching almost 50% (compared to barely 25% for the earlier placement group). After the programme, then, half of the 14-group individuals fall back into compensated unemployment. Their probability of benefit collection then begins to fall, though not only does it remain significantly higher than the one for the other group over the whole observation period, but it also rises again in correspondence of 25 months. If we allow 5 months on the treatment programme, another 14 in compensated unemployment, another 5 in a further programme to renew eligibility, we arrive at the start of their 25th month, from which time almost one third of the original 14-group participants is again claiming benefits. Once again, the story continues, with another hump in correspondence of 14 months of compensated unemployment, after which the benefit receipt probability drops, only to start rising again after 5 months.

Thus far we have seen that starting a cycle from the first treatment is far from uncommon, and that this applies in particular to those who enrol into programmes in their 14th month in unemployment, a period suspiciously coinciding with unemployment compensation exhaustion. The supposition that these individuals may just view the programme as a passport to renew benefit entitlement is supported by the finding that the main treatment effect is a stronger incentive to keep switching between compensated unemployment spells and subsequent programme participation. Additional pieces of evidence point in fact to entitlement (renewability) considerations as a weighty driving force behind the observed programme effects. In particular, UIentitled participants are found to be alternating between unemployment and programme participation for an average of 654 days, 20% longer than KAS-entitled and 35% longer than non entitled treated. Focusing on the subset of fully compensated cycles, the ranking of the durations on the basis of the strength of the incentives provided by UI and by KAS is again confirmed.⁴⁸ Furthermore, not only do UIcompensated unemployment spells in cycles last on average significantly longer than KAS-compensated ones, but there is an unmistakable clustering of durations in correspondence of the maximum duration of benefits for both types of entitlement: at 210 days for KAS and at 420 and 630 days for UI.

⁴⁸ KAS pays out on average around half of the cash benefits provided by UI, and for only half the period (see Section II.2).

II.8 Programmes for unemployed adults: The role of entitlement status

Various pieces of evidence have thus consistently pointed to entitlement (renewability) considerations as a weighty driving force behind the direction and strength of the observed programme effects. To shed more light on the linkages between entitlement to unemployment compensation and programme effects, this section presents some further analyses that single out those individuals whose incentives are most likely to be affected: job-seekers entitled to unemployment benefits. In particular, it directly contrasts the programme effects for that sub-group of adults who were entitled to unemployment compensation when first registering as unemployed to the programme effects for unemployed adults who were not entitled to benefits at the start of their unemployment spell.⁴⁹

Focus of the evaluation is on the overall effectiveness of the six main Swedish programmes available to adult individuals at the height of the economic recession in 1994: labour market training, workplace introduction (API), work experience placement (ALU), relief work, trainee replacement and employment subsidies.⁵⁰

Figure 2.12A (B) plot the average effect on employment probability for an entitled (respectively for a non-entitled) unemployed individual of joining one of the above programmes, rather than searching in open unemployment for at least a while longer. The two figures offer a striking contrast. Although joining a programme initially has a significantly negative lock-in effect for both sub-samples, programmes appear to reduce their entitled participants' job search intensity much more severely, with a substantial lock-in effect of almost 20 percentage points compared to 5 for the non-entitled sub-sample.

⁴⁹ When disaggregating by entitlement status, the data no longer allow estimation of separate propensity scores by month. The less data hungry procedure followed is the one used and described in Chapter 3, III.4.2.

⁵⁰ A description of these programmes is postponed to the next chapter, III.2. Also note that results for the non-entitled sub-group relate to five programmes, since ALU is reserved to entitled individuals only.

However even after the typical 4-5 months duration of a programme, entitled participants do not seem to enjoy higher employment rates than if they had postponed their participation decision further. In fact, the negative programme effect persists for up to three years after entry into the programme, after which former participants become just as likely to be in employment as if they had further searched in open unemployment. These results are in sharp contrast to those for the non-entitled subsample, for whom a significant and substantial positive effect of the programmes was already visible just after 6 months.

Figure 2.12C plots the programme effect on the probability of compensated unemployment for the entitled sub-sample. The comparison with the effect for the non-entitled sub-sample (Figure 2.12D) is again very revealing, with twice as much of a positive effect on benefit receipt probability for entitled individuals (peaking to 30 percentage points) compared to non-entitled individuals (up to 15 percentage points). Finally, results for the entitled sub-group display a pronounced second 'hump' starting around the 24th month (=5 months on the programme + 14 months of maximum benefit collection + 5 months on another programme) and lasting for another 14 months – a feature likely to be linked to cycling incentives, whereby programmes are simply viewed and used as a gateway to renewed benefit eligibility.

In conclusion, the results – both in terms of employment rates and of benefit collection probability over time – for the sub-sample of entitled adults just considered are considerably worse than those obtained for the sub-group of non-entitled adults. Contrasting these two sets of results would thus lend support to the conjecture that for individuals entitled to unemployment compensation, the eligibility renewability rules are likely to significantly distort the incentives for participation and thus wipe out potential productivity-enhancing effects of several types of programmes.

Figure 2.12 Average effect for participants of joining any of the programmes^a compared to waiting longer in open unemployment on



Notes: ^aTraining, API, ALU, relief, replacement and subsidies for entitled; all but ALU for nonentitled.

Time in months, with *t*=0 at programme start. 95% confidence intervals bands

II.9 Discussion and conclusions

When assessing the effectiveness of the Swedish programme system, it is felt that the most important issue concerns the co-ordination and interaction between labour market programmes and the unemployment insurance system. A labour market programme in Sweden effectively comes as a bundle of two conflicting components: it is
intended to equip job-seekers with marketable skills which should improve their opportunities on the labour market, but at the same time it allows to renew eligibility to generous unemployment compensation (and until 1996 even to become eligible for the first time), thus reinforcing the work disincentive associated with the unemployment insurance system.⁵¹ In order to display a positive effect, any productivityenhancing component of the programmes would thus need to be strong enough to outweigh the reinforced work disincentive associated with the entitlement renewability that participation allows.

The results from the paper relate to how unemployed individuals joining a programme perform compared to a hypothetical state where they would have waited longer in open unemployment. Overall, the impact of the programmes appears to have been mixed, with evidence for both of the programmes' components being at work. Unemployed individuals who go sooner on a programme (compared to later or never) have a higher probability of being in employment from six months after joining the programme, for up to at least five years, an effect which seems quite robust to the misclassification problem of the 'lost' individuals. At the same time, there is visible evidence of the work disincentive element embedded in the institutional set-up of the programmes: joining a programme prolongs the current unemployment spell by a couple of months and greatly increases the probability of being in benefitcompensated unemployment over time, of participating in further programmes over time, as well as of being in the midst of a chain of alternating programme participation and compensated unemployment spells. When looking at the detailed mechanism, the positive effect on employment arises because the programmes considerably reduce the probability of being unemployed *outside* the official unemployment system, and to a lesser extent of exiting the labour force. Overall, joining a programme compared to waiting longer would thus 'swap' uncompensated, unregistered unemployment not only for unemployment-benefit compensated (and thus registered) un-

⁵¹ Several mechanisms may contribute to weaken the incentives to escape unemployment, such as a higher reservation wage, a lower search intensity or a lower geographical mobility.

employment as well as for further programme participation, but also for employment in the regular labour market.

Nevertheless, further results seem to indicate that the human capital-enhancing component of the programmes may not in general be strong enough to always outcompete the work disincentives provided by the system. In particular, for individuals entering a programme around benefit exhaustion – a time when entitled job-seekers are found to unmistakably enjoy preferential access to the programmes – the various programme effects are found to be among the worst than for any other group of participants. In addition to the disappointing results in terms of the probability of employment, de-registration from the unemployment office, regular studies and especially benefit receipt over time, by far the worst programme effect in terms of compensated cycling probability is displayed by those joining a programme at benefit exhaustion. Even more revealing, the evidence for individuals entitled to unemployment benefits provided a sharp contrast to the findings relating to non-entitled individuals, quite unmistakably pointing to distorted incentives behind programme participation as a most likely force behind the disappointing programme effects.

The present analysis has looked at the programme-benefit system in its entirety, lumping all kinds of programmes into one 'treatment'. Overall, the system may not be seen as being satisfactorily fulfilling its aim. Different treatments may however have differential effects, so that some programmes may indeed endow individuals with marketable transferable skills, whose return on the labour market may turn out to be large enough to outweigh the work disincentives created by the system. In the next chapter we apply the multiple-treatment matching framework recently developed by Imbens (2000) and Lechner (2001) to explore such a possibility in the Swedish set-up.

II.10 Appendix

2A.1 Matching protocol

- I) The following procedure is separately repeated for each month up to one and a half year since registration (i.e. m=1,2, ... 18):
 - 1. The relevant units are selected: all those treated who enter a programme within their m^{th} month in unemployment and all those non-treated who are still unemployed after m months since first registering at the employment office.
 - 2. The propensity score $e(X; m) \equiv Pr\{D^m=1 \mid X, T^0 \ge m\}$ is estimated on the two sub-groups, giving an estimate of the conditional probability of joining a programme in one's m^{th} month of unemployment.⁵²
 - 3. This sub-group of participants is matched to the corresponding sub-group of (at least yet) non-participants by nearest-neighbour with replacement within caliper δ ; treated unit *i* is matched to that non-treated unit *j* such that:

$$\delta > e(X_i;m) - e(X_j;m) = \min_{k \in \{D^m = 0\}} \{ |e(X_i;m) - e(X_k;m)| \}.$$

- 4. The differential performance of the sub-groups of the treated and of the matched comparisons is used to calculate the effect of entering a programme after *m* months *versus* waiting longer.
- II) All the M=18 the results by month of entry are finally aggregated to obtain the average expected effect of joining in a given month of unemployment compared to waiting longer, where the average is taken with respect to the observed joining distribution (i.e. the T^{1} distribution) of the treated.
- III) The bias-corrected percentile confidence intervals are obtained via bootstrapping.

⁵² Given that any order-preserving transformation of the propensity score is sufficient to matching purposes, the predicted linear index rather than the predicted probability has been used, as the former allows one to be more discriminating on individuals with predicted probabilities in the tails of the distribution.

Month	Treated	Non-Treated	Month	Treated	Non-Treated
1	4,141	99,992	10	982	21,417
2	4,004	85,377	11	861	19,047
3	4,739	68,669	12	747	17,115
4	3,665	53,478	13	823	15,093
5	2,602	43,976	14	968	13,247
6	1,922	37,283	15	852	11,474
7	1,689	32,101	16	672	10,064
8	1,367	27,494	17	501	8,938
9	1,140	24,159	18	384	7,998

2A.2 Treated and pool of potential controls by time in unemployment

2A.3 Matching quality in terms of the propensity score

Support of linear prediction:Min = -3.54066Max = 0.99597Difference in linear prediction between treated and matched controls:Mean = 0.00037Median = 0.0000295th percentile = 0.00056

2A.4 Matching quality in terms of imbalance of the most important covariates between treated and matched controls: standardised differences after matching (% bias)

Age at entry	-2.86	Sector	
Gender female)	1.48	professional and technical work	0.24
Foreign	0.38	health, nursing and social work	-0.03
Education		administrative, managerial, clerical	-0.92
compulsory	0.78	sales	-0.13
upper secondary	-0.30	production	-1.55
vocational upper secondary	0.06	services	-0.08
university	-0.60	Part-time	-0.71
Education for job sought (yes)	-0.79	Interlocal	-0.17
Experience for job sought		Month of registration	-2.82
some	-0.65	January	0.54
good	-0.91	June	0.74
Missing information	0.63	August	-0.36
Entitlement status		First register as part-time unemployed	1.16
UI	-2.09	Part-time unemployment	1.00
KAS	-2.30	Type of unemployed	
County		able to take a job immediately	-1.16
Stockholm	0.16	offered a labour market programme	2.23
Göteborg and Bohus	0.06	need guidance	-0.21
Malmöhus	-0.55		
Average abs. standardised difference	0.86	Propensity score	0.95
Median abs. standardised difference	0.71	Number of treated individuals	31,975

Notes: Following Rosenbaum and Rubin (1985), for a given covariate X, the standardised difference after matching is the difference of the sample means in the treated and matched non-treated sub-samples as a percentage of the square root of the average of the sample variances in the treated and non-treated groups. Average and median taken over all variables in estimation.

CHAPTER III

Differential effects of Swedish active labour market programmes for unemployed adults during the '90s

III.1 Introduction

A feature that makes the 'Swedish model' of active labour market programmes particularly interesting to examine is that unemployed individuals can potentially choose among a wide array of different programmes. While all aimed at improving jobseekers' labour market opportunities, some types of programmes provide direct incentives to move back into employment (e.g. by facilitating individuals' job search, providing wage subsides or fostering the acquisition of work contacts and references), whilst other measures try to make the working option more attractive by providing incentives to improve individual productivity and skills (e.g. *via* formal teaching or work experience).

Since different programmes may in fact have differential effects, a natural question concerns the relative effectiveness of the various programmes, ideally with the aim of singling out the best performing ones. Such an exercise could prove instructive for other countries as well, in particular those who have recently been focusing on active labour market policy¹; although with the obvious care required by cross-country comparisons, general lessons as to which type of programme is more effective could be shared (cf. e.g. Martin and Grubb, 2001).²

¹ Examples include the UK, where the 'New Deal for the Young Unemployed', introduced in April 1998 and sharing some of the features of the Swedish set-up, offers five types of programme; France, where a series of measures targeted at unemployed youth were introduced during the late '80s; or Switzerland, where an ambitious array of programmes was set up during the '90s. In fact, both at the OECD (OECD, 1996) and European Union (European Commission, 1998) levels, labour market programmes are increasingly viewed as important measures to reduce long-term unemployment.

² Microeconometric studies looking at the relative effects of Swedish programmes include among others Carling and Gustafson (1999) for self-employment subsidies *versus* subsidised jobs, Melkersson (1999a, b) and Frölich, Heshmati and Lechner (2000) for programmes targeted at the disabled, Larsson

The present evaluation focuses on the six main Swedish programmes that were available to adult individuals at the height of the economic recession in 1994: labour market training, workplace introduction, work experience placement, relief work, trainee replacement and employment subsidies.

The differential performance of these programmes is investigated both relative to one another and vis-à-vis more intense job search in open unemployment. More precisely, the evaluation concerns the effect, for participants in a given programme, of joining that programme compared to joining another available programme, as well as compared to waiting longer in open unemployment.

When looking at the relative effectiveness of one programme compared to another, one needs to consider a group of unemployed job-seekers who, at least formally, could have chosen any of the measures under consideration. Focus of this analysis are individuals entitled to unemployment benefits: they have exclusive access to some types of programmes and enjoy 'special' conditions on programmes of wider access (e.g. they are in principle granted the right to some types of programme when approaching benefit exhaustion). Special policy interest in examining this group arises from the fact that one of the programmes was created just for entitled individuals, so that a natural question concerns the actual effectiveness of this special measure. Furthermore, since up to February 2001 participation in a Swedish programme used to renew job-seekers' eligibility to unemployment compensation, we are focusing on that one group whose participation incentives have been most affected and for whom the trade-off between productivity-enhancing components of the programmes and the reinforced work disincentive associated with the benefit system should have been at its sharpest. In particular, Chapter 2 has shown how both in terms of employment rates and of benefit collection probability over time, programme effects for the sub-

⁽²⁰⁰⁰⁾ for youth programmes, Johansson and Martinson (2000) for two types of labour market training programmes and Carling and Richardson (2001) for the relative efficiency of eight of the Swedish programmes. Evaluations of differential programme impacts outside the Swedish context include the recent work by Gerfin and Lechner (2000) for Switzerland and by Brodaty, Crépon and Fougère (2000) as well as Bonnal, Fougère and Sérandon (1997) for France, and the earlier work by Ridder (1986) for the Netherlands.

sample of entitled adults were considerably worse than those obtained for the subgroup of non-entitled adults.

Still, even when focusing on entitled individuals, the programmes considered may well have heterogeneous effects: while some of them may simply lock participants in rather useless or low-qualified tasks that will give them no subsequent edge on the labour market, some others may endow participants – and even participants entitled to unemployment benefits – with skills marketable enough to make the working option sufficiently attractive. This chapter thus moves a step further and disaggregates the composite 'programme' analysed in Chapter 2 into its six distinct components.

We consequently concentrate on two important types of outcomes. Given that an explicit aim of the active labour market policy is to improve the employability of unemployed workers, employment rates over time will be considered, summarising possible programme effects on both job finding probability and survival in employment once an occupation has been found. This will allow to address the issue of what type of programme – if any – is most beneficial to participants in terms of their employment prospects in the short and in the long run. To capture the influence that benefit renewability considerations have been shown to exercise on the impacts of the programmes, special attention is also devoted to the differential programme effects on individuals' benefit collection probability over time.³

The next section outlines the relevant features of the six Swedish programmes being evaluated. Section III.3 presents the data and sample choice and offers a 'naïve' first evaluation of the programmes based on the raw data. Section III.4 highlights the evaluation problem in a multiple-treatment framework and how it has been addressed in the Swedish context, as well as discussing the plausibility of the underlying identi-

³ We do not consider differential programme effects on wages or earnings. Although this would provide interesting information on potential programme effects on individual productivity, as highlighted by Carling and Richardson (2001) increased income has never been an explicit objective of the Swedish labour market policy; programmes have by contrast traditionally represented a measure to keep a compressed wage structure.

fying assumption. Section III.5 presents the findings, before the concluding Section III.6.

III.2 The Swedish labour market policy

As described in Chapter 2, unemployment compensation takes on two forms in Sweden, the primary one being the relatively generous (up to 80% of the previous wage) unemployment insurance (UI), while the roughly half as generous cash assistance (KAS) was mainly designed for new entrants into the labour market. The duration of compensated unemployment is in principle capped at 60 weeks for UI- and 30 for KAS-recipients. However, up to February 2001, participation in a labour market programme for five months (or completion of a training course) would allow participants to renew their benefit eligibility. Given such a close link between the passive unemployment compensation system and the active programme system, object of the evaluation of this chapter are the six main programmes which in the middle of the 90s were open to adult unemployed workers entitled to unemployment benefits: labour market training, workplace introduction, work experience placement, relief work, trainee replacement and employment subsidies.⁴

To gain access to any programme, one needs to be registered at a local official employment office. The six programmes under consideration are additionally open to adults only (over 20 or 25), while work experience requires the individual to be entitled to unemployment compensation, and employment subsidies are targeted at the long-term unemployed. The latter may often be regarded as a mere guideline, though, since 20 percent of the employment subsidy participants in our data have spent less

⁴ Two programmes are excluded from the analysis on the basis that they are targeted to (or attract) quite specific sub-groups of unemployed individuals: self-employment grants (for individuals wishing to establish their own new business, with both a business idea and a financial plan approved by the offices) and vocational rehabilitation (for persons with occupational disabilities needing specialised resources for in-depth counselling and job-preparation measures). Findings by Carling and Richardson (2001) do in fact support the view that participants in self-employment grants may have better employment prospects due to unobserved characteristics than participants in the other programmes.

than the required six months in open unemployment prior to joining.⁵ All the individuals in our chosen sample satisfy the eligibility rules in terms of registration, age and entitlement criteria, while we shall control very carefully for unemployment duration prior to programme start.

Whilst on a programme, participants either receive the stipulated wage and other benefits on their 'temporary' workplace, or the equivalent of the unemployment benefit they would have enjoyed as openly unemployed. Most programmes have a maximum duration of six months (under special circumstances renewable for another six), though participants stay an average of four to five months.

Table 3.1 contrasts the main features of the programmes being evaluated.

Labour market training (AMU), by far the most expensive measure, is intended to augment participants' human capital with formal, full-time vocational⁶ teaching of new skills.

A second type of programme offers workplace traineeship to maintain and enhance contact with working life and gain practical experience, good working habits and references from which to later benefit on the regular labour market. *Work experience placement* (ALU) was introduced at the deepening of the recession in 1993 with the explicit aim to prevent entitled individuals from exhausting their benefits. In fact, individuals need to be eligible to either UB or KAS to participate in this scheme, which can involve almost any kind of activity (the most frequent tasks being in administration and construction). *Workplace introduction* (API), which replaced a number of older job-experience programmes, offers unemployed individuals a period of

⁵ Larsson (2000) finds the waiting period rule to be *de facto* regarded as a formal requirement for youth practice too.

⁶ To reduce the heterogeneity in courses offered, the focus of this evaluation is on vocational training. Like Carling and Richardson (2001), we exclude participants in non-vocational courses, which are aimed at helping workers with basic educational insufficiencies to move on to further education or to other programmes, rather than directly into a job.

PROGRAMME	AIM	ELIGIBLE	EMPLOYER	TRAINING	TASK	COMPENSATION ^a	EMPLOYER INCENTIVES	COST ^b
EMPLOYMENT SERVICES	fill job openings quickly, job search assistance and training			job seeker activi- ties		UI/KAS if enti- tled		
LABOUR MARKET TRAINING (AMU)	equip individuals with skills to find jobs more easily	>20	priv. and publ. providers	vocational class- room training		TA/BA course free		13,940
WORK PRACTICE								
Work experience placement (ALU)	prevent exhaustion of benefits while maintaining contact with the regular labour market and enhancing good working habits	entitled ≥20	90% public and non-profit		otherwise not performed	TA/BA	free labour	9,294
Workplace introduction (API)	contact with working life to get work- place training, job-experience and refer- ences	≥20	private and public	practical voca- tional training	otherwise not performed	TA/BA	pay tuition to government (2,000 SEK/month)	6,993
TEMPORARY JOB								
Relief work	specially created temporary jobs to maintain working skills and habits, also to avoid benefit exhaustion	>25	2/3 in public sec- tor (municipalities and state organiza- tions)		otherwise not performed	according to collective agree- ment	grant 50% of labour cost up to fixed amount (SEK 7,000/month)	9,201
Trainee replacement	enhance skills of employee while pro- viding an unemployed individual with work experience in a regular job	≥20	80% in public sector	on-the-job practice	e replaces regu- lar employee	according to collective agree- ment	grant 50% of labour cost up to fixed amount (SEK 7,000/month); deduction of train- ing costs; educational grant of up to 20,000 SEK per employee	7,665
EMPLOYMENT SUBSIDIES	establish permanent employment rela- tion	≥20 ≥6months unem- ployed	private sector only; from 97 some industries excluded	on-the-job practice	e normal	according to collective agree- ment	grant 50% of labour cost up to fixed amount (SEK 7,000/month)	5,968

Table 3.1 Synoptic table of the main features of the programmed	es
---	----

Notes: Information has been gleaned from various sources, in particular, Swedish Institute (1997). ^aTA is training allowance equivalent to the UI or KAS the individual would have been entitled to; BA is the basic amount (SEK 103 per day) if the individual is not entitled. ^b Total monthly cost per participant (SEK); such information is from AMS (1998) and has been taken from Carling and Richardson (2001, Table 1).

workplace training.

A third kind of measure provides unemployed workers with a temporary job. *Relief work* involves specially created temporary jobs, mostly in the public sector. The oldest measure (dating back to 1933) to create employment, it has diminished in importance during the 1990s, remaining primarily used for individuals at risk of losing their unemployment benefits (Swedish Institute, 1997); in particular, unemployed UI fund members who run out of compensation are in principle granted the right to a relief job. In a *trainee replacement scheme*, an unemployed individual replaces a regularly employed worker who is on leave for education. This measure thus allows an unemployed worker to acquire valuable work experience, while creating an opportunity for firms to update the skills of their employees.

Finally, *employment subsidises* not only represent a temporarily subsidised job opportunity to acquire job-specific human capital, but they are aimed at influencing an employer's hiring process: the engagement is implicitly expected to continue after completion of the programme.

Thus while all the programmes aim at improving participants' employment prospects, two important dimensions that distinguish the various types is the kind of skills provided and the way they are provided. At the one end of the spectrum, labour market training provides vocational classroom training of new skills deemed in demand. API has a strong emphasis on practical vocational training; similarly, ALU and relief work may provide participants with job experience and improve their working habits. Participants in these three kinds of programme are however prevented – at least formally – from performing tasks that a regularly employed individual would otherwise do. Although it is likely for such a rule to be often interpreted more as a recommendation than as a strict guideline⁷, to the extent it is adhered to, the type of on-the-job practice acquired may not be expected to be particularly marketable.

Like the two work practice schemes and relief work, trainee replacement and employment subsidies offer the opportunity to invest in job-specific human capital; in these cases, though, the participant does in fact replace ordinary labour. Finally, while trainee replacement – a deputyship for the employee on study leave – is intrinsically a temporary opportunity to gain job-specific experience, employment subsidies, with the implicit agreement that the employer will then hire the individual on a regular and indefinite basis, almost entail the 'promise' of a permanent job.

A final consideration relates to the first row in Table 3.1, which highlights that in Sweden the state to which programme participants can be compared to is in fact not one of being completely left on one's own to look for a job, but rather the baseline 'package' offered by the employment offices. Simply being registered as openly unemployed gives access to the various employment services offered by the offices, not only in terms of the increasingly computerised job information and matching of vacancies to applicants, but also in terms of the 'job-seeker activities', which include search-skill-enhancing activities such as training courses on how to apply for a job and motivation-raising activities. In some countries this kind of assistance is in fact considered a programme in its own right.⁸

III.3 Data, sample selection, and a preliminary look at the raw data

The base dataset is the one constructed for the analyses of Chapter 2, obtained by combining the unemployment register recording each individual's labour market status information over time (Händel), with the additional information for individuals entitled to UI or KAS provided by the unemployment insurance funds (Akstat).

As to sample choice, we need individuals who are homogeneous in those basic characteristics which determine eligibility to the programmes under examination; only then will it be relevant to examine their outcomes had they chosen a competing type of programme. As motivated in the previous sections, the choice of this paper is to focus on adult individuals entitled to unemployment benefits. An additional advantage compared to non-entitled indi-

⁷ Circumstantial evidence in Hallström (1994; reported in Ackum Agell, 1995) shows that all parties involved (sponsors, participants and the employment officers) believe that these projects often do replace jobs that are part of the organisers' normal activity.

⁸ An example is the Gateway period of the new UK New Deal programme for the unemployed.

viduals is in terms of data quality and availability: since registration at an employment office is a pre-requisite for drawing benefits, our chosen sub-sample is a particularly representative one of the sub-population of interest. The information for benefit recipients is thus especially reliable, but also much richer, since it includes all the information from the Akstat dataset (in particular, amount and type of compensation received, previous wages and working hours).

The sample of over 30,800 adult individuals has thus been selected who entered the employment offices for their first time and in the same calendar year 1994 (when unemployment was still at its highest), registered as openly unemployed⁹ and were entitled to either UI or to KAS. Additionally, individuals whose first programme was start-up grants, vocational rehabilitation or non-vocational training are dropped from the analysis (see *** footnotes 4 and 6). Our individuals, all in the 25-54 age group and with no occupational disabilities, are then followed until the end of November 1999.

An exploratory first look at the raw data allows one to gather a general idea of the paths participants in the various programmes follow after their respective programme. A few interesting features emerge from Table 3.2, showing the share of each type of participant moving on to a different labour market state directly after the programme. The exceptional performance of employment subsidies jumps to the eye: three quarters of participants *directly* after programme completion exit the unemployment register, and practically all for a regular job. The ranking of the various other programmes, lagging far behind and with replacement schemes as second best, is in line with *a priori* expectations about the degree of relevance of the experience gained on the programmes. If we accept that after such schemes participants would often need to spend some time job-seeking, the superiority of replacement schemes remains, although training now also performs quite well. Quite interestingly, a large fraction of former participants (around a third of the remaining unemployed pool) return to the *same* kind of programme.

⁹ In particular, given that the main purpose of the programmes is to enhance the re-employability of the unemployed, those registering as employed or directly entering as programme participants (possibly anticipating a risk of unemployment) are excluded from the sample.

	Type of programme							
	Train-	ALU	API	Relief	Replace-	Sub-		
	ing				ment	sidies		
Number of participants	1,387	2,983	425	654	483	426		
(as percent of participants)	(21.8)	(46.9)	(6.7)	(10.3)	(7.6)	(6.7)		
Directly after the programme								
(a) found employment	8.9	11.3	14.3	15.3	22.5	72.8		
(b) other exit ^a	2.0	2.6	3.8	6.6	5.6	2.6		
Out of those who after resuming their unemployment spell								
(a) found employment	29.5	23.8	25.0	35.4	47.1	42.0		
(b) other exit ^a	10.9	14.3	18.2	12.2	14.6	18.0		
(c) same type of programme	27.5	35.4	34.8	16.7	21.4	8.0		

Table 3.2 Transitions from the first programme onwards (% of respective participants)

Notes: ^a Other exit = exit from the labour force (including for regular education) or de-registered for 'contact lost'.

After this crude 'tracking' of participants' early moves, the raw data can be further explored by looking at employment rates over time for participants in the six programmes, starting from the moment they join and following them up to five years. The raw differential outcomes visualised in Figure 3.1 again clearly confirm the 'star' performance of employment subsidies. Still in line with *a priori* expectations, the second-best performer appears to be trainee replacement. It is interesting to note that labour market training, though not one of the best performing measures in the short-term, seems to catch up later on: employment rates of former trainees equal if not surpass those of former participants in replacement schemes. API and ALU seem to perform roughly equally well, with API offering slightly better outcomes in the short term.

Figure 3.1 Raw data: employment probability over time, by type of programme (Time in months; *t*=0 at programme start)



This simple picture emerging from the raw data, though interesting and in line with expectations, cannot however be taken as showing the causal effects of the programmes. Such differential performance may be wholly or partially attributable to a selection effect: individuals going into the different programmes are likely to systematically differ in terms of characteristics that also influence their labour market performance. Visual inspection of selected average characteristics in Table 3.3 clearly shows that participants in the different programmes are not a random sample from the population, but are in fact quite distinctive groups. There seem to be several variables – such as skills, qualifications and employment histories – that influence programme assignment and which are most likely to affect subsequent outcomes.

	Programme participants					Exits from unemployment				
	Train-	ALU	API	Relief	Replace-	Subsidy	Em-	Exit labour	Regular	Attri-
	ing				ment		ployed	force	educat.	tion
Age at entry (years)	38	41	37	40	38	39	38	35	33	37
Gender (% female)	43.5	43.7	50.4	26.8	79.3	33.8	47.0	77.4	66.2	49.4
Foreign (%)	6.1	7.4	21.2	9.2	3.7	4.7	4.6	7.1	5.9	10.2
Education (%): compulsory	29.6	33.5	35.3	33.5	18.0	32.6	26.0	24.9	22.2	30.4
vocational upper secondary	53.2	41.1	35.3	50.3	48.9	47.4	45.3	45.0	41.2	39.5
University	8.5	17.6	22.1	9.0	26.7	12.9	22.3	19.7	19.8	20.9
Educat. for job sought (% yes)	64.5	64.5	59.1	66.8	77.4	67.6	73.4	67.8	54.6	64.4
Experience (%): some	9.8	12.0	15.5	11.8	17.0	12.7	11.5	15.2	19.6	14.2
good	83.6	79.9	64.2	82.7	71.6	83.1	82.7	76.2	59.3	77.9
KAS (%)	5.8	6.3	13.6	14.8	5.2	12.7	8.2	6.1	6.1	15.8
Previous wage (SEK, daily)	641	667	602	665	555	647	665	591	587	617
Prev. working hours (% 40)	84.0	83.1	79.8	86.2	67.7	87.6	81.1	75.2	72.6	76.7
Sector (%)										
admin., manag. and clerical	19.8	16.5	18.1	6.4	8.7	16.4	13.4	17.6	16.2	12.0
sales	12.0	13.5	15.1	9.0	5.2	23.0	10.5	13.6	10.6	12.6
production	31.5	25.2	18.6	48.9	6.0	22.5	26.2	11.0	11.0	18.9
services	10.2	10.1	14.6	9.9	9.1	8.9	9.6	13.6	9.6	15.0
Looks for part-time job (%)	3.9	6.5	5.6	4.0	9.7	4.2	7.2	11.9	6.7	7.8
Part-time unemployment (%)	11.5	12.3	22.1	7.3	35.4	15.5	33.2	36.6	21.4	38.3
Needs guidance (%)	16.8	11.4	19.5	10.6	4.6	7.7	3.3	6.6	6.4	5.9
Unempl. duration (days)	232	349	507	277	217	319	249	329	208	413
Observed days on programme	116	148	141	137	125	146				
Number	1,387	2,983	425	654	483	426	15,972	2,680	2,456	2,739
Percent of total (= 30,863)	4.5	9.7	1.4	2.1	1.6	1.4	51.8	8.7	8.0	8.9

Table 3.3 Selected individual descriptive statistics, by type of exit from first unemployment spell

III.4 Methodology

III.4.1 The evaluation problem in a multiple-treatment framework

In the prototypical evaluation problem, the effect on some outcome of a single 'treatment' of interest is assessed relative to another comparison treatment (the latter generally corresponding to the non-administration of the treatment of interest).

When it comes to the evaluation of a country's active labour market policy, however, the 'treatment' is no longer homogeneous, but is made up of various kinds of programmes which may well differ in terms of their effects on the outcome of interest. In such a context, a natural question arises as to the relative effectiveness of the different types of measures.

This sub-section sketches the framework recently developed by Imbens (2000) and Lechner (2001), which generalises Rosenbaum and Rubin's (1983) potential outcome approach for the case of a single treatment to the case where a whole range of treatments is available.

More precisely, let a set of K+1 different kinds of mutually exclusive treatments¹⁰ be available to any given individual. As a concrete example, the choice set of an unemployed individual may contain K types of programmes as well as a 'no-programme' option.

Interest lies in the causal average effect of a treatment relative to another treatment on some outcome Y. A set of potential outcomes is correspondingly associated to each of the K+1 states: Y^0 , Y^1 , ..., Y^K , with Y_i^k denoting the outcome Y for individual *i*, if *i* were to receive treatment *k*. Let $T \in \{0, 1, ..., K\}$ denote the actual assignment to a specific treatment, so that $T_i = k$ if individual *i* receives treatment *k*. Since each individual receives only one of the treatments, his remaining K potential outcomes are unobserved counterfactuals.

 $^{^{10}}$ Or equivalently, in a dose-response model, the treatment of interest is allowed to take on integer values between 0 and *K*.

Note that for this representation to be meaningful, the stable-unit-treatment-value (SUTVA)¹¹ assumption has to be fulfilled, requiring treatment status as well as all the potential outcomes of a given individual to be independent from the treatment status of others, the latter condition ruling out the possibility of general equilibrium or cross-effects.

A number of interesting parameters can now be defined (see Lechner, 2000), but in what follows, the focus will be on the generalisation of the popular 'effect of treatment on the treated': the (K+1)•*K* pair-wise comparisons of the average effect of treatment *k* relative to treatment *k*' conditional on assignment to treatment *k*, for all combinations of *k* and *k*':

$$E(Y^{k}-Y^{k'}|T=k) = E(Y^{k}|T=k) - E(Y^{k'}|T=k) \text{ for } k, k' \in \{0, 1, ..., K\}, k \neq k'.^{12}$$

In our case, this amounts to assessing the average effect for an individual registering as unemployed in Sweden of participating in programme k compared to a hypothetical state in which he received treatment k'.

The first term, the average outcome following treatment k for individuals who have participated in k, is observed in the data. This is however not the case for all the counterfactuals of the type $E(Y^{k'}|T=k)$, i.e. all the outcomes participants in k would have experienced, on average, had they taken any treatment other than k.

Identifying assumptions thus need to be invoked to overcome the fundamental missing data problem that since no individual can be in more than one state at the same time, all but one of the K+1 potential outcomes are not observed for any given individual.¹³ One such assumption often invoked in evaluation exercises is the condi-

¹¹ First expressed by Rubin (1980) and further discussed in Rubin (1986) and Holland (1986).

¹² Note that in general this parameter is not symmetric: $E(Y^{k}-Y^{k'}|T=k) \neq -E(Y^{k'}-Y^{k}|T=k')$ if participants in the two programmes systematically differ in characteristics related to the outcome.

¹³ Identification assumptions and estimation of treatment effects in non-experimental studies have been extensively looked at. Standard references in the evaluation literature include the comprehensive survey by Heckman, LaLonde and Smith (1998), as well as Heckman and Robb (1985), Heckman, Ichimura and Todd (1997, 1998), Heckman, Ichimura, Smith and Todd (1998), Rosenbaum and Rubin (1983, 1985) and Rubin (1974).

tional independence assumption (CIA), an extension of which would allow us to identify all the counterfactuals:¹⁴

$$T \perp (Y^0, Y^1, \dots, Y^K) \mid X = x, \forall x \in C^*$$

This identifying assumption (termed 'strong unconfoundedness' by Imbens, 2000) requires the existence of a set of observable characteristics X (variables unaffected by the treatments, defined as 'attributes' by Holland, 1986) such that, conditional on their values x, the treatment indicator T is independent of the entire set of potential outcomes (over the set C* of X values for which the treatment effect is defined).

Note however that a weaker form would in fact suffice to identify the conditional treatment effects we are interested in: ¹⁵

$$T \perp (Y^k, Y^{k'}) \mid X = x, \forall x \in C^*, T \in \{k, k'\}$$
 for $k, k' \in \{0, 1, ..., K\}, k > k'$ (*)

Since we are just interested in the pair-wise comparison of the various kinds of treatments, we can relax strong unconfoundedness by requiring conditional independence to hold only for the sub-populations receiving either treatment k or treatment k' (see Lechner, 2000): all the (outcome-relevant) differences between individuals choosing treatment k and those selecting into treatment k' need to be captured by variables the evaluator can control for.

The unobserved counterfactuals can thus be identified as:

$$E(Y^{k'}|T=k) = E_X [E(Y^{k'}|T=k, X)|T=k] = E_X [E(Y^{k}|T=k', X)|T=k]$$

where the inner expectation is identified due to CIA (*) and the outer expectation is taken with respect to the distribution of X for participants in k.

The latter highlights how in order to adjust for differences in X, sufficient overlap is required in the distribution of X by treatment status. In particular, all participants in k need to have a counterpart in the k'-group for each X for which we seek to make a comparison. If there are regions where the support of X does not overlap for the two

 ¹⁴ Its weaker form in terms of conditional mean independence would suffice.
 ¹⁵ Again, the requirement could just be in terms of conditional mean independence.

groups, matching has to be performed over the common support region¹⁶; the estimated treatment effect has then to be redefined as the mean treatment effect for those treated *k* falling within the common support.¹⁷

Formally, define the (generalised) propensity score as the conditional probability of receiving a given type of treatment given *X*:

$$P^{k}(X) \equiv P(T = k | X)$$

The common support requirement for all pair-wise conditional parameters then translates into:

$$0 < P^{k}(X) < 1$$
 for $X \in C^{*}$ and $k=0, 1, ..., K$.¹⁸

An important practical result by Rosenbaum and Rubin (1983) for the single treatment case ($T \in \{0,1\}$) is that the propensity score P(T=1|X), a single variable giving the probability of being treated conditional on X, provides a parsimonious way to adjust for differences in a (generally large) set of pre-treatment characteristics between treatment and non-treatment groups, formally: $T \perp X \mid P(T=1|X)$.

More generally, a balancing score b(X) is a function of X, such that conditional on it, the characteristics X are 'balanced' across the treatment groups, i.e. $T \perp X \mid b(X)$. A necessary and sufficient condition for a function of X to be a balancing score is to be at least as fine as the (generalised) propensity score $P^k(X)$:¹⁹

 $E[Pr(T=k|X)|b(X)] = P(T=k|X) \equiv P^{k}(X)$

$$0 < P^{k}(X) < 1$$
, for $k=0, 1, ..., K$.

Since we are however just interested in the separate pair-wise comparisons of the various treatments, we need to find a balancing score ensuring the balancing of the

¹⁶ Alternatively, identification would rely on (parametrically) extrapolating from regions of C^* that have positive probabilities for both the treatment states being compared to occur.

¹⁷ Note that if the treatment effect varies among individuals, restricting to the common subset may actually change the parameter being estimated.

¹⁸ To just compare treatment k with k' for participants in k, one would need to have some participants in k' with those X's at which there are participants in k, i.e. $P^{k'}(X) > 0 \quad \forall X \in C^*: P^k(X) > 0$.

¹⁹ Cf. Theorem 2 by Rosenabaum and Rubin (1983) and proposition 1 in Lechner (2000).

X's in the two sub-populations of interest for each separate comparison, say for k and k':

$$T \perp X \mid b(X), T \in \{k, k'\}$$

which is verified iff

$$E[P(T=k|X, T \in \{k, k'\})|b(X)] = P(T=k|X, T \in \{k, k'\}) \equiv P^{k/kk'}(X)$$
$$0 < P^{k/kk'}(X) < 1.$$

In our case of separate pair-wise comparisons of the various treatments, the conditioning variable (balancing score) of minimal dimension which ensures the balancing of observables in the two sub-populations of interest k and k' is thus still given by a scalar, the conditional choice probability of treatment k given either treatment k or k': ²⁰

$$P^{k|kk'}(X) = \frac{P(T = k \mid X)}{P(T = k \mid X) + P(T = k' \mid X)} \equiv \frac{P^k(X)}{P^k(X) + P^{k'}(X)}$$

Under the CIA, the required counterfactual can thus be estimated as follows:

$$E(Y^{k'}|T=k) = E_{Pk/kk'}[E(Y^{k'}|T=k', P^{k/kk'}(X))|T=k].$$

One way to apply such results is to control for systematic differences between treatment groups' observed characteristics by matching participants in k to individuals receiving treatment k' based on a balancing score b(X). For any pair of treatments k and k', under the CIA assumption that all the outcome-relevant differences between the two groups are captured by their observable characteristics, the average outcome experienced by the matched pool of k'-participants thus identifies the counterfactual outcome participants in k would have experienced, on average, had they taken treatment k' instead.

²⁰ Cf. also Brodaty, Crépon and Fougère (2000).

III.4.2 Multiple-treatment matching in the Swedish institutional set-up

As extensively discussed in the previous chapter (cf. II.5.1), an important initial clarification concerns the definition of the 'no-programme' state in Sweden. In general, sooner or later an unemployed individual will go on a programme, provided he remains unemployed 'long enough'. In other words, if unemployed individuals in Sweden are not observed to go into a programme, it can be argued that it is because they have found a job (before). Using as no-programme group those individuals who are observed not to enter any programme (thus *de facto* observed to leave the unemployment register) would *a priori* set programme participants at a disadvantage. A connected important feature in the programme selection process in Sweden is the fact that unemployed job-seekers and case-workers are most likely to take their decisions sequentially over time in unemployment. In particular, at any given moment the relevant decision is between joining a programme now or not participating for now, in the knowledge that one can always join later on. The key choice faced by the unemployed in Sweden is thus a decision between either participating in a programme now or else searching longer in open unemployment whilst availing themselves of the services offered by the employment offices. Correspondingly, when looking at the inflow into unemployment, what one can evaluate in the Swedish institutional set-up (in addition to the pair-wise comparisons of the various programmes) is the average effect of joining a given programme compared to further postponing the participation decision by not joining any programme at least up to the **T** he aim of the paper thus consists in quantifying the differential effectiveness on subsequent labour market performance (e.g. employment probability over time) of seven different types of treatments: labour market training, work experience, job introduction, relief work, trainee replacement, job subsidies and searching longer in open unemployment. As to the latter, it may be worth reminding that the 'treatment' effectively consists of the baseline assistance offered by the employment offices to the openly unemployed (Table 3.1, first row).

Implementation

In Section III.4.1 both the identification conditions and the balancing scores have been defined just taking account of the two sub-samples participating in the two treatments which are the object of a given comparison, *de facto* ignoring the multiprogramme nature of the environment the individuals face. As Lechner (2000) clearly points out, when interested in comparing two programmes for participants in one of those two, the existence of multiple treatments can in fact be ignored, since individuals who do not take part in either programme considered are not needed for identification.

However, considerable attention should be devoted to the specification of the treatment probabilities, and it is in fact their estimation which offers an opportunity to capture and take account of the multiplicity of options open to individuals.

In the Swedish context in particular, it was argued above that it is also important to model the sequential decision-making process of the individual/caseworker. A way to accomplish this is to model the effect of unemployment duration (as well as of both fixed and time-varying characteristics) on the various options open to an individual at any given point of time. In particular, all our individuals start by registering as (first-time) unemployed. At any given point U=u in their first unemployment spell (our empirical units will be months), they can 'decide' between a set of 11 exhaustive and mutually exclusive options: to participate in one of the six available programmes, to continue searching for a job full-time as openly unemployed, to find (or decide to accept) a job, to leave the labour force (in particular to go on education in the regular system), or to drop out of the unemployment register for reasons unknown to the officials. By modelling the effect of unemployment duration on exit type, one can thus simultaneously take account of the various exit routes from unemployment, of right-censoring and of the effect of time-varying characteristics on individual choices.

As to the practical implementation, the data no longer allow to model the joining decision by stratifying the sample by month in unemployment, as done in Chapter 2. The less data hungry procedure we devised consists in splitting each single individual unemployment spell of a given number of days is into monthly spells. Each of these

new sub-spells is characterised by the duration month *u* the new sub-spell refers to, by a corresponding treatment indicator and by those characteristics pertaining to, and events taking place during that u^{th} month of unemployment. The conditional probability of choosing option *k* after having spent *U* months in unemployment, $P(T_i=k \mid U_i, X_i)$, is then estimated²¹ and the corresponding balancing scores constructed.²²

To compare programme k and programme k' for participants in programme k, each k-participant is then matched²³ to that k'-participant based on this balancing score. The differential performance of the two matched groups then starts being evaluated from entry into the respective programme.

To estimate the average effect of joining a given programme k compared to waiting longer (than they have) for participants in programme k, the corresponding balancing score is calculated for each k-participant and each waiting spell. The procedure then follows closely the 'stratification' approach in proposed in Chapter 2. In particular, k-participants and waiting individuals are stratified by unemployment duration $U=1, 2, ..., U_{max(k)}$. For a given unemployment duration U=u, those kparticipants who enter the programme in their uth month are matched to the most similar individuals who are still unemployed after u months. The evaluation of the effect of joining programme k in one's uth month of unemployment compared to wait-

²¹ The conditioning set of observables X denotes fixed individual characteristics as well as timevarying characteristics both of the individual and of the macro local conditions he faces. Time-varying observables other than elapsed unemployment duration U are defined conditional on U_i or on calendar time, and include two main sets of controls: those relating to the unemployment experience of the individual so far (i.e. up to U=u) and those capturing the local conditions prevailing at U=u at the employment office of the individual. A thorough discussion of the conditioning variables is deferred to the next sub-section.

²² Since our interest in the estimation of the balancing score purely lies in its ability to balance the characteristics of the sub-groups being pair-wisely compared, the resulting matching quality has led to the choice, for each pair-wise comparison, of the 'best' specification among: multinomial logit on the full set of exits, on an aggregation of some of them; a series of binomial probits; matching finer on unemployment duration if it resulted to be balanced unsatisfactorily; matching on both participation probabilities in the case of the multinomial logit, imposing a caliper when differences in the matched scores were deemed excessive (see the Appendix for the final choice in each case).

²³ Matching estimators can be implemented in wide variety of ways (e.g. Heckman, Ichimura and Todd, 1997 and 1998, Heckman, Ichimura, Smith and Todd, 1998, Dehejia and Wahba, 1999, Rosenbaum and Rubin, 1985, Cochran and Rubin, 1973). The analyses of this paper are based on one-to-one matching, performed with replacement (since pair-wise comparisons are performed across all

ing longer than *u* months starts from entry into programme *k*, i.e. from U=u. All the $U_{max(k)}$ effects of programme *k* by unemployment duration are then averaged over the observed entry distribution into programme *k* to derive an overall average effect.

III.4.3 Plausibility of the matching approach in the Swedish context

The method just outlined relies on the central assumption that we observe - and thus can match on - all those differences between the various treatment groups that are likely to affect their outcomes.

The plausibility of the CIA should be discussed in relation to the richness of the available dataset as well as the selection process into the Swedish programmes. To this end it may be useful to separately consider:

- the decision between waiting further in open unemployment or joining a (i.e. any) programme;
- (2) the decision to choose one specific programme among the available ones.

Figure 3.2 should help clarify the discussion that follows by highlighting the agents – the unemployed job-seeker, his caseworker and the local conditions prevailing at his employment office – whose interactions determine the outcome of the selection process (i.e. whether an individual joins a programme and if yes which one), as well as how these respective influences are captured in the available data.

As to decision (1), we need to control for all variables that, conditional on having spent a given amount of time in unemployment, influence both the decision to join a programme as well as potential future labour market performance were such decision to be postponed further. Since this was the basis for the analyses in the previous chapter and was extensively discussed in Section II.5.3, here we only summarise the main issues. The decision between waiting further in open unemployment or joining a (i.e. any) programme appears to be driven by the individual's subjective likelihood of em-

⁽differently-sized) sub-samples, each sub-group will act both as a treated group and as (several) com-

Figure 3.2 Selection process into the Swedish programmes and key available *regressors*



ployment (Harkman, 2000, as reported in Carling and Richardson, 2001), which could in turn be proxied by several pieces of information characterising the recent

parison groups, entailing the need to use a given individual more than once in a given comparison).

employment history of the individuals (in particular, elapsed unemployment duration in their first unemployment spell, entitlement status, and the additional piece of information in terms of pre-unemployment wage – a summary statistic of the worker's past labour market situation). Similarly, we have controlled for factors relating both to employment prospects and either to potential returns from programme participation or affecting the opportunity cost or psychological cost of participation (age, gender, previous stock of human capital in terms of both specific and general education and job-specific experience, occupation being sought, citizenship, part-time unemployment status). Additional useful information allowing us to capture caseworker selection relates to the officials' own subjective, synthetic and evolving overall evaluation of the situation and character of their unemployed client, summarising individual traits that are potential indicators of unobserved heterogeneity.

A possible source of violation of the 'selection on observables' assumption would be the presence of hidden job offers, that is if individuals waiting longer have decided to do so because they know they will be hired shortly. This would however not constitute a serious problem if the typical time span between job offer and de-registration from the unemployment office is not too long.

Turning now to decision (2), i.e. the selection mechanism into the various programmes, the CIA requires the evaluator to have access to all the variables that influence both the choice between the programmes as well as potential future outcomes that would occur had the individual chosen an alternative programme. Note that all our individuals have access to the same choice set, the only relevant recommendation being the one requiring a certain length of the unemployment period prior to enrolment; benefit renewability rules and individual compensation while on the programmes are similarly comparable across programmes.

Harkman (2000) finds that while individual self-selection into different programmes is likely to be a minor issue in Sweden (unemployed workers tend to value the various programmes equally), the caseworkers do seem to have clear ideas about which type of programme is suitable for their clients, based on individual characteristics. Since the relevant decision-maker thus appears to be the caseworker, the only issue we need to focus on is whether he acts upon information which is unobserved to us and correlated with labour market outcomes. We do however observe not only important characteristics of the unemployed client, but also the caseworker's own subjective and synthetic evaluation of the overall situation and needs of service of his unemployed client as described above. In a sense, the caseworker reveals, updates and records in the data a synthetic appraisal of various factors, including some which may have been originally unobserved to us. Our assumption then translates into the requirement that conditional on all this information, programme assignment is unrelated to outcomes; caseworkers or employment offices act idiosyncratically given worker characteristics (based e.g. on their preferences, incentives, experiences, colleagues' opinions). Carling and Richardson (2001), who carefully examine the factors that determine in which programme the job seeker ends up into, do in fact provide reassuring evidence that the administrative selection process appears to be unrelated to the outcome.

Finally, in addition to county indicators, a set of local indicators at the individual's municipality / employment office level over time have been included to further control for the possibility that individual joining decisions and/or office-specific programme selection criteria may be based on local unobserved characteristics in turn correlated with individuals' potential labour market performance. In addition to the local 'programme-rate' (the share of registered unemployed job-seekers participating in any programme) and the local 'offer-rate' (the proportion of unemployed workers who have been offered a programme out of all openly unemployed) used in the previous chapter, we have constructed a series of single programme ratios, reflecting the programme mix at that office and at that time.

III.5 Empirical findings

In this section, differential programme effects and the effect of joining a given programme vis-à-vis waiting longer in unemployment are assessed in relation to two important outcomes: individuals' employment rates over time and the probability of being in a compensated unemployment spell over time.

III.5.1 Employment probability over time

The effect of joining programme *A* (compared either to joining another programme or to searching longer in open unemployment) on the employment probability of programme *A* participants' is calculated from the start of the programme to five years on and summarises various components: a 'lock-in' effect, an effect on the probability of finding a job and an effect on job longevity.

The differential lock-in effect of the programme vis-à-vis the comparison treatment originates from a differential job search while on the programme. Compared to open unemployment, job search is clearly reduced because less time is left due to participation itself. Different programmes may however also differentially reduce the intensity of job search while participating in the respective programme: they may for instance leave different amounts of time and energy for job search or may entail different 'promises' once completed (e.g. employment subsidies may induce participants to focus on the job at hand to 'impress' the employer in order to increase the likelihood to remain with the firm afterwards).

Differential treatment effects on job finding probabilities may originate from various channels: improved (e.g. *via* contacts and references from an employment programme) or more intense (e.g. while in full-time open unemployment) job-search; the acquisition of new marketable skills making the working option more attractive and/or the individual more in demand (e.g. *via* training); and the revelation of previously unknown individual productivity to temporary or potential employers.

Finally, a differential degree of job longevity may be the result of the different extent to which the programmes improve the individual's working habits, skills, adaptability or ability to learn on the job.

As to the effect from participating in a given programme compared to longer jobsearch as openly unemployed, all the programmes considered are found to have a negative impact on their respective participants' short-term employment prospects. As shown in Figure 3.3, joining any of these programmes initially locks participants in, reducing their chances of being in employment by an over 15 percent probability in each case. However, the more long-term effect of joining a programme is found to critically depend on the type of programme the individual has entered. In particular, for our sample of entitled adult unemployed workers it seems more worthwhile to intensively search longer in open unemployment rather than joining labour market training, ALU or relief work. Even after the programme typically ends, these participants subsequently enjoy lower employment rates than if they had postponed the joining decision further. These negative effects persist over a substantial time horizon before turning insignificant (around one and a half years in the case of training, almost three years for ALU and over four and a half years for relief work). A possible explanation, to be explored below, is that these programmes may not provide participants and especially participants entitled to unemployment benefit – with skills marketable enough to make the working option sufficiently attractive; these programmes may thus end up being typically used by entitled individuals simply as a passport to renewed eligibility.

Participants in API and in trainee replacement on the other hand are just as well off as if they had waited longer. By contrast, the decision to join a job subsidy programme rather than searching further in open unemployment results in significantly and persistently higher employment rates (up to 40 percentage points) soon after the programme typically ends.

Figure 3.3 Average effect on employment probability over time of joining the specified programme compared to waiting longer in open unemployment for participants in the specified programme, with 95% confidence intervals bands (Time in months, from programme start)



Notes: ^a Employment probability obviously refers to a regular (i.e. non-subsidised) job.

Table 3.4 summarises the above results as well as the main picture that emerges from the series of graphs plotting the differential programme effects on employment probability over time for all the pair-wise comparisons of the programmes.²⁴ Note that although later in the section ALU and API will be explicitly contrasted, in the table and the following discussion these two programmes centred on work experience have been lumped into one type of treatment, 'work practice'. The two measures have in fact a very similar overall aim, nature and implementation (in particular the requirement of not performing regular tasks), this at least formal equivalence having being sanctioned by the employment offices themselves in January 1999, when the two measures were collapsed into the new work practice scheme. ALU's additional eligibility requirement is also not binding in our sample of unemployment-benefit entitled individuals.

Compari- son↓	Training	Work practice ^a	Relief	Replacement	Subsidies
Waiting	 lock-in negative up to 30m then 0 	 lock-in, negative up to 30m for ALU only then 0 	 lock-in then negative 	lock-inthen 0	 short lock-in then large positive
Training		0	mostly 0	positive	large positive
Work practice	0		mostly 0	positive	large positive
Relief	0	0		mostly 0 positive up to 15m	large positive
Replace- ment	negative then zero from 30m	negative	0 (neg. but in- significant)		large positive
Subsidies	large negative	large negative	large negative	mostly negative	

Table 3.4 Informal summary of the various conditional average treatment effects on

 employment probability over 5-year horizon since programme start

Notes: This summary takes informal account of the statistical significance of the estimated effects; for the complete set of results, see the Appendix. m = month(s).^a ALU and API combined.

²⁴ See the Appendix for the full set of results.

Turning to the results concerning the relative performance of the different programmes, both *a priori* expectations and raw data outcomes appear to be confirmed. The star programme is again clearly job subsidies – not surprisingly, given the job promise they generally entail. Individuals having joined this programme enjoy a much higher (20 to 40 percentage points) employment probability over time than if they had joined an alternative programme. In addition, participants in any of these other programmes (with the possible exception of trainee replacement schemes) would have fared considerably better had they gone on job subsidies instead. The second best performing programme is confirmed to be trainee replacement. Since the task performed is by construction a useful one, for which the firm was willing to pay a regular employee, the presumption that this programme should teach marketrelevant skills is corroborated by the result that former deputies have considerably better outcomes than if they had joined any other of the remaining programmes (in particular, training or work practice). Conversely, trainees and work practice participants would have improved their labour market performance had they joined a replacement scheme. As to the remaining programmes – labour market training, work practice and relief work, they do not seem to perform much differently from one another.

III.5.2 Unemployment-benefit collection probability over time

Since we are looking at individuals who are entitled to unemployment benefits and for whom the eligibility-renewability property of the programmes is likely to represent a particularly attractive feature likely to affect incentives, we additionally consider the differential treatment effects on the probability of being effectively drawing unemployment compensation over time.²⁵

The performance of job subsidy participants stands out again: they are significantly less likely to be on unemployment benefits over time than if they had partici-

²⁵ The results are displayed in the Appendix.

pated in any other programme, and participants in the other programmes would have been less likely to be drawing benefits over time had they gone on a subsidised job, the only exception again being replacement schemes, participants in which do not seem to perform substantially differently in this dimension than if they had gone on subsidised jobs.²⁶ What is even more striking is the negative, mostly significant effect on the likelihood of compensated unemployment of joining employment subsidies compared to waiting longer in open unemployment.

In fact, employment subsidies is the only programme to display a negative effect on benefit collection probability compared to postponing the participation decision. While replacement schemes have a zero effect beyond the initial five months²⁷, participants in training, API, ALU and relief work all have a significantly higher likelihood of compensated unemployment over time than if they had waited longer in unemployment – clear evidence in favour of the likely role played by benefit renewability considerations in the above finding of a negative treatment effect on employment rates displayed by these latter measures.

Coming back to the pair-wise comparison of the programmes in terms of compensated unemployment probability, replacement schemes have a negative effect compared to training, but no effect compared to the other programmes. Conversely, participants in training, work practice and relief work would have been less likely to be in compensated unemployment had they joined a replacement scheme instead. Again, these three kinds of programme do not perform significantly differently from one another in terms of benefit collection probability. Interestingly, in the case of relief work and especially work practice participants, clear evidence of unemploymentprogramme 'cycling' effects is visible, with significant positive effects (compared both to some other programmes and especially to the waiting longer option) arising between the 6th and 20th month (i.e. after programme end and up to the maximum 14

²⁶ The initial positive effect from start of the replacement programme to up to 5 months reflects the fewer *direct* programme-employment transitions that deputies experience compared to subsidised workers.

²⁷ By construction, individuals do not draw unemployment compensation while on the programme.

months of compensated unemployment), and often between the 27th and 38th month (a second cycling spell, starting from the end of a second programme).

III.5.3 API versus ALU

As to the two work practice measures, their potentially different effectiveness is of particular interest, since while sharing the basic features of API, ALU is exclusively reserved to individuals entitled to unemployment benefits and has been explicitly introduced to prevent them from running out of compensation. In terms of employment probability over time, while participants in one of the two programmes would not have fared better had they joined the other programme instead (Figure 3.4A), compared to waiting longer in open unemployment the performance of ALU is considerably worse than the one of API (Figure 3.3). In addition, ALU participants display an even stronger propensity to be drawing benefits on a visibly 'cycling' basis compared to waiting longer than do API participants had they waited longer too (Figure 3.4B). The explicit, close link between entitlement renewability and programme (as institutionalised in the case of ALU) would thus seem to severely impact on the programme's effectiveness on the labour market performance of its participants.²⁸

²⁸ For more analyses of the linkages between entitlement, programme participation, benefit exhaustion and 'cycling' behaviour, see Chapter 2.

Figure 3.4 Differential performance of ALU and API





(B) Average effect on benefit collection probability over time of joining the specified programme compared to waiting longer in open unemployment for programme participants



API



Notes: Time in months, from programme start. 95% confidence intervals bands.

III.5.4 The problem of the 'lost' individuals

A final issue concerns an attrition problem in the Händel dataset, whereby a registered unemployed individual, having first missed an appointment at the official employment office and subsequently failing to contact the agency within a week, is sim-
ply de-registered – thus lost from the data – without information on whether a job has been found or whether the individual is still unemployed. Bring and Carling (2000), who have tried to trace back a sample of 'lost' individuals, have found that around half of them had in fact found a job, highlighting how seriously under-reported employment status is in the official data. More critically, though, it is quite possible that the probability of being in a lost spell over time, as well as the true status (employed *versus* unofficially unemployed) once in a lost spell may be systematically different among individuals taking the various treatments, i.e. entering one of the available programmes or searching longer in open unemployment. Although in our sample of entitled individuals this attrition problem is considerably less severe than in the full sample, almost 9% of our individuals do become 'lost' after their first (registered) unemployment spell (see Table 3.3), while the probability of being lost over time steadily rises to 12% over our 5-year horizon. It would thus seem important to check the robustness against these lost spells of the findings on employment rates.

Following the procedure of Chapter 2²⁹, the additional information from the Bring and Carling survey has been exploited to perform best- and worst-case bounds analysis on all the pair-wise comparisons of the treatments. As shown in the Appendix (Figure 3A.11, to be contrasted with Figure 3.3 above)³⁰, the conclusions discussed above remain in fact virtually unaffected, in particular regarding the positive employment effect of job subsidies and the negative ones of relief work, ALU and training.

²⁹ The conditional probability that a lost individual (*L*=1) with characteristics *X* has in reality found employment (*Y*=1) can be decomposed as: P(Y=1|X=x, L=1)=P(Y=1|X=x, L=1, D=1) P(D=1|X=x, L=1)+P(Y=1|X=x, L=1, D=0) [1–P(D=1|X=x, L=1)], where for each pair-wise treatment comparison, D=1 denotes the treatment and D=0 the comparison treatment. For each lost individual, we know his treatment status *D*, we can estimate his treatment probability given the lost status $P(D=1|X_i, L_i=1)\equiv p_i^D$ and based on the survey we can impute his misclassification probability $P(Y_i=1|X_i, L_i=1)\equiv p_i^V$. The procedure to derive worst- and best-case bounds consists in assigning $P(Y_i=1|X_i, L_i=1, D=d_i)$ by

setting $P(Y_i=1|X_i, L_i=1, D=1-d_i)$ to its maximum or minimum, compatible with the given $p_i^{D_i}$ and $p_i^{Y_i}$, as well as with all probabilities being in [0; 1].

³⁰ The full set of results is available upon request.

III.6 Discussion and conclusions

The analyses in this paper have investigated the differential performance of six main types of Swedish programmes both relative to one another and vis-à-vis more intense job search in open unemployment.

Starting from this latter comparison, the results concerning programme effects on employment and compensated unemployment have been discouraging for all the programmes considered except job subsidies (and possibly replacement schemes).

Several factors (in addition to a possible violation of the identifying assumption underlying the method chosen for analysis) may account for such disappointing findings. It might for instance be more difficult to put participants back into stable work in periods of high unemployment³¹ (though it may be argued that it is exactly in such difficult times when effective labour market programmes would be most needed). There is also the connected issue of the scale of the programmes; the massive use of large-scale programmes in the 1990s is likely to have resulted in inefficient programme administration.³²

An additional most likely explanation however relates to the use of the programmes simply as a way to re-qualify for unemployment benefits, with programmes ending up locking their participants – and in particular those entitled to unemployment compensation – in the unemployment system.

In fact, when looking at these six programmes taken as a whole compared to waiting longer in open unemployment, the results – both in terms of employment rates and of benefit collection probability over time – for the sample of entitled adults considered here are considerably worse than those obtained for the full sample and all the Swedish programmes in the previous chapter. Contrasting these two sets of results would thus lend support to the conjecture that for individuals entitled to unemploy-

³¹ Cf. e.g. the switch from positive effects for Swedish labour market training in the 1980s to negative ones in the 1990s. For more details, see Calmfors, Forslund and Hemström (2001).

³² In principle there could also be a stigma effect linked to participation in these programmes; this is however not confirmed by Swedish evidence, according to which employers view former programme participants more favourably than openly unemployed individuals. For a review of the relevant survey studies, see Calmfors, Forslund and Hemström (2001).

ment compensation, the eligibility renewability rules are likely to significantly distort the incentives for participation and thus wipe out potential productivity-enhancing effects of several programmes.

In particular, the present analysis has found that individuals joining labour market training, workplace practice schemes or relief work subsequently display lower employment rates coupled with a higher benefit collection probability than if they had searched further as openly unemployed.

As to the pair-wise comparison of the effectiveness of the six programmes, it is interesting to start by considering the work by Carling and Richardson (2001), a Swedish study most similar in aim and sample selection³³ to the present one. The present study can be seen as a 'robustness' analysis (using a different methodology from their hazard regression model), as well as complementing the previous one, in which programmes are evaluated along one dimension: their ability to reduce unemployment duration (measured from start of the programme), thus ignoring what happens once a job is found.³⁴

It is thus both reassuring and interesting to notice how their main finding is confirmed in our analyses looking at further types of outcomes. Those programmes providing (subsidised) workplace experience and on-the-job training at an employer are relatively more effective in terms of participants' subsequent labour market performance than vocational classroom training courses. In addition, the more relevant the kind of task performed, the higher the programme ranks. More specifically, the top six programmes (from the eight) emerging from their results in term of unemploy-

³³ They examine the relative efficiency of eight Swedish programmes – the same six programmes examined here plus self-employment grants and computer/activity centres – for adult unemployed becoming unemployed for their first time in slightly later years than ours (between 1995 and 1997).

³⁴ They also do not consider the impact of the option of intensive job search in open unemployment, and thus do not investigate whether participation in any programme is better or worse than postponing the participation decision. On the other hand, they examine (providing a negative answer) the issue of whether the programmes' relative efficiency is affected by how long an individual has been unemployed before joining, or if it depends on participants' demographics and skills.

ment duration (cf. their Table 3) are: 1. job subsidies, 2. trainee replacement, 3. work practice (API), 4. labour market training, 5. relief work and 6. work practice (ALU).³⁵

Even more generally, the underlying similarity of results across studies looking at different countries with varying labour market structures and policies³⁶ may indicate a general validity of the overall conclusions.

Coming back to the present evaluation, the best performer overall is undisputedly employment subsidies, followed by trainee replacement. As to the remaining types of programmes, they do not seem to perform in a significantly different way between one another.³⁷

Turning to the cost side, it is quite remarkable to notice how the ranking of the programmes in terms of their effectiveness is almost perfectly reversed when taken in terms of their expensiveness (1. labour market training, 2. ALU, 3. relief work, 4. trainee replacement, 5. API and 6. job subsidies).

It is however important not to jump at the hasty conclusion that employment subsidies are *the* solution – the most effective programme as well as the cheapest. Several types of issues can be raised to point out potential problems both in terms of the effective magnitude of the uncovered effects and in terms of their general applicability should the scope of the programme be extended.

As to the scope of the analysis, it is important to bear in mind that the programme's effects have been evaluated for a rather specific sub-group of the population – the declared target group of individuals who have been relatively long in unemployment (although note that Carling and Richardson (2001) find their results unaffected by time spent in unemployment prior to participation).

A second issue concerns the validity of the identifying CIA assumption for participants in this programme: since job subsidies generally entail the 'promise' of a job, it

³⁵ It is not surprising that ALU performs worst in terms of (subsequent) unemployment duration. Cf. end of Section 5.

³⁶ For a summary of other Swedish evidence in line with the present results, see the review by Calmfors, Forslund and Hemström (2001). For OECD countries see the review by Martin and Grubb (2001) and e.g. Gerfin and Lechner (2000) for Switzerland, Brodaty, Crépon and Fougère (2000) and Bonnal, Fougère and Sérandon (1997) for France, and Ridder (1986) for the Netherlands.

is likely that potential candidates are considered quite carefully. Even though we control for a host of factors likely to underlie the case-worker's judgement and despite Carling and Richardson's (2001) finding of no selection bias for this programme, it may still be the case that subsidised participants are slightly 'better' on average than matched comparisons. Nonetheless, it would be hard to argue that selection bias (also possibly in the form of anticipatory effects) could account for all of the large positive effects seemingly displayed by job subsidies in terms of all comparisons and outcomes considered.

Even if the direction of the estimated effects may appear reliable, however, it may not be possible or even desirable to focus attention and funds on this kind of measure. As to the sheer possibility of extending it, scope is in fact limited: the public sector cannot use such grants, and following EU regulations in 1997 neither do employers in the synthetic fibre, automotive, steel, shipyard, fishery and transport industries.

Apart from legal feasibility, the desirability of a widespread use of this measure may not be warranted once it is considered that our estimates ignore potential indirect and general equilibrium effects which may spill over to other groups. In particular, substitution would take place if participants in the employment subsidy programme were to take (some of) the jobs that participants in the other programmes or 'waiting' unemployed individuals would have been offered in the absence of the subsidies. The impact of the subsidy would thus be at the expense of worsened conditions either for participants in the other programmes or for openly unemployed individuals finding it more difficult to get jobs or getting worse jobs. The estimated effect would in this case overestimate the net impact of the subsidy programme. Both survey and econometric Swedish studies do in fact find sizeable (around 65-70 percent) direct displacement effects arising from those Swedish programmes that generate subsidised employment.³⁸

Finally, it is obviously unthinkable to generalise such a measure to *all* unemployed job-seekers: it would simply become just a way to subsidise firms' hirings, resulting

³⁷ As to the two work practice programmes, the only significant (and quite predictable) difference lies in a stronger likelihood for ALU participants to be drawing benefits on a 'cycling' basis.

in huge dead-weight effects (i.e. subsidising hiring that would have taken place anyway).

In the light of the present and previous results and of the above considerations, a more promising measure might appear to be trainee replacement schemes. Still among the cheapest programmes, it was shown to perform quite satisfactorily. In fact, it shares some of the features likely to be at the root of the success of job subsidies (short of the job promise): in terms of the present temporary employment, it provides relevant job-specific training and can be used as a cheap screening device of individ-ual unobserved productivity. At the same time it sends out a message that the individual has been gaining (or maintaining) relevant skills, thus making the job seeker more attractive to potential future employers, who value the fact that a job is being performed in the regular competitive market. Finally, our partial-equilibrium estimates are likely to be an underestimate of the programme's effect, since they do not take into account the 'double-dividend' effect arising from the possibility offered to the replaced employees of increasing their human capital through training.

Nevertheless, even though at first sight the potential of this programme appears particularly promising, a few issues need once again to be considered. The 'double-dividend' from the subsidised training of the replaced employee may in fact often turn out to be dead-weight loss instead³⁹, while Harkman, Johansson and Okeke (1999) found evidence of dead-weight in terms of the deputies as well, with a large share of participants alternating between regular short-term jobs and trainee replacement with the *same* employer. Finally, survey studies have in fact uncovered displacement effects of the same order as employment subsidies (e.g. AMS, 1998).⁴⁰

In conclusion, the present analysis unambiguously joins previous micro studies in finding that the closer to regular, relevant employment in the competitive labour market, the higher the programme's benefits to its participants. It is however essential to

³⁸ For more details, see Calmfors, Forslund and Hemström (2001).

³⁹ Since 80 to 90 percent of employers taking part in the scheme are within sectors (health care and related branches in the public sector) with a long-standing system for further training funded by the employer, it seems likely that a good part of the sponsored training would have occurred anyway. (I thank Anders Harkman for this information.)

⁴⁰ 42 percent as an average across survey studies, see Calmfors, Forslund and Hemström (2001).

consider these findings in the light of those arising from the macroeconomic literature, which has widely documented that exactly for these types of programmes the potential for negative crowding-out and dead-weight effects is largest. Taken together⁴¹, the various results clearly highlight the difficult trade off faced by labour market policy.

⁴¹ See in particular Calmfors, Forslund and Hemström (2001).

III.7 Appendix

Treated	Comparisons	Specification	Median bias	Bias in U
		÷		
training	waiting	MNL, all states	1.85	_
experience	waiting	MNL, all states	2.12	_
introduction	waiting	MNL, all states	3.30	_
relief	waiting	MNL, all states	2.55	_
replacement	waiting	MNL, all states	2.35	_
subsidy	waiting	Probit	3.07	-
training	work practice	Probit, finer on U	3.90	0.41
training	relief	MNL, all states	6.57	3.58
training	replacement	Probit, finer on U	5.55	2.77
training	subsidy	Probit, finer on U	7.98	2.74
work practice	training	Probit, finer on U	4.20	0.73
work practice	relief	MNL, fewer states	4.86	1.78
work practice	replacement	Probit, finer on U	6.08	2.63
work practice	subsidy	Probit, finer on U	4.67	1.04
relief	training	Probit, finer on U	3.16	1.48
relief	work practice	Probit, finer on U	2.79	0.09
relief	replacement	Probit	7.00	1.36
relief	subsidy	MNL, all states, finer on U	7.83	3.22
replacement	training	Probit, finer on U	4.18	0.06
replacement	work practice	Probit, finer on U	5.26	0.98
replacement	relief	Probit, finer on U	11.16	1.53
replacement	subsidy	Probit	9.85	5.17
subsidy	training	Probit, finer on U	3.14	1.18
subsidy	work practice	Probit, finer on U	3.40	0.66
subsidy	relief	Probit, finer on U	6.14	0.29
subsidy	replacement	MNL, all states, finer on U	9.29	0.93
experience	introduction	MNL, all states, finer on U	10.87	2.5
introduction	experience	MNL, all states, finer on U	4.36	0.31

Table 3A.1 Specification chosen and indicators of resulting matching quality

Notes: MNL: multinomial logit model. Finer on U: Mahalanobis-metric matching on the balancing score and unemployment duration. Median bias: median overall absolute percentage bias, where the median is taken over the post-matching absolute standardised differences of 70 variables in estimation of the choice model (the various programme rates are excluded from calculation of the median). For a given regressor, the standardised difference after matching is defined as the difference of the sample means in the treated and matched comparison sub-samples as a percentage of the square root of the average of the sample variances in the treated and comparison groups (cf. Rosenbaum and Rubin, 1985).

Figures 3A.1-5

Differential average effects on **employment probability** over time of the specified programme compared to the various alternatives for participants in the specified programme.

(percentage points; e.g. 0.2 is a 20 percentage points higher probability; *t*-axis: months since joining the programme)

Figure 3A.1 TRAINING compared to ... for individuals taking training



Figure 3A.2 WORK EXPERIENCE compared to ... for individuals taking work experience



Figure 3A.3 RELIEF compared to ... for individuals taking relief





Figure 3A.4 REPLACEMENT compared to ... for individuals taking replacement TRAINING WORK EXPERIENCE

Figure 3A.5 SUBSIDIES compared to ... for individuals taking subsidies TRAINING WORK EXPERIENCE



Figures 3A.6-10

Differential average effects on **compensated unemployment probability** over time of the specified programme compared to the various alternatives for participants in the specified programme.

(percentage points; *t*-axis: months since joining the programme)

Figure 3A.6 TRAINING compared to ... for individuals taking training



Figure 3A.7 WORK EXPERIENCE compared to ... for individuals taking work experience



Note: See Figure 3.4 in the main text for the effect of the two work practice schemes compared to waiting longer.



Figure 3A.8 RELIEF compared to ... for individuals taking relief

120

50 55 60

-.6 -0 5 10 15 20 25 30 35 40 45



Figure 3A.9 REPLACEMENT compared to ... for individuals taking replacement



Figure 3A.10 SUBSIDIES compared to ... for individuals taking subsidies

Figure 3A.11 Average effect on employment probability over time of joining the specified programme compared to waiting longer in open unemployment for participants in the specified programme: estimated effect and best- and worst-case bounds (Time in months, from programme start)



Summary and conclusions to Part One

The first research question set out in the introduction to Part One and explored in the first chapter concerned the effectiveness of the Swedish unemployment-programme system in improving individual labour market opportunities during the recession of the 1990s. The evidence has proved in fact rather mixed; individuals joining a programme are found to subsequently enjoy higher employment rates but also to be more likely to draw unemployment benefits over time than if they had searched longer in open unemployment.⁴²

Comparing the main lessons arising from studies performed at different times (in particular the switch from positive effects of Swedish labour market training in the 1980s to negative ones in the 1990s⁴³), it may thus seem that the collection of measures that appeared to be quite effective in a low-unemployment environment may no longer be so successful if applied – and on a massive scale – in periods of severe economic downturns. On the other hand, it may also be argued that it is exactly in such difficult times of high unemployment when effective labour market programmes are be most needed to place participants back into work. Similarly, the large scale at which programmes have been administered may have prevented the efficient management and tailoring of the various measures.

Nevertheless, possibly the most critical factor appears to have been the link between the programme system and the unemployment benefit system, an interaction quite likely to intensify in periods of high unemployment and unstable labour market conditions. Various pieces of evidence concerning this link have been combined and discussed. In particular, the evidence for individuals entitled to unemployment benefits provided a sharp contrast to the findings relating to non-entitled individuals, quite

⁴² It may be worthwhile to stress again that these programme effects do not relate to the effect of joining a programme compared to *never* joining any, but rather compared to delaying participation at least some more time further while searching for a job in open unemployment.

⁴³ For more details, see Calmfors, Forslund and Hemström (2001).

unmistakably pointing to distorted incentives behind programme participation as a most likely force behind the disappointing programme effects.

All of these considerations thus raise the important issues as to whether there may be more efficient means of providing (sustained) unemployment compensation, as well as whether some programme expenditure could be more effectively redirected, for instance towards market-based incentives to stimulate labour demand (e.g. by decreasing payroll taxes).

The second question motivating the first part of the thesis concerned the possibility of scrutinizing the Swedish experience in order to derive some general lessons as to which type of programme works best. The answer that has emerged from Chapter 3 is that those programmes most similar to regular employment rank unambiguously highest, an overall conclusion not only in line with other Swedish analyses, but also with studies looking at different countries with varying labour market structures and policies.

In particular, while on employment subsidies or trainee replacement the participant performs a task that is by construction a useful one, one for which the firm is willing to pay a regular employee. These programmes should thus teach demonstrably market-relevant skills, in contrast to e.g. labour market training with its emphasis on the classroom-based acquisition of new skills which are deemed to be – or soon to become – in demand. A second advantage of these programmes is that they can be used as a cheap screening device of the participant's initially unknown productivity in a regular task. For employment subsidies – with their informal promise of a job – there is thus a valuable opportunity for mutually trying out the likely future employment relationship on a low-cost basis. For trainee replacement, a signal is sent out to potential employers that the individual has been gaining (or at least maintaining) relevant skills. Participants in either type of programme are thus likely to become more attractive to potential employers, who value the fact that a job is being performed in the regular competitive market.

By contrast, it appears that the formal vocational skills taught by labour market training, as well as the working skills, additional work experience, improved working habits, fresh contacts and references that relief work and the two work practice schemes are intended to provide may not be relevant – and thus valuable – enough to fetch a return on the labour market, or at least not one high enough to make the work-ing option more attractive, this being particularly the case for individuals entitled to unemployment benefits. These types of programme are then likely to be regarded just as a gateway to renewed benefit eligibility, ending up locking their participants – in particular their entitled participants – in the unemployment system.

Despite the ranking of the programmes in terms of their effectiveness being almost perfectly reversed when viewed in terms of their expensiveness, it is however important not to jump to hasty conclusions as to which programmes should attract most public funds. As discussed in the previous chapter, for job subsidies, apart from an increasingly restricted legal possibility of extension, both survey and econometric Swedish studies have found sizeable direct displacement effects. Similarly, broadening such a measure is bound to lead to substantial dead-weight effects. For trainee replacement too, dead-weight losses have been both suspected in terms of the sponsored training, as well as documented in terms of the deputies, while survey studies have uncovered displacement effects of the same order as employment subsidies.

In conclusion, labour market policy-makers are confronted with a difficult trade off: although the results obtained, perfectly in line with previous micro evidence, have found that a programme's benefits to its participants are highest the more it resembles regular employment, several macroeconomic studies have uncovered large and negative displacement and dead-weight effects exactly for this type of programme.

Although increased income has never been an explicit objective of the Swedish labour market policy, further research to corroborate and expand the results obtained will consider post-programme earnings or wages – an important measure of potential programme effects on individual productivity, which would allow a more complete picture of the comparative performance of former participants who have found a job. In addition, the results obtained in these two chapters rely on a non- (or semi-) para-

126

metric technique which assumes selection on observables. Despite the richness of the available dataset, the discussion in Sections II.5.3 and III.4.3 has highlighted some potentially remaining sources of bias. The robustness of the conclusions obtained should thus be assessed by resorting to an alternative approach aimed at identifying a structural econometric model explicitly modelling the sequence of choices facing individuals and taking into account the endogeneity of the selection of unemployed workers to specific programmes, which are intertwined with unemployment benefits eligibility and renewability.

References

- Ackum Agell, S. (1995), "Swedish labour market programmes: Efficiency and timing", *Swedish Economic Policy Review*, 2, 65-98.
- Ackum Agell, S., Björklund, A. and Harkman, A. (1995), "Unemployment insurance, labour market programmes and repeated unemployment in Sweden", *Swedish Economic Policy Review*, 2, 101-128.
- Agell, J. and Lundborg, P. (1999), "Survey evidence on wage rigidity and unemployment: Sweden in the 1990s", IFAU Working Paper 1999:2, Office of Labour Market Policy Evaluation, Uppsala.
- AMS (1998), "Undanträngningseffekter av arbetsmarknadspolitiska åtgärder enenkätundersökning ur både arbetssökande- och arbetsgivarperspektiv", Ura 1998:8, AMS.
- Bonnal, L., Fougère, D. and Sérandon, A. (1997), "Evaluating the impact of French employment policies on individual labour market histories", *Review of Economic Studies*, 67, 683-713.
- Bring, J. and Carling, K. (2000), "Attrition and misclassification of drop-outs in the analysis of unemployment duration", *Journal of Official Statistics*, 4, 321-330.
- Brodaty, T., Crépon, B. and Fougère, D. (2000), "Using matching estimators to evaluate alternative employment programmes: Evidence from France, 1986-8", Labour Economics Discussion Paper 2604, CEPR, London.
- Calmfors, L. Forslund A. and Hemström, M. (2001), "Does active labour market policy work? Lessons from the Swedish experiences", *Swedish Economic Policy Review*, forthcoming ***.
- Carling, K. and Gustafson, L., (1999), "Self-employment grants *versus* subsidised employment: Is there a difference in the re-unemployment risk?", IFAU Working Paper 1999:6, Office of Labour Market Policy Evaluation, Uppsala.
- Carling, K. and Larsson, L. (2000a), "Utvärdering av arbetsmarknadsprogram i Sverige: Rätt svar är viktigt, men vilken var nu frågan?", *Arbetsmarknad&Arbetsliv*, 6, 185-192.
- Carling, K. and Larsson, L. (2000b), "Replik till Lars Behrenz och Anders Harkman", *Arbetsmarknad&Arbetsliv*, 6, 278-281.
- Carling, K. and Richardson, K. (2001), "The relative efficiency of labour market programmes: Swedish experience from the 1990s", IFAU Working Paper 2001: 2, Office of Labour Market Policy Evaluation, Uppsala.

- Carling, K., Edin, P.-A., Harkman A. and Holmlund, B. (1996), "Unemployment duration, unemployment benefits, and labour market programmes in Sweden", *Journal of Public Economics*, 59, 313-334.
- Carling, K., Holmlund, B. and Vejsiu, A. (2001), "Do benefit cuts boost job findings?", IFAU Working Paper 1999: 8, Office of Labour Market Policy Evaluation, Uppsala, forthcoming in the *Economics Journal*.
- Cochran, W. and Rubin, D.B. (1973), "Controlling bias in observational studies", *Sankyha*, 35, 417-446.
- Dehejia, R.H. and Wahba, S. (1988), "Propensity score matching methods for nonexperimental causal studies", NBER Working Paper No.6829.
- Dehejia, R.H. and Wahba, S. (1999), "Causal effects in non-experimental studies: reevaluating the evaluation of training programmes", *Journal of the American Statistical Association*, 94, 1053-1062.
- European Commission (1998), From Guidelines to Actions: The National Action Plans for Employment, Directorate-General for Employment, Industrial Relations and Social Affairs, Luxemburg.
- Fisher, R.A. (1935), *The Design of Experiments*, Edinburgh: Oliver&Boyd.
- Forslund, A. and Krueger, A.B (1995), "An evaluation of the Swedish active labour market policy New and received wisdom", *NBER/SNS Project: Reforming the Welfare State*, Occasional Paper 65.
- Frölich, M., Heshmati, A. and Lechner, M. (2000), "A microeconometric evaluation of rehabilitation of long-term sickness in Sweden", SSE/EFI Working Paper Series in Economics and Finance, No.373.
- Gerfin, M. and Lechner, M. (2000), "Microeconometric evaluation of the active labour market policy in Switzerland", Discussion Paper 2000-10, Volkswirtschaftliche Abteilung, Universität St. Gallen.
- Hägglund, P. (2000), "Effects of changes in the unemployment insurance eligibility requirements on job duration Swedish evidence", IFAU Working Paper 2000:4, Office of Labour Market Policy Evaluation, Uppsala.
- Hallström, N.-E. (1994), "Genomförandet av åtgärden arbetslivsutveckling (alu). En studie i sex kommuner i tre län", Mimeo, Department of Culture and Social Sciences, Linköping University.
- Ham, J.C. and LaLonde, R.J. (1996), "The effect of sample selection and initial conditions in duration models: Evidence from experimental data on training", *Econometrica*, 64, 175-205.
- Harkman, A. (2000), "Vem placeras i åtgärd?", Mimeo, Office of Labour Market Policy Evaluation, Uppsala.

- Harkman, A., Johansson, A. and Okeke, S. (1999), "Åtgärdsundersökning 1998", Ura 1999:1, AMS.
- Heckman, J.J. and Robb, R. (1985), "Alternative methods for evaluating the impact of interventions", in Heckman, J.J. and Singer, B. (eds.), *Longitudinal Analysis of Labour Market Data*, Cambridge University Press, 156-246.
- Heckman, J.J., Ichimura, H. and Todd, P.E. (1997), "Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme", *Review of Economic Studies*, 64, 605-654.
- Heckman, J.J., Ichimura, H. and Todd, P.E. (1998), "Matching as an econometric evaluation estimator", *Review of Economic Studies*, 65, 261-294.
- Heckman, J.J., Ichimura, H., Smith, J.A. and Todd, P.E. (1998), "Characterising selection bias using experimental data", *Econometrica*, 66, 1017-1098.
- Heckman, J.J., LaLonde, R.J. and Smith, J.A. (1998), "The economics and econometrics of active labour market programmes", in Ashenfelter, O. and Card, D. (eds.), *The Handbook of Labour Economics*, 3, Ch.31, Amsterdam: North-Holland.
- Holland, P.W. (1986a), "Statistics and causal inference", *Journal of the American Statistical Association*, 81, 396, 945-960.
- Holland, P.W. (1986b), "Rejoinder", *Journal of the American Statistical Association*, 81, 396, 968-970.
- Imbens, G. (2000), "The role of propensity score in estimating dose-response functions", *Biometrika*, 87, 706-710.
- Jansson, F. (1999), "Rehires and unemployment duration New evidence of temporary layoff on the Swedish labour market", Ura 1999:10, AMS.
- Johansson, P. and Martinson, S. (2000), "The effect of increased employer contacts within a labour market training programme", IFAU Working Paper 2000:10, Office of Labour Market Policy Evaluation, Uppsala.
- Larsson, L. (2000), "Evaluation of Swedish youth labour market programmes", IFAU Working Paper 2000:1, Office of Labour Market Policy Evaluation, Uppsala.
- Layard, R., Nickell, S. and Jackman, R. (1991), Unemployment, Macroeconomic Performance and the Labour Market, Oxford University Press.
- Lechner, M. (1996), "The Effects of Enterprise-Related Continuous Vocational Training in East Germany on Individual Employment and Earnings", Discussion Paper 542-96, Mannheim University: Beiträge zur angewandten Wirtschaftsforschung.
- Lechner, M. (1999a), "Earnings and Employment Effects of Continuous Off-the-Job Training in East Germany after Unification", *Journal of Business and Economic Statistics*, 17, 74-90.

- Lechner, M. (1999b), "An Evaluation of Public-Sector Continuous Vocational Training Programmes in East Germany", University of St. Gallen, Mimeo, September.
- Lechner, M. (2001), "Identification and estimation of causal effects of multiple treatments under the conditional independence assumption", in Lechner, M., Pfeiffer, F. (eds), *Econometric Evaluation of Labour Market Policies*, Heidelberg: Physica/Springer, 43-58.
- Lundin, M. and Skedinger, P. (2000), "Decentralisation of active labour market policy: The case of Swedish Local Employment Service Committees", IFAU Working Paper 2000:6, Office of Labour Market Policy Evaluation, Uppsala.
- Manski, C.F. (1990), "Non-parametric bounds on treatment effects", *The American Economic Review*, 80, 2, Papers and Proceedings of the Hundred and Second Annual Meeting of the American Economic Association, 319-323.
- Martin, J.P. and Grubb, D. (2001), "What works and for whom: A review of OECD countries' experiences with active labour market policies", *Swedish Economic Policy Review*, forthcoming***.
- Melkersson, M. (1999a), "Policy programmes only for a few? Participation in labour market programmes among Swedish disabled workers", IFAU Working Paper 1999:1, Office of Labour Market Policy Evaluation, Uppsala.
- Melkersson, M. (1999b), "Unemployment duration and heterogeneous search behaviour among Swedish disabled workers", IFAU Working Paper 1999:5, Office of Labour Market Policy Evaluation, Uppsala.
- Neyman, J. (with co-operation by Iwaszkiewicz, K. and Kolodziejczyk, S.) (1935), "Statistical Problems in Agricultural Experimentation" (with discussion), Supplement of the Journal of the Royal Statistical Society, 2, 107-180.
- OECD (1996), The OECD Jobs Strategy: Enhancing the Effectiveness of Active Labour Market Policies, OECD, Paris.
- Quandt, R. (1972), "Methods for Estimating Switching Regressions", *Journal of the American Statistical Association*, 67, 306-310.
- Regnér, H. (1997), "Training at the job and training for a new job: Two Swedish studies", Dissertation Series 29, Swedish Institute for Social Research, Stockholm University.
- Ridder, G. (1986), "An event history approach to the evaluation of training, recruitment and employment programmes", *Journal of Applied Econometrics*, 1, 109-126.
- Rosenbaum, P.R. and Rubin, D.B. (1983), "The central role of the propensity score in observational studies for causal effects", *Biometrika*, 70, 41-55.

- Rosenbaum, P.R. and Rubin, D.B. (1985), "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score", *The American Statistician*, 39, 33-38.
- Roy, A. (1951), "Some thoughts on the distribution of earnings", *Oxford Economic Papers*, 3, 135-146.
- Rubin, D.B. (1974), "Estimating causal effects of treatments in randomised and non-randomised studies", *Journal of Educational Psychology*, 66, 688-701.
- Rubin, D.B. (1980), "Discussion of 'Randomisation analysis of experimental data in the fisher randomisation test" by Basu, *Journal of the American Statistical Association*, 75, 591-593.
- Rubin, D.B. (1986), "Discussion of 'Statistics and causal inference" by Holland, Journal of the American Statistical Association, 81, 396, 61-962.
- Schmidt, C.M. (2000), "Arbeitsmarktpolitische Maβnahmen und ihre Evaluierung: Eine Bestandaufnahme", IZA Discussion Paper 207, Bonn.
- Swedish Institute (1997), Swedish Active Labour Market Policy, *Fact Sheets on Sweden*, Stockholm, June.

PART TWO

Estimating the Returns to Education

Introduction

There are at least three distinct ways of defining the 'returns to education': the private return, the social return and the labour productivity return. The first of these is made up of the costs and benefits to the individual and is clearly net of any transfers from the state and any taxes paid. The second definition highlights any externalities or spill-over effects and includes transfers and taxes. The final definition simply relates to the gross increase in labour productivity (or growth). A key component of each of these measures is the impact of education on earnings. This is perhaps the aspect of returns to education measurement where statistical methods have been most developed and most fruitfully deployed and is the central focus of the second part of the thesis.

With extensive data available over time and individuals on schooling and on earnings, the measurement of the effect of education on earnings is one area where we might expect agreement. However, a casual look through the literature on the impact of education on earnings reveals a wide range of estimates and an equally wide range of empirical approaches that have been adopted to estimate the return. So why do the estimates vary so widely and what is the most appropriate empirical method to adopt? The answer to these two questions provides the central motivation for this paper. It is illustrated using the sample of men from the NCDS Birth Cohort data for the UK. This data source provides a uniquely rich source of non-experimental data on family background, educational attainment and earnings. We argue that it is ideally suited to analyse statistical methods for the measurement of the effect of education on earnings.

The appropriate statistical method to adopt will depend, in a rather obvious way, on the chosen model for the relationship between education and earnings. We distinguish two broad characterisations of this specification. The first relates to the measurement of education. In particular whether we can summarize education, or human capital more generally, in a single measure – years of schooling, for example. This is commonly referred to as a one factor model. It is a restrictive framework since it assumes that, as returns to education change over time, it is

only the single aggregate that matters and there are no differential trends in returns for different education levels. It is convenient though since we can simply include a single education measure in an earnings equation.

We will refer to different education levels as to different treatments borrowing a common notation from the evaluation literature. A single treatment specification refers to the impact of a specific educational level – such as undertakings higher education. A multiple treatment model will distinguish the impact of many different education levels, thus allowing different educational levels to have separate effects on earnings. In general the multiple treatment model would seem a more attractive framework since we will typically be interested in a wide range of education levels with very different returns. However, we will also consider models with a single discrete treatment such as the impact of a specific qualification and models with a single overall education level such as years of schooling.

The second characterisation relates to the distinction between heterogeneous and homogeneous returns. In simple terms, whether the response coefficient on the education variable(s) in the earnings equation is allowed to differ across individuals. To allow this to happen according to observables is a relatively straightforward extension of the homogeneous model, but to allow the heterogeneity to be unobservable to the econometrician but acted upon by individuals completely changes the interpretation and the properties of many common estimation approaches. Chapter IV starts with this distinction between model specifications and use it to define parameters of interest in the earnings education relationship.

Even where there is agreement on the model specification there are alternative statistical methods which can be adopted. With experimental data the standard comparison of control and treatment group recovers an estimate of the average return for the treated under the assumption that the controls are unaffected by the treatment. Although experimental design is possible and growing in popularity in some studies of training, for large reforms to schooling and for measuring the impact of existing educational systems, nonexperimental methods are essential. There are broadly two categories of nonexperimental methods: those that attempt to control for correlation between individual factors and schooling choices by way of an excluded instrument, and those that attempt to measure all individual factors

that may be the cause of such dependence and then match on these observed variables. Whilst the feasibility of these alternative methods clearly hinges on the nature of the available data, their implementation and properties differ according to whether the model is one of heterogeneous response and whether schooling is represented through a single or multiple measure. The different properties of these estimators and the drawbacks to each method are the object of Chapter IV.

Going back to our initial questions, the above discussion should have highlighted how no given nonexperimental estimator is always superior to all others or always appropriate for any application; the choice between the various estimation methods should by contrast be guided by the postulated model for the outcome and selection processes, the corresponding parameter of interest to be recovered, as well as the richness and nature of the available data in the application at hand.

The various models and non-experimental estimators are then compared in Chapter 5 using the British NCDS data. In particular, we first consider the return from undertaking some form of higher education, then move on to estimating the returns to education for three levels: 1) leaving after completing O-levels or its vocational equivalent; 2) leaving after completing A-levels or its vocational equivalent; 3) undertake some form of higher education (including sub-degree level higher education). Only results for men are presented so as to conserve space and to focus on the earnings effect versus the employment effect.

CHAPTER IV

Models and Methods

IV.1 The earnings-education relationship: A general set-up and alternative models

For each individual i=1, ..., n, we let y_i represent their earnings or hourly wage opportunities in work (generally expressed in logs). To begin with we will assume that we are measuring earnings at one point in time for a sample of individuals who have completed formal schooling. A good illustration to keep in mind is from the British cohort studies where a single cohort is followed through education and employment and sampled at specific intervals usually several years apart. We measure their earnings when they are 33 years of age and ask: what is the impact on earnings at age 33 of different schooling outcomes?

This returns to education problem can be fruitfully framed in the context of the evaluation literature (see in particular Heckman and Robb, 1985, and Heckman, LaLonde and Smith, 1999). The problem of estimating the effect of schooling on earnings can in fact be viewed as a specific application of the general evaluation framework: the measurement of the causal impact of a generic 'treatment' on an outcome of interest. Indeed, some of the more recent developments in the returns literature, for example those which use matching estimators or social experiments, relate explicitly to similar approaches in evaluation.

In order to cover a fairly flexible representation of schooling we will consider the *multiple treatment* case of a finite set of schooling levels available to any given individual. Write the exhaustive set of J + 1 treatments (schooling levels) under examination as 0,1,...,J and denote actual receipt of schooling level j by individual i as $S_{ji} = 1$. These will typically be defined in some natural sequence of binary indicator variables: $S_{0i} = 1$ would represent the base educational level, $S_{1i} = 1$ if the individual stopped at the first stage of schooling, $S_{2i} = 1$ if the next stage in the sequence is completed and so on. For example, in the UK context $S_{0i} = 1$ may refer to dropping out at the minimum school leaving age, $S_{1i} = 1$ stopping at O-levels, $S_{2i} = 1$ finishing with at least one A-level and $S_{3i} = 1$ achieving a first degree.

Of course, one can imagine a finer sequence, with S_{ji} representing completion of *j* years of schooling by individual *i*. Similarly, one could envisage a possible set of non-sequential outcomes. All the methods discussed below are easily extended to more complicated situations but will typically require more demanding data requirements and modelling assumptions to estimate the "causal" impact on earnings.

One can think of a set of potential outcomes associated to each of the *J*+1 treatments: y_i^0 , y_i^1 , ..., y_i^j , where y_i^j denotes the (log) earnings of individual *i* were *i* to receive schooling level *j*. The problem of estimating the returns to education can be phrased as the evaluation of the causal effect of one schooling level *j* relative to another (without loss of generality, let this be treatment 0) on the outcome considered, *y*. In terms of the notation established above, interest thus lies in recovering quantities of the form $(y_i^j - y_i^0)$, averaged over some population of interest. The researcher might for example be interested in the average return to schooling level *j* for the whole population, $E[y_i^j - y_i^0]$, or alternatively for those who did achieve that level, $E[y_i^j - y_i^0 | S_{ji} = 1]$.

Since however each individual receives only one of the treatments, his remaining *J* potential outcomes – the earnings he would command had he attained a different educational level – are unobserved counterfactuals, and $(y_i^j - y_i^0)$ as well as the parameters defined above are not observed in the data. At the core of the evaluation problem, including its application to the returns to education framework, is thus the attempt to estimate missing data.

The observed outcome of individual *i* can be written as:

$$y_{i} = y_{i}^{0} + \sum_{j=1}^{J} (y_{i}^{j} - y_{i}^{0}) S_{ji}$$
(1)

It is worth pointing out that for this representation to be meaningful, the stable unit-treatment value assumption (SUTVA) needs to be satisfied (Rubin, 1980 and for further discussion Rubin, 1986, and Holland, 1986). This assumption requires that an individual's potential outcomes as well as the chosen education level are independent from the schooling choice of other individuals in the population (thus ruling out cross-effects or general equilibrium effects, as well as peer effects in educational choices). A related feature worth highlighting is that each of the relationships being discussed will typically be specific to a particular time period and location. For example, if equation (1) refers to the impact of education levels on the earnings of British men aged 33 in 1991, it will be unlikely to be stable across time periods and countries. The returns will depend on the earnings set in the labour market and will in turn depend on the demand and supply of individuals with these differing human capital attributes. This point, although quite obvious, is often misunderstood in the context of predicting returns to education.

Equation (1) is extremely general; to relate the alternative models and estimation methods discussed below, it is appropriate to specify it a bit further. Suppose there are a set of observed covariates X_i (e.g. early test scores, demographic variables, aspects of the local labor market). Letting potential outcomes depend on both observable factors and unobserved ones, write the still general functions:

$$y_i^j = f_j(X_i, u_{ji})$$
 for $j = 0, 1, ..., J$. (2)

Note that implicit in (2) is the requirement that the observables X be exogenous in the sense that their potential values do not depend on treatment status, or equivalently, that their potential values for the different treatment states coincide $(X_{ji} = X_i \text{ for } j=0,1,...J)$. Natural candidates for X that are not determined or affected by treatments S are time-constant factors, as well as pre-treatment characteristics.

Assuming additive separability between observables and unobservables, we can write

$$\mathbf{y}_i^j = m_j(X_i) + u_i^j$$

with $E[y_i^j | X_i] = m_j(X_i)$, i.e. assuming that the observable regressors X are unrelated to the unobservables u. We will maintain these exogeneity assumptions on the X 's throughout.

Let now the state-specific unobservable components of earnings be written as

$$u_i^j = \alpha_i + \varepsilon_i + \beta_{ji}$$
 for $j = 0, 1, ..., J$

with α_i representing some unobservable individual trait, like ability or motivation, which affects earnings for any given level of schooling, β_{ji} measuring the individual-specific unobserved marginal return to schooling level *j* relative to level 0 in terms of the particular definition of earnings y_i (for convenience let us normalised β_{0i} to 0) and ε_i capturing measurement error in earnings (measurement error in the schooling or education variable S_i is also likely to be important and will be discussed in terms of the alternative approaches to estimation).

Given the still rather general specification, equation (1) for observed earnings becomes:

$$y_{i} = m_{0}(X_{i}) + \sum_{j=1}^{J} (m_{j}(X_{i}) - m_{0}(X_{i}))S_{ji} + \sum_{j=1}^{J} (u_{i}^{j} - u_{i}^{0})S_{ji} + \alpha_{i} + \varepsilon_{i}$$

$$= m_{0}(X_{i}) + \sum_{j=1}^{J} b_{j}(X_{i})S_{ji} + \sum_{j=1}^{J} b_{ji}S_{ji} + \alpha_{i} + \varepsilon_{i}$$

$$= m_{0}(X_{i}) + \sum_{j=1}^{J} \beta_{ji}S_{ji} + \alpha_{i} + \varepsilon_{i}$$
(3)

with $\beta_{ji} \equiv b_j(X_i) + b_{ji}$.

In this still very general set-up, β_{ji} , the private return to schooling level *j* (relative to schooling level 0) is allowed to be heterogeneous across individuals in both observable and unobservable dimensions: $b_j(X_i)$ represents the return for individuals with characteristics X_i and thus captures observable heterogeneity in returns, while b_{ji} represents the individual-specific unobserved return to schooling level *j*, conditional on X_i . Typically we would assume the α_i and b_{ji} to have a finite population mean and variance. In the following the population means are labelled α_0 and b_{j0} respectively.

Two central aspects in the empirical investigation of the earnings return to human capital investments are now considered. First among these is the distinction between the *homogeneous returns* and *heterogeneous returns* model. In the homogeneous returns model the rate of return to gross earnings of a particular human capital investment is the same for all individuals. Growing statistical evidence and causal empiricism suggests that the homogeneous returns restriction is unwarranted.

Secondly, it is appropriate to set apart the *one factor* model of human capital. In this specification, all schooling is thought of as an investment in a single homogeneous construct called human capital, of which each additional unit has the same return.

An example of an empirical model that is both one-factor and homogeneousreturns is the popular linear regression equation – log earnings regressed on years of schooling. The constant parameter on the schooling variable is equivalent to homogeneous returns and the use of years of schooling as a single measure of schooling is equivalent to a single measure of human capital.

IV.1.1 The homogeneous returns model

In the homogeneous returns framework, the rate of return to a given schooling level *j* is the same across individuals, that is $\beta_{ji} = \beta_j$ for all individuals *i*.

In the case of a finite set of schooling levels (specific discrete educational levels as in our application, or even finer with each level representing a year of education), the *multiple treatment model* (3) becomes:

$$y_{i} = m_{0}(X_{i}) + \beta_{1}S_{1i} + \beta_{2}S_{2i} + ... + \beta_{J}S_{Ji} + \alpha_{i} + \varepsilon_{i}$$
(4)

where α_i represents differing relative levels of earnings across individuals for any given level of schooling and the β_i 's measure the impact of schooling level *j* relative to the base level. Although the returns to a given level are homogeneous across individuals, the different schooling levels are allowed to have different impacts on earnings.

The latter is no longer the case in the *one factor human capital model*, where it is assumed that education can always be aggregated into a single measure, say years of schooling, $S_i \in \{0, 1, ..., J\}$. To see why, consider how the popular specification

$$y_i = m_0(X_i) + \beta S_i + \alpha_i + \varepsilon_i \tag{5}$$

can be obtained from our general set-up (3) with the various treatment levels as

years of education (so that $S_i = \sum_{j=1}^{J} j \cdot S_{ji}$ with $S_{ji} \equiv 1_{(S_i=j)}$) by assuming the linear relationship $\beta_{ji} = \beta_j = j \cdot \beta$, that is that the (homogeneous) return to *j* years of schooling is simply *j* times the return to one year of schooling, or equivalently, $\beta_{j+1,i} - \beta_{ji} = \beta$ for all j = 0, 1, ..., J, that is that each additional year of schooling has the same marginal return.

A last useful specification which can be obtained from (3) by setting J=1 is the *single treatment model*, the aim of which is to recover the causal impact of a single type of schooling level. Let schooling S_{Ii} for individual *i* be defined as a binary indicator variable representing the successful achievement of a particular education level – such as obtaining a qualification, obtaining an A level, or undertaking higher education. Compared to the full multiple-treatment specification (3), this model allows us to directly formulate a specific question by the appropriate choice of the treatment and the benchmark; for example, one might be interested in the return to staying on relative to dropping out, or in the return to achieving a degree compared to *not* obtaining a degree, and so on.

In the homogeneous returns model, this single treatment specification can thus be expressed as:

$$y_i = m_0(X_i) + \beta S_{1i} + \alpha_i + \varepsilon_i$$

where β is the return to achieving the education level under consideration (relative to educational level 0 as chosen for $S_{Ii}=0$).

Note that although in these homogeneous returns models β_{ji} is constant across all individuals, α_i is allowed to vary across *i* to capture the differing productivities (or abilities or earnings levels) across individuals with the same education levels. Since educational choices and thus attained educational levels are likely to differ according to productivity (or expected earnings levels more generally), the schooling variable *S* is very likely to be correlated with α_i and this in turn will induce a bias in the simple least squares estimation of β . In addition, if *S* is measured with error there will be some attenuation bias. We will return to these estimation issues in more detail below.
IV.1.2 The heterogeneous returns model

Despite the preponderance of the homogeneous returns model in the early literature, the recent focus has been on models allowing for heterogeneous returns (examples include Heckman, Smith, and Clements, 1997, Dearden, 1999a and 1999b, and Blundell, Dearden, Goodman and Reed, 2000). Once the return is allowed to vary across individuals, the immediate question in this type of model concerns the parameter of interest. Is it the average of the individual returns? If so what average? Is it the average in the population whether or not the educational level under consideration is achieved – the *average treatment effect* – or the average among those individuals actually observed to achieve the educational level – the *average treatment on the treated* – or the average among those who have not achieved that educational level – the *average treatment on the non-treated*? In some cases a particular estimation method will recover a *local average treatment effect*, measuring the return for an even smaller subgroup of individuals. We discuss all these in greater detail in the next section.

In the general framework (3), the return to schooling level j is allowed to be heterogeneous across individuals in both observable and unobservable dimensions.

It should be pointed out that it is rather straightforward to generalise models such as (4) or (5) to allow for the *observable heterogeneity* $b_j(X_i)$. Returns can be made to depend on observables X_i in a quite arbitrary way, with the precise form chosen depending on the richness of the data set and the particular problem at hand. For instance, if the returns are assumed to depend on X_i in a linear fashion, the interactions of X_i with the educational variable(s) will enter the regression specification (this was done e.g. in Dearden, 1999a, using the same NCDS data as in this paper).

The heterogeneity we thus now focus our attention on is *unobserved heterogeneity* across individuals in the response parameter β . This person-specific component of the return may be observed by the individual but is unobserved by the analyst.

Consider first the single treatment model. A general relationship between the level of education under examination and earnings is then written as

$$y_{i} = m_{0}(X_{i}) + \beta_{i}S_{1i} + \alpha_{i} + \varepsilon_{i}$$

$$= m_{0}(X_{i}) + (b(X_{i}) + b_{0})S_{1i} + (b_{i} - b_{0})S_{1i} + \alpha_{i} + \varepsilon_{i}$$
(6)

where b_i can be thought of as random coefficients representing the heterogeneous relationship between educational qualification S_{1i} and earnings, conditional on observables X_i ($b_0 \equiv E[b_i]$ denoting its population mean).

The parameter of interest will be some average of $b(X_i) + b_i$, where the average is taken with respect to the distribution of X in the sub-population of interest; the resulting parameter will thus measure the average return to achieving education level S_i for this group. Examples are the average effect of treatment on the treated, $\beta_{ATT} \equiv E[b(X_i) + b_i | S_{1i} = 1]$, the average treatment effect in the population, $\beta_{ATE} \equiv E[b(X_i)] + b_0$, or the average effect on the non-treated, $\beta_{ATNT} \equiv E[b(X_i) + b_i | S_{1i} = 0]$.

As we mentioned in the homogeneous models above, the dependence of the schooling level(s) on the unobserved "ability" component α_i is critical in understanding the bias from the direct comparison of groups with and without education level S_i . An additional central issue in determining the properties of standard econometric estimators in the heterogeneous effects model is whether or not schooling choices S_{1i} depend on the unobservable determinants of the individual's marginal return from schooling b_i , conditional on observables X_i . If given the information in the available X, there is some gain b_i still unobserved by the econometrician but known in advance (or predictable) by the individual when making his or her educational choices, then it would seem sensible to assume that choices will, in part at least, reflect the return to earnings of that choice. As mentioned before, however, b_i is likely to vary over time and will depend on the relative levels of demand and supply, so that the dependence of schooling choices on marginal returns is not clear-cut. Some persistence in returns is however likely and so some correlation would seem plausible.

The discussion of heterogeneous returns extends easily to the multiple treatment model (4) as:

$$y_{i} = m_{0}(X_{i}) + \beta_{1i}S_{1i} + \beta_{2i}S_{2i} + ... + \beta_{Ji}S_{Ji} + \alpha_{i} + \varepsilon_{i}$$
(7)

We will also want to discuss the one factor model in which S_i enters as a single continuous variable:

$$y_i = m_0(X_i) + \beta_i S_i + \alpha_i + \varepsilon_i \tag{8}$$

In fact, the three basic specifications (6), (4) and (7) will form the main alternatives considered here, the single discrete treatment case (6) being the baseline specification.

IV.2 The earnings-education relationship: Alternative methods

IV.2.1 An overview

The aim of this section is to investigate the properties of alternative estimation methods for each of the model specifications considered above.

It is useful to start by considering a naive estimator in the general framework (3) of the returns to educational level *j* (relative to level 0) for individuals reaching this level: the simple difference between the observed average earnings of individuals with $S_{ii}=1$ and the observed average earnings of individuals with $S_{0i}=1$.

This observed difference in conditional means can be rewritten in terms of the average treatment on the treated parameter ATT (what we are after) and the bias potentially arising when the earnings of the observed group with $S_{0i}=1$ ($y_0 | S_{0i}=1$) are used to represent the counterfactual ($y_0 | S_{ji}=1$):

Naive estimator
$$\equiv E(y_i | S_{ji}=1) - E(y_i | S_{0i}=1)$$

= $E(y_{ji} - y_{0i} | S_{ji}=1) - \{ E(y_{0i} | S_{ji}=1) - E(y_{0i} | S_{0i}=1) \}$
= ATT - $\{ \text{ bias } \}$

The key issue is that when estimating the return to a particular educational choice we are likely to observe only optimal decisions, resulting in the sample of individuals who make each choice being not random. If this is ignored and individuals for whom the choice was optimal are simply compared to those for whom it was not, the estimates will suffer from bias. Using experimental data, Heckman, Ichimura, Smith and Todd (1998) provide a very useful breakdown of this bias term:

$$bias \equiv E(y_{0i} | S_{ii}=1) - E(y_{0i} | S_{0i}=1) = B_1 + B_2 + B_3 \quad (9)$$

The first two components arise from differences in the distribution of observed characteristics *X* between the two groups: B_1 represents the bias component due to non-overlapping support of the observables and B_2 is the error part due to misweighting on the common support, as the resulting empirical distributions of observables are not necessarily the same even when restricted to the same support. The last component, B_3 , is the true econometric selection bias resulting from "selection on unobservables", in our notation α_i , b_{ji} and ε_i .

Of course, a properly designed and ideally implemented randomized experiment would eliminate the bias discussed above. Pure education or schooling experiments are however very rare. It is difficult to persuade parents or the students themselves of the virtues of being randomised out of an education programme – except for rather minor programmes. Our application will be to the main stages of educational level in the UK and randomised assignment is unavailable.

Instead, our focus will be on non-experimental approaches. Many alternative methods are available and all have been widely used. Each of these methods uses observed data together with some appropriate identifying assumptions to recover the missing counterfactual. Depending on the richness and nature of the available data and the postulated model for the outcome and selection processes, the researcher can thus choose among the alternative methods the one most likely to avoid or correct the sources of bias outlined above.

More common than randomised experiments are the socalled "natural" experiments. The idea is to find real-world events which assign individuals to different levels of schooling in a random way, that is independently of any characteristics that affect earnings. This is for instance the case where some educational rule or qualification level (say minimum schooling leaving age) is exogenously changed for one group but not another. Provided the groups are representative samples from the population, then this simple comparison can recover a parameter of interest like the average treatment effect. Where the samples differ in their ability levels or other characteristics it may still be possible to recover an average effect for those who experience the change in rules.

The essence of this natural experiment approach is thus to provide a close ap-

proximation to a randomized trial by exploiting an exogenous change in schooling that only affects a subsample of the target group. This relates to instrumental variable methods more generally, where some variable (or transformation of the data) has to be found that can vary schooling independently of the heterogeneity terms.

The control function approach too takes advantage of the existence of a determinant of schooling which can legitimately be omitted from the earnings equation. This variable is exploited to estimate an additional equation that determines which educational choice is made, which is then used to augment the earnings regression with selectivity variables reflecting the selection bias.

An alternative to using instruments to control for correlation between individual factors and schooling choices is the matching method. This method attempts to measure all individual factors that may be the cause of such dependence and then purge the relationship between schooling and earnings of any important observed heterogeneity that would lead to bias.

The initial setting for the discussion of the three broad classes of alternative methods we consider – instrumental variable, control function and matching – will be based on the biases that occur from the simple application of ordinary least squares to the estimation of each of the model specifications described in the previous section. As mentioned, the primary model specification will be the single discrete treatment heterogeneous returns model (6), but the extension to the multiple treatment model (7) will also be considered and so will the specific issues that occur in the one factor-years of schooling specification (8). In each of these, the complications that are engendered by allowing the return parameter β to be heterogeneous – and acted upon by individuals – will be central to the discussion.

IV.2.2 Least Squares

Consider the single treatment model looking at the impact of a given educational level S_I . For example, in the British context $S_{Ii}=1$ may refer to those undertaking some form of higher education, with $S_{Ii}=0$ identifying those who drop out of the education system before. The model is to be estimated for a given population (defined for instance as all those individuals entering schooling at a particular date).

In the heterogeneous case, specification (6) is

$$y_i = m_0(X_i) + b(X_i)S_{1i} + b_iS_{1i} + \alpha_i + \varepsilon_i$$
(10)

There are several potential sources of bias in the least squares regression of log earnings on schooling to recover average treatment effects. Borrowing on the bias decomposition highlighted in (9):

Bias due to observables: mis-specification

First of all, note that to parametrically implement (6), the functional form for both $E[y_i^0 | X_i] \equiv m_0(X_i)$ and $E[y_i^j - y_i^0 | X_i] \equiv b(X_i)$ needs to be specified. The usual least squares specification would generally control linearly for the set of observables $\{S_{1i}, X_i \equiv [X_{1i} \dots X_{Mi}]'\}$, that is, it would be of the form:

$$y_i = \gamma' X_i + b S_{1i} + \eta_i$$

Now, in any situation where the true model included higher-order terms of the *X* variables, or interactions between the various *X*'s, or where the effect of schooling varied according to the *X*'s, the OLS coefficient on the schooling dummy variable S_I would be biased. The bias would be given by $\gamma_k \phi_{kS_1.X}$, where γ_k is the impact of the omitted term *k* on earnings *y* (conditional on the included regressors *X*) and $\phi_{kS_1.X}$ is the effect of S_I on the omitted term given *X*. Note that if the return to schooling did depend say on a variable X_m , so that the true model required the inclusion of $X_{mi}S_{1i}$ (i.e. $\gamma_{X_mS_1} \neq 0$), the OLS estimate of *b* would certainly be biased, since by construction $\phi_{(X_mS_1)S_1.X} \neq 0$.

This issue of mis-specification is linked to the source of bias B_2 – not appropriately reweighing the observations to fully control for the difference in the distribution of X over the common region, as well as to source B_1 – lack of sufficient overlap in the two groups' densities of X. The OLS approximation of the regression function over the non-overlapping region, is purely based on the chosen (in our example linear) functional form; in other words, for individuals outside the common support, the OLS identification of the counterfactual crucially relies on being based on the correctly specified model.

Bias due to unobservables

Gathering the unobservables together in equation (6), we have

$$y_i = m_0(X_i) + \beta_{ATE}(X_i)S_{1i} + e_i \quad \text{with} \\ e_i \equiv \alpha_i + (b_i - b_0)S_{1i} + \varepsilon_i$$
(11)

$$y_{i} = m_{0}(X_{i}) + \beta_{ATT}(X_{i})S_{1i} + w_{i} \text{ with} w_{i} \equiv \alpha_{i} + (b_{i} - E[b_{i} | X_{i}, S_{1i} = 1])S_{1i} + \varepsilon_{i}$$
(12)

where $\beta_{ATE}(X_i) \equiv b(X_i) + b_0$ and $\beta_{ATT}(X_i) \equiv b(X_i) + E[b_i | X_i, S_{1i} = 1]$.

Running a correctly specified OLS regression will produce a biased estimator of either parameter of interest if there is correlation between S_{1i} and the error term e_i or w_i , i.e. $E[e_i | X_i, S_{1i}]$ and $E[w_i | X_i, S_{1i}]$ may be non-zero. Such correlation may arise from different sources:

(i) Ability bias. This arises due to the likely correlation between S_{Ii} and the α_i term. If higher-ability or inherently more productive individuals tend to acquire more education, the two terms will be positively correlated, inducing an upward bias in the estimated average return β_{ATE} or β_{ATT} .

(ii) **Returns Bias**. This occurs when the individual returns b_i themselves are correlated with the schooling decision S_{Ii} . The direction of this bias is less clear and will depend on the average returns among the sub-population of those with schooling level $S_{Ii}=1$. Indeed, if ability bias is negligible, $E[\alpha_i | X_i, S_{1i}]=0$, the ability heterogeneity is unrelated to the unobserved return and returns bias is the only remaining bias present, $E[b_i | X_i, S_{1i} = 1] \neq 0$, then (11) and (12) show how the least squares coefficient on S_{Ii} will be biased for β_{ATE} but will recover the average treatment on the treated β_{ATT} , that is the average returns in the sample of those with $S_{Ii}=1$.

(iii) Measurement Error Bias. This refers to measurement error in the schooling variable S_{1i} . This may be due to random misclassification error. As usual, measurement error of this kind will induce attenuation bias in the regression coefficient and an under-estimate of the returns parameter. For the purposes of much of the discussion we can redefine ε_i to include measurement error in the schooling variables(s). In the homogeneous returns model the second bias is, by definition, absent. This is the case that is much discussed in the literature (especially in the one factor-years of schooling model (5)), where the upward ability bias may be partially offset by the attenuation measurement-error bias, and this trade-off was at the heart of the early studies on measuring gross private returns (see Griliches, 1977, and Card, 1999, for example). Indeed, there is some evidence of a balancing of biases, in which case OLS fortuitously consistently estimates the return coefficient $b(X_i) + b_0$.

Much of the practical discussion of the properties of least squares bias depends on the richness of other control variables that may be entered to capture the omitted factors. Indeed, the method of matching, described below, takes this one step further by trying to eliminate the imbalance of observables by matching observations with similar covariates. One simple recommendation in the use of least squares seems thus worth following: add in a rich set of controls, carefully specify the model and try to find separate measures of the education variable that may not suffer from the same measurement error. The rich set of controls may help reduce the ability bias; allowing a flexible model specification and interacted $X \cdot S_1$ terms may lessen bias arising from differences in the distribution of observed characteristics between the education groups and from omitted (observably) heterogeneous returns; and the second measure of schooling may be used to purge the measurement error. But, in the end, a comparison with the alternative methods - instrumental variables, control functions and matching – is always helpful in assessing how to interpret a least squares estimate of education returns. As mentioned above, the first two require some excluded instrument which determines education choices but not earnings¹, while the matching method requires an extensive set of observable characteristics on which to match. All place strong demands on data.

¹ Strictly speaking, the control function method could rely on functional form assumptions only.

IV.2.3 Instrumental variable methods

The Instrumental Variable (IV) estimator seems a natural method to turn to in estimating returns – at least in the homogeneous returns model. The third source of bias in (9) – and the most difficult to avoid in the case of least squares – arises from the correlation of observable schooling measures with the unobservables in the earnings regression. If an instrument can be found that is correlated with the true measure of schooling and uncorrelated with the unobservable ability, heterogeneity and measurement error terms, then surely a consistent estimator of the returns is achievable. This turns out to be true in the homogeneous returns model but not, except for certain special cases described below, for the heterogeneous returns model.

Even in the homogeneous returns model, finding a suitable instrument is not an easy task, since it must satisfy the Instrumental Variable criteria of being correlated with the schooling choice while correctly excluded from the earnings equation.

To more formally investigate the properties of the IV estimator, consider the general heterogeneous model (6) which also allows for $b(X_i)$. Note that without loss of generality this observably-heterogeneous return $b(X_i)$ can assumed to be linear in the *X* variables, so that $b(X_i) S_{1i} = b_X X_i S_{1i}$, where b_X is the vector of the additional returns for individuals with characteristics *X*. Note again that in this framework b_i captures the individual idiosyncratic gain (or loss) and has population mean of b_0 . The model can thus be written as:

$$y_{i} = m_{0}(X_{i}) + b_{X}X_{i}S_{1i} + b_{0}S_{1i} + e_{i}$$
(13)
with $e_{i} = \alpha_{i} + \varepsilon_{i} + (b_{i} - b_{0})S_{1i}$

Now define an instrumental variable Z_i and assume that it satisfies the orthogonality conditions:

IV:A1
$$E[\alpha_i | Z_i, X_i] = E[\alpha_i | X_i] = 0$$

IV:A2 $E[\varepsilon_i | Z_i, X_i] = E[\varepsilon_i | X_i] = 0$
IV:A3 $E[b_i - b_0 | Z_i, X_i] = E[b_i - b_0 | X_i] = 0$

With a valid instrument Z_i , one may envisage two ways of applying the IV method to estimate model (13).

<u>IV method (A)</u> uses the extended set of instruments Z_i and $Z_i X_i$ to instrument S_{Ii} and $X_i S_{Ii}$. It needs sufficient variation in the covariance of the interactions of X_i and S_{Ii} and the interactions of X_i and Z_i . Note however that this approach does not fully exploit the mean independence assumptions IV:A1 and IV:A2.

<u>IV method (B)</u> by contrast recognises that under the conditional mean independence assumptions, application of IV is equivalent to replacing S_{1i} with its prediction in *both* its linear and its interactions terms. To see this, assume:

IV:A4 $E[S_{1i} | Z_i, X_i]$ is a non-trivial function of Z for any X.

Taking the conditional expectation of (13) under assumptions IV:A1, A2, and A4 and noting that $E[X_iS_{1i} | Z_i, X_i] = X_i E[S_{1i} | Z_i, X_i]$ yields:

$$E[y_i | Z_i, X_i] = m_0(X_i) + (b_X X_i + b_0) E[S_{1i} | Z_i, X_i] + E[(b_i - b_0)S_{1i} | Z_i, X_i]$$
(14)

First of all, note that in the absence of interactions $b(X_i)$, the two IV methods are identical. A second crucial remark is that irrespective of the method chosen, there is nothing in assumptions IV:A1–A4 that makes the final term in (14) disappear. Since the error term e_i in (13) contains the interaction between the endogenous schooling dummy the unobserved individual return, neither way of applying IV would produce consistent estimates. In fact, assumption IV:A3 that the instrument be uncorrelated with the unobservable components has not been used, and nor would it help further on its own. Further and stronger assumptions are in fact needed. An alternative venue is to redefine the parameter to be identified – the Local Average Treatment Effect. We discuss this alternative strategy below. First, we consider some special cases based on stronger assumptions.

IV.2.3.1 IV in the homogeneous one factor model

Consider the one factor or "years of schooling" model (8). This is a case where specifying assumption IV:A4 as follows would seem plausible:

$$S_i = f(Z_i, X_i; \pi) + v_i$$
 where $E(v_i | Z_i, X_i) = 0$ (15)

and a consistent estimate of π can be obtained from OLS on the reduced form. Now assume that returns are homogeneous, at least conditional on *X*: b_i is constant for all *i* and equal to its average value, b_0 . Consequently, the problematic last term in (14) is zero by definition and under IV:A1, A2, and A4, instrumental variable estimation can produce a consistent estimator of $b(X_i)+b_0$.

Note how in general the IV estimators needs to deal with the specification of $m_0(X_i)$ and $b(X_i)$, just like least squares, and is thus subject to potential misspecification bias.

A final remark applies to the case of a 'pure' homogeneous returns model, that is a model where returns do not vary even by X, so that $b(X_i) = b$. In this case, the exact same IV estimator can be obtained from either a regression of log earnings y_i on predicted schooling \hat{S}_i , or, equivalently, from a regression of log earnings y_i on schooling S_i including the reduced form error v_i as an additional regressor:

 $y_i = m_0(X_i) + (b + b_0)S_i + \rho_{ev}\hat{v}_i + \zeta_i$ where $E[\zeta_i | S_i, X_i, v_i] = 0$ and $\hat{v}_i = S_i - \hat{S}_i$.

This augmented regression framework for IV is popular for testing the exogeneity assumption ($H_0: \rho_{ev} = 0$) and generalises to binary choice and censored regression settings (see Smith and Blundell, 1986).

The estimation of the return to an additional year of schooling by the inclusion of v_i in this 'pure' homogeneous returns specification is also exactly equivalent to the control function approach. As we will show below, this analogy between IV and control function breaks down outside the one-factor, no-interactions homogeneous returns model.

IV.2.3.2 IV in the heterogeneous one factor model: A special case

Consider again the one factor or "years of schooling" specification (8), this time allowing for heterogeneous individual returns b_i . To simplify the discussion, assume however that there is no heterogeneity in the returns according to X, that is assume $b(X_i) = b$. Now, following Wooldridge (1997), use all the assumptions IV:A1-A4 and further assume that S_i – determined by (15) – relates to b_i according to: IV:A5 $v_i = \rho_{vb}(b_i - b_0) + \eta_i$ with $E[(b_i - b_0)^2 | Z_i, X_i] = \sigma_b^2$.

These assumptions imply

$$E[(b_i - b_0)S_i | Z_i, X_i] = \rho_{vb}\sigma_b^2$$

so that (14) becomes

$$E[y_i | Z_i, X_i] = m_0(X_i) + \rho_{\nu\beta}\sigma_{\beta}^2 + (b+b_0)\hat{S}_i = \tilde{m}_0(X_i) + (b+b_0)\hat{S}_i$$
(16)

and the IV estimator will, under these assumptions, consistently estimate the average return $b+b_0$ (but not the intercept of the m_0 function).

Note that this is very specific to the continuous schooling measure S_i in the one factor model, since then the additively separable model for (15) is a reasonable specification for the reduced form. Even so the homoskedasticity assumption imbedded in IV:A5 is very strong.

IV.2.3.3 IV in the heterogeneous single treatment model

We now turn to the general heterogeneous returns model with a single binary treatment (6). As discussed above, assumptions IV:A1-A4 are not enough in this case. Incidentally, it may be worth pointing out that the second equalities in IV:A1 and A2 – $E[\alpha_i | X_i] = E[\varepsilon_i | X_i] = 0$ – follow from the exogeneity assumption made on X at the start, but are however not required for the IV estimation in the case of the binary treatment S_{Ii} we are considering. Let us reiterate IV:A4:

IV:A4 For each X, $E[S_1 | Z, X] = P[S_1 = 1 | Z, X]$ is a non-trivial function of Z, in particular, the instrument takes on at least two distinct values z' and z'' which differently affect the schooling participation probability.

Add now the additional property that for the treated, the instrument Z is not correlated with the individual-specific component of the return b_i (conditional on *X*). Formally:

IV:A6 $E[b_i | Z_i, X_i, S_{1i} = 1] = E[b_i | X_i, S_{1i} = 1]$

Under IV:A1, A2, A4b and A6, taking expectations we get:

$$E[y_i | Z_i, X_i] = m_0(X_i) + (b(X_i) + E[b_i | X_i, S_{1i} = 1]) P(S_{1i} = 1 | Z_i, X_i)$$

From which we can recover the conditional effect of treatment on the treated:

$$\widehat{\boldsymbol{\beta}}_{IV}(X) = \frac{E[y_i \mid X_i, Z_i = z'] - E[y_i \mid X_i, Z_i = z'']}{P[S_{1i} = 1 \mid X_i, Z_i = z'] - P[S_{1i} = 1 \mid X_i, Z_i = z'']} = b(X_i) + E[b_i \mid X_i, S_{1i} = 1] = \widehat{\boldsymbol{\beta}}_{ATT}(X)$$
(17)
$$= E[y_i^1 - y_i^0 \mid X_i, S_{1i} = 1] = \widehat{\boldsymbol{\beta}}_{ATT}(X)$$

Assumption IV:A6 is however a strong one: while allowing for heterogeneous returns b_i , it requires schooling decisions to be unrelated to these individual gains. In particular, since IV:A4b requires the schooling participation probability to depend on Z, IV:A6 rules out that this probability depends on b_i as well. To see why, model by contrast the schooling decision as:

$$S_{1i} = 1(Z_i \pi_z + X_i \pi_x + b_i + v_i \ge 0)$$

For individuals choosing to acquire the schooling level under examination, we know that $b_i \ge -Z'_i \pi_z - X'_i \pi_x - v_i$.

Thus, for the treated, Z is correlated with b_i given X, clearly violating IV:A6. In other words, for the S_{1i} =1 individuals, the value of the instrument reveals something on their average expected gains b_i and thus on their expected earnings y_i . This violates the basic IV assumption that the instrument, conditional on S_{1i} , does not affect outcomes.

Before turning in the next sub-section to the issues that emerge when schooling choices are allowed to depend on b_i , it may be worth pointing out that even where the IV estimator does produce a consistent estimate of the returns parameter of interest – as in the three cases just described –, there remains the issue of efficiency and of weak instruments. Efficiency concerns the imprecision induced in IV estimation when the instrument has a low correlation with the schooling variable. The weak instrument case is an extreme version of this where the sample correlation is very weak and the true correlation is near to zero. In this case IV will tend to the biased OLS estimator even in very large samples (see Bound, Jaeger and Baker, 1995 and Staiger and Stock, 1997).

The Local Average Treatment Effect

In the general heterogeneous returns model with a single treatment (6), even when we allow individuals to partly base their education choices on their individualspecific gain b_i , it is still possible to provide an interesting interpretation of the IV estimator – although it does not estimate the average treatment on the treated or average treatment parameter. The interpretation of IV in this model specification was precisely the motivation for the Local Average Treatment Effect of Imbens and Angrist (1994).

Suppose there is a single discrete binary instrument $Z_i \in \{0,1\}$. For example, a discrete change in some educational ruling that is positively correlated with the schooling level S_{1i} in the population. There will be four subgroups of individuals: those who do not take the education level under consideration whatever the value of the instrument (the 'never-takers'), those who always choose to acquire it (the 'always-takers'), and those who are induced by the instrument to change their behaviour: either in a perverse way (the 'defiers') or in line with the instrument (the 'compliers'). This last group is of particular interest: it is made up of those individuals who are seen with education level $S_{1i}=1$ after the rule change ($Z_i=1$) but who would not have had this level of schooling in the absence of the rule change ($Z_i=0$). To be more precise we define the events

$$D_{1i} \equiv \{S_{1i} \mid Z_i = 1\}$$
$$D_{0i} \equiv \{S_{1i} \mid Z_i = 0\}$$

and assume, in addition to the exclusion restrictions concerning the unobservables in the base state (IV:A1 and A2) and to the non-zero causal effect of Z on S_{1i} (IV:A4b – i.e. the instrument must actually change the behaviour of some individuals):

LATE:A1 For all *i*, either $[D_{1i} \ge D_{0i}]$ or $[D_{1i} \le D_{0i}]$ (note that due to IV:A4b, strict inequality must hold for at least some *i*).

This 'monotonicity' assumption requires the instrument to have the same directional effect on all those whose behaviour it changes, *de facto* ruling out the possibility of either defiers or compliers. Assume in particular that $D_{1i} \ge D_{0i}$ (*Z* makes it more likely to take S_I and there are no defiers); in this case the standard IV estimator (17)

$$\frac{E[y_i \mid X_i, Z_i = 1] - E[y_i \mid X_i, Z_i = 0]}{P[S_{1i} = 1 \mid X_i, Z_i = 1] - P[S_{1i} = 1 \mid X_i, Z_i = 0]}$$

reduces to

$$= b(X_i) + \frac{E[b_i S_{1i} | X_i, Z_i = 1] - E[b_i S_{1i} | X_i, Z_i = 0]}{P[D_{1i} = 1, D_{0i} = 0]}$$

$$= b(X_i) + \frac{E[b_i (D_{1i} - D_{0i}) | X_i]}{P[D_{1i} > D_{0i}]}$$

$$= b(X_i) + \frac{E[b_i (D_{1i} - D_{0i}) | X_i, D_{1i} > D_{0i}]}{P[D_{1i} > D_{0i}]} P[D_{1i} > D_{0i}]$$

$$= b(X_i) + E[b_i | X_i, D_{1i} > D_{0i}]$$

$$= E[y_i^1 - y_i^0 | X_i, D_{1i} > D_{0i}]$$

This provides a useful interpretation for IV: it estimates the average returns among those induced to change behaviour because of a change in the instrument – the Local Average Treatment Effect (LATE). For example, suppose $Z_i=0$ reflected a bad financial event for the family at the time the education decision was being made, while $Z_i=1$ would denote the absence of such a shock. Then IV would pick out the average marginal return among those taking schooling level S_1 when their family does not experience financial difficulties, but who would not achieve that schooling level in the presence of a bad shock.

Note that LATE simply avoids invoking the strong assumption IV:A6 (LATE:A1 is not intended to replace it; indeed, this monotonicity assumption highlighted by LATE is implicit in typical schooling choice models of the form $S_{1i} = 1(Z'_i \pi_z + X'_i \pi_x + v_i \ge 0)$). In fact, Angrist, Imbens and Rubin (1996) note that IV:A6 would here amount to assuming that the return is the same for always-takers and compliers, or in other words, that it is the same for all the treated, which are made up of these two groups. If one is not willing to make this assumption, which would identify the treatment on the treated parameter as in (17), then the only causal effect to be identified is LATE, that is the effect for compliers.

An interesting special case when panel data are available is worth mentioning. Consider Z_i as a data transformation, in particular a differencing instrument, applied to the treated. Allowing for the general case of heterogeneous returns on which individuals potentially base their choices, the *before-after* estimator simply takes the difference in mean outcomes for the group of treated individuals before and after the treatment occurs. In our framework, this transformation sweeps away any common individual component fixed over time (α_i and ε_i in our framework) and just picks out the average of b_i among those individuals with $S_{Ii}=1$ – precisely the treatment on the treated parameter. In this case, the LATE assumptions are satisfied by construction for the instrument denoting the before ($Z_i=0$) and after ($Z_i=1$). For everyone, $D_{0i}=0$ by construction, while $D_{Ii}=1$ for the treated and $D_{Ii}=0$ for the non-treated. Thus LATE:A1 is satisfied, and the group for whom $D_{Ii}>D_{0i}$ is exactly the treated group. In this case, then, the local average treatment effect corresponds to the effect of treatment on the treated, and LATE to the simple before-after estimator, $\beta_{LATE} = (\overline{y}_{t_i} - \overline{y}_{t_0})/(1-0)$.

As we mentioned in Section IV.1, our framework abstracts from time. A more sophisticated version of the before-after estimator would however be the *difference in differences* estimator. This compares the group of individuals with S_{Ii} =1 to the group with S_{Ii} =0 before and after the treatment S_{Ii} occurs, and thus controls for common macroeconomic shocks in addition to unobserved time-invariant individual effects. These two estimators are however confined to the evaluation of treatments where there is an outcome observation both before and after the treatment occurs (e.g. earnings before and after a training spell); it is by contrast less applicable in the schooling/formal education evaluation problem where a "before" observation of earnings is very unlikely to be available, since for most individuals formal education is completed before labour market participation.

IV.2.3.4 Some drawbacks to IV

The first requirement to perform IV estimation is of course the availability of a suitable and credible instrument. Although often quite ingenious instruments have been put forward (from selected parental background variables, to birth order, to smoking behaviour when young, to distance to college, etc.), they have all been subject to some criticism, since it is indeed very hard to fully justify the untestable exclusion restriction they must satisfy. Alternatively, natural experiments have often been used. For example researchers have compared the outcomes among two groups that have a similar distribution of abilities but who, from some exogenous reform, experience different schooling outcomes (for example, see the papers by Angrist and Krueger, 1991 and 1992, Butcher and Case, 1994, Harmon

and Walker, 1995, Meghir and Palme, 2000). A classic example of this is the comparison of adjacent cohorts one of which experiences a school reform (say a change in the minimum school leaving age) and the other who does not. As we have seen, in the homogenous treatment effects model this can be used to estimate the average treatment effect, but in the heterogeneous model where individuals act on their heterogeneous returns, it will estimate the average of returns among those induced to take more schooling by the reform - the Local Average Treatment Effect. The LATE discussion highlights the point that the IV estimate will typically vary depending on which instrument is used. Moreover, it could vary widely according to the local average it recovers. In the example above, the IV estimator will estimate the average returns among those induced to achieve $S_{1i}=1$ by the school leaving age reform. These could be a group with very high (or very low) returns. If those who now achieve S_{1i} were those who had little to gain, then the local average could be low. If on the other hand, they are individuals who had previously left education earlier because of a lack of information or family resources, the local average return for them could be quite high.

In any case the lesson to be learned from the discussion of IV in the heterogeneous returns model is that the nature of the incidence of the instrument within the distribution of returns b_i is critical in understanding the estimated coefficient.

IV.2.4 Control function methods

If individuals make optimal educational choices on the basis of their unobserved characteristics, for the observed sub-sample of high-education individuals the error in the earnings equation will have a nonzero expectation (cf. (11) and (12)). In particular, if individuals who select into schooling have higher average unobserved ability and/or if individuals with higher unobserved idiosyncratic returns from schooling invest more in education, the residual in the earnings equation of high-education individuals will have a positive mean.

Since this bias arises because individual optimisation truncates the underlying disturbances, the idea at the basis of the control function approach is to control directly for the part of the error term in the outcome equation that is correlated

with the schooling regressor, using an explicit model of the schooling selection process. More precisely, the method exploits an additional equation determining which educational choice is made to augment the earnings regression with selectivity variables providing an estimate of the conditional mean of the unobservables in the outcome equation.

IV.2.4.1 The heterogeneous single treatment model

Suppose that in the heterogeneous single treatment model (6),

$$y_i = m_0(X_i) + (b(X_i) + b_0)S_{1i} + (b_i - b_0)S_{1i} + \alpha_i + \varepsilon_i$$

 S_{1i} is determined according to the binary response model:

CF:A1
$$S_{1i} = 1(m_s(Z_i, X_i) + v_i \ge 0)$$
 where $v_i \sim N(0, 1)$

Now assume that the unobserved productivity or ability term α_i and the unobserved individual residual return b_i relate to S_{1i} according to

CF:A2
$$\alpha_i - \alpha_0 = \rho_{\alpha v} v_i + \xi_{\alpha i}$$
 with $v_i \perp \xi_{\alpha i}$
CF:A3 $b_i - b_0 = \rho_{bv} v_i + \xi_{bi}$ with $v_i \perp \xi_{bi}$

where again b_0 is the mean of b_i in the population.

Note that given CF:A1-A3

$$E[(\alpha_{i} - \alpha_{0}) | Z_{i}, X_{i}, S_{1i} = 1] = \rho_{\alpha\nu} \lambda_{1i}(X_{i}, Z_{i})$$

$$E[(\alpha_{i} - \alpha_{0}) | Z_{i}, X_{i}, S_{1i} = 0] = \rho_{\alpha\nu} \lambda_{0i}(X_{i}, Z_{i})$$

$$E[(b_{i} - b_{0}) | Z_{i}, X_{i}, S_{1i} = 1] = \rho_{b\nu} \lambda_{1i}(X_{i}, Z_{i}) \quad (18)$$

where

$$\lambda_{0i} \equiv -\frac{\phi(m_s(Z_i, X_i))}{1 - \Phi(m_s(Z_i, X_i))} \quad \text{and} \quad \lambda_{1i} \equiv \frac{\phi(m_s(Z_i, X_i))}{\Phi(m_s(Z_i, X_i))}$$

are the standard inverse Mills ratios from the normal selection model (Heckman, 1979) – or control functions. (Note that we have implicitly assumed that the ρ 's do not depend on the observables *X*.)

The idea is that once these terms are inserted in the outcome equation (6) and implicitly subtracted from its error term $(b_i - b_0)S_{1i} + \alpha_i + \varepsilon_i$, the purged disturbance will be orthogonal to all of the regressors in the new equation (cf. Heckman and Robb, 1985).

Formally, under CF:A1-A3 the outcome model can be written as

$$y_{i} = \alpha_{0} + m_{0}(X_{i}) + (b(X_{i}) + b_{0})S_{1i} + \rho_{\alpha\nu}(1 - S_{1i})\lambda_{0i} + (\rho_{\alpha\nu} + \rho_{\beta\nu})S_{1i}\lambda_{1i} + \omega_{i}$$
(19)
with $E[\omega_{i} \mid X_{i}, S_{1i}, (1 - S_{1i})\lambda_{0i}, S_{1i}\lambda_{1i}] = 0.$

Consequently, least squares estimation of the augmented log earnings regression which includes the additional terms $(1-S_{1i})\cdot\lambda_{0i}$ and $S_{1i}\cdot\lambda_{1i}$ will produce a consistent estimator of the conditional average treatment effect $b(X_i)+b_0$ and thus of $\beta_{ATE} = b_0 + E[b(X_i)]$. These additional control functions terms thus eliminate the bias induced by the endogeneity of schooling. It is interesting to observe that under the structure imposed on the model, the estimated ρ coefficients are informative on the presence and direction of the selection process (ρ_{av} for selection on unobserved 'ability' and ρ_{bv} for selection on unobserved returns). Moreover, not only does the model readily estimate the average treatment effect for a *random* individual *even* when individuals select into education based on their unobserved individual assumptions made allow us to also recover the other parameters of interest:

$$\beta_{ATT} = b_0 + E[b(X_i) | S_{1i} = 1] + \rho_{\beta \nu} E(\lambda_{1i} | S_{1i} = 1)$$

$$\beta_{ATNT} = b_0 + E[b(X_i) | S_{1i} = 0] + \rho_{\beta \nu} E(\lambda_{0i} | S_{1i} = 0)$$

where $\rho_{\beta\nu}$ is identified from the difference of the coefficients on $S_{1i}\lambda_{1i}$ and on $(1-S_{1i})\lambda_{0i}$.

Finally note that the control function terms depend on the unknown reduced form parameter of the $m_s(\cdot)$ function, which can however be consistently estimated at a first stage Probit step – again analogous to the selection model.²

IV.2.4.2 The homogeneous returns model

In the special case where b_i is constant for all *i* or where individuals do not select on the basis of their unobserved gain (b_i and v_i are uncorrelated, so that $\rho_{bv} = 0$),

² An early example of this can be found in Willis and Rosen (1979).

the control function terms reduce to a single term

$$\rho_{\alpha\nu}[(1-S_{1i})\lambda_{0i}+S_{1i}\lambda_{1i}] \tag{20}$$

IV.2.4.3 The multiple treatment model

The extension to the multiple treatment case is reasonably straightforward. As in (7), write the exhaustive set of *J* treatments (schooling levels) under examination as S_{1i} , S_{2i} , ..., S_{Ji} . Extend the control function assumptions to obtain (where now a bar rather than a zero subscript denotes means to avoid confusion):

$$E[(\alpha_{i} - \overline{\alpha}) | Z_{i}, X_{i}, S_{ji} = 1] = \rho_{\alpha\nu} \lambda_{ji} (X_{i}, Z_{i}) \quad \text{for } j = 0, 1, ..., J$$
$$E[(b_{ji} - \overline{b_{j}}) | Z_{i}, X_{i}, S_{ji} = 0] = \rho_{\beta_{j}\nu} \lambda_{ji} (X_{i}, Z_{i}) \quad \text{for } j = 1, ..., J$$

The heterogeneous returns model (19) is then extended to

$$y_{i} = \overline{\alpha} + m_{0}(X_{i}) + \sum_{j=1}^{J} (b_{j}(X_{i}) + \overline{b_{j}}) S_{ji} + \sum_{j=0}^{J} \rho_{j} S_{ji} \lambda_{ji} + \omega_{i}$$
(21)

with $S_{0i} = 1 - \sum_{j=1}^{J} S_{ji}$, $\rho_j = \rho_{\alpha\nu} + \rho_{\beta_j\nu}$ for all j (with $\rho_{\beta_0\nu} = 0$) and $E[\omega_i \mid X_i, S_{1i}, ..., S_{Ji}, S_{1i}\lambda_{1i}, ..., S_{Ji}\lambda_{Ji}] = 0$

However, note that to avoid multicollinearity problems the λ_{ji} terms will need to have independent variation, suggesting that at least *J*-1 excluded instruments will be required for identification. Typically finding such a large set of "good" instruments is difficult. An alternative identification strategy is to link the λ_{ji} terms together. For example, if the schooling outcomes follow an ordered sequence then it may be that a single ordered probit model could be used for all λ_{ji} terms. In this case,

CF:A1'
$$S_{ji} = 1(\mu_{j-1} < m_s(Z_i, X_i) + v_i \le \mu_j)$$
 where $v_i \sim N(0, 1)$ and

$$\lambda_{ji} = \frac{\phi(\mu_{j-1} - m_s(Z_i, X_i)) - \phi(\mu_j - m_s(Z_i, X_i))}{\Phi(\mu_j - m_s(Z_i, X_i)) - \Phi(\mu_{j-1} - m_s(Z_i, X_i))}$$

All the parameters of interest can then easily be obtained; the generic average return to schooling level j compared to schooling level 0 (the return to which is normalised to zero) for those individuals with highest achieved schooling qualification k is:

$$E(\beta_{ji} | X_i, S_{ki} = 1) = ATE_{j0}(X_i) + \rho_{\beta_j v} E(\lambda_{ki} | S_{ki} = 1)$$

= $\{b_j(X_i) + \overline{b_j}\} + \rho_{\beta_j v} E(\lambda_{ki} | S_{ki} = 1)$

IV.2.4.4 The heterogeneous one-factor model

Consider the one factor or "years of schooling" model (8). As mentioned above in the discussion of IV for this case, one plausible specification of assumption IV:A4 would be

$$S_i = f(Z_i, X_i; \pi) + v_i$$
 where $E(v_i | Z_i, X_i) = 0$

Now given the control function assumptions CF:A2 and A3 we may write

$$y_i = \alpha_0 + m_0(X_i) + (b(X_i) + b_0)S_i + \rho_{\alpha\nu}v_i + \rho_{\beta\nu}S_iv_i + \omega_i$$

where $E[\omega_i | X_i, S_i, v_i, S_i v_i] = 0$ and now the inclusion of the control functions v_i and $S_i v_i$, renders least squares consistent (see e.g. Garen, 1984).

Again note that this is very specific to the continuous schooling measure S_i in the one factor model and the additively separable model for S_i . Finally, in the pure homogeneous one factor model (i.e. with no X-heterogeneous returns either), the control function approach reduces to a regression of log earnings y_i on schooling S_i including the reduced form error v_i as an additional regressor, and is exactly equivalent to the IV approach:

$$y_i = \alpha_0 + m_0(X_i) + (b + b_0)S_i + \rho_{\alpha\nu}v_i + \omega_i$$
 where $E[\omega_i | X_i, S_i, v_i] = 0$.

IV.2.4.5 Some drawbacks to control function

The control function approach allows for heterogeneity in a multiple treatment model but at the cost of being able to construct a set of control function – one for each treatment – that have independent variation. This places strong demands on instrument availability, in that an excluded instrument is required for each treatment.³

³ Strictly speaking, the model would be identified even if X=Z, though identification would then be based purely on the postulated functional form (e.g. the Mill's ratio are non-linear in the regressors X in the outcome equation).

Moreover, a functional form assumption is typically made on the control function. This is equivalent to making an assumption on the distribution of unobservables.

It is true that the distributional assumptions can be relaxed, following the recent developments in the semiparametric selection model literature, but the requirement on excluded instruments can only be weakened by strengthening the model for the treatment choices. For example, when treatments are sequential, as in the case of educational qualifications, one could exploit their sequential nature to estimate an ordered Probit model from which the control functions for each qualification level can be derived. In this example a single instrument would be sufficient.

IV.2.5 Matching methods

The general matching method is a non-parametric approach to the problem of identifying the treatment impact on outcomes.

As discussed earlier, in the case of a social experiment, random assignment of individuals to treatment ensures that potential outcomes are independent of treatment status, which allows one to compare the treated and the non-treated directly, without having to impose any structure on the problem. To recover the average treatment effect on the treated, the matching method tries to *ex post* mimic an experiment by choosing a comparison group from all the non-treated such that the selected group is as similar as possible to the treatment group in terms of their observable characteristics. Under the matching assumption that all the outcome-relevant differences between any two individuals are captured in their observable attributes, the only remaining difference between the two groups is their treatment status, so that the average outcome of the matched non-treated individuals constitutes the correct sample counterpart for the missing information on the outcomes the treated would have experienced, on average, had they not been treated.

The central issue in the matching method is choosing the appropriate matching variables. We will point out that this is a knife edge decision as there can be too many as well as too few to satisfy the identifying assumption for recovering a consistent estimate of the treatment effect. In some ways this mirrors the issue of choosing an appropriate excluded instrument in the IV and Control Function approaches discussed above. However, it will become clear that instruments do not make appropriate matching variables and vice versa. Instruments should satisfy an exclusion condition in the outcome equation conditional on the treatment whereas matching variables should impact on both the outcome and treatment equations.

IV.2.5.1 General matching methods

To illustrate the matching solution for the average impact of treatment on the treated in a more formal way, consider the completely general specification of the earnings outcomes (2) in the single discrete treatment case (J=1). In line with what anticipated above, among the set of variables X in the earnings equations it is useful to distinguish those affecting *both* outcomes y (directly via $m_0(X)$ or indirectly through returns b(X)) and schooling choices S from those affecting outcomes alone. Denote the former subset of X by \tilde{X} .

The solution to the missing counterfactual advanced by matching is based on a fundamental assumption of conditional independence between non-treatment outcomes and the schooling variable S_{Ii} :

MM:A1 $y_i^0 \perp S_{1i} | \tilde{X}_i$

This assumption of *selection on observables* requires that, conditional on an appropriate set of observed attributes, the distribution of the (counterfactual) outcome y^0 in the treated group is the same as the (observed) distribution of y^0 in the non-treated group. For each treated observation $(y_i : i \in \{S_{Ii}=1\})$ we can look for a non-treated (set of) observation(s) $(y_i : i \in \{S_{Ii}=0\})$ with the same \tilde{X} -realisation. Under the matching assumption that the chosen group of matched controls – i.e. conditional on the \tilde{X} 's used to select them – does not differ from the treatment group by any variable which is systematically linked to the non-participation outcome y^0 , this matched control group constitutes the required counterfactual. Actually, this is a process of re-building an experimental data set: given the right observables, the distribution of the observations of the non-treated represents statis-

tically what the treated group's observations would have been had they not been treated.

As it should be clear, the matching method avoids defining a specific form for the outcome equation, decision process or either unobservable term. Still, translated into the more specialised framework of equation (6), MM:A1 becomes: $(\alpha_i, \varepsilon_i) \perp S_{1i} | \tilde{X}_i$. Note that the individual-specific return to education b_i is allowed to be correlated with the schooling decision S_{1i} , provided in this case $(\alpha_i, \varepsilon_i) \perp b_i | \tilde{X}_i$ also holds. In particular, individuals may decide to acquire schooling on the basis of their (from the analyst unobserved) individual gain from it, as long as this individual gain is not correlated to their non-treatment outcome y_i^0 .

For the matching procedure to have empirical content, it is also required that

MM:A2 $P(S_{1i} = 1 | \tilde{X}_i) < 1$ for $\tilde{X} \in C^*$

which prevents \tilde{X} from being a perfect predictor of treatment status, guaranteeing that all treated individuals have a counterpart on the non-treated population for the set of \tilde{X} values over which we seek to make a comparison. Depending on the sample in use, this can be quite a strong requirement (e.g. when the education level under consideration is directed to a well specified group). If there are regions where the support of \tilde{X} does not overlap for the treated and non-treated groups, matching has in fact to be performed over the common support region C^* ; the estimated treatment effect has then to be redefined as the mean treatment effect for those treated falling within the common support.

Note that to identify the average treatment effect on the treated over C^* , this weaker version in terms of conditional mean independence, implied by MM:A1 and MM:A2, would actually suffice:

MM:A1' $E(y^0 | \tilde{X}, S_1 = 1) = E(y^0 | \tilde{X}, S_1 = 0)$ for $\tilde{X} \in C^*$

Based on these conditions, a subset of comparable observations is formed from the original sample, and with those a consistent estimator for the treatment impact on the treated (within the common support C^*) is the empirical counterpart of:

$$E(y^1 - y^0 | S_1 = 1, C^*)$$

$$= \frac{\int_{C^*} [E(y^1 \mid \tilde{X}, S_1 = 1) - E(y^0 \mid \tilde{X}, S_1 = 1)] dF(\tilde{X} \mid S_1 = 1)}{\int_{C^*} dF(\tilde{X} \mid S_1 = 1)}$$

=
$$\frac{\int_{C^*} [E(y^1 \mid \tilde{X}, S_1 = 1) - E(y^0 \mid \tilde{X}, S_1 = 0)] dF(\tilde{X} \mid S_1 = 1)}{\int_{C^*} dF(\tilde{X} \mid S_1 = 1)}$$

=
$$\frac{\int_{C^*} [E(y \mid \tilde{X}, S_1 = 1) - E(y \mid \tilde{X}, S_1 = 0)] dF(\tilde{X} \mid S_1 = 1)}{\int_{C^*} dF(\tilde{X} \mid S_1 = 1)}$$

If the second assumption is fulfilled and the two populations are large enough, the common support is the entire support of both. Note that this estimator is, simply, the mean conditional difference in earnings on the common support, appropriately weighted by the distribution of \tilde{X} in the treated group.

The preceding discussion has referred to the estimation of the average treatment effect on the treated. If we are also interested in using matching to recover an estimate of the treatment on the non-treated, as we do in our application to the NCDS data, a symmetric procedure applies, where MM:A2 needs to be extended to $0 < P(S_{1i} = 1 | \tilde{X}_i)$ for $\tilde{X} \in C^*$ and MM:A1 to include y^1 . In terms of the framework of equation (6), the strengthened MM:A1 thus becomes $(\alpha_i, \varepsilon_i, b_i) \perp S_{1i} | \tilde{X}_i$, highlighting how now possibly heterogeneous returns b_i are prevented from affecting educational choices by observably identical agents. Under these strengthened assumptions, the average treatment effect $E[y^1-y^0]$ can then be simply calculated as a weighted average of the effect on the treated and the effect on the non-treated.

As to the potential sources of bias highlighted by the decomposition in (9), matching corrects for the first two, B_1 and B_2 , through the process of choosing and re-weighting observations within the common support. In fact, in the general nonparametric matching method a quite general form of $m_0(X)$ and of interactions $b(X)S_{1i}$ is allowed (note the use of X rather than \tilde{X} – matching would balance also the variables affecting outcomes alone, since by construction they would not differ between treatment groups), avoiding the potential mis-specification bias highlighted for OLS. Arguing the importance of the remaining source of bias – the one due to unobservables – amounts to arguing the inadequacy of the conditional independence assumption (MM:A1) in the specific problem at hand, which should be done in relation to the richness of the available observables (i.e. the data \tilde{X}) in connection to the selection/outcome processes.

IV.2.5.2 Propensity score matching

It is clear that when a wide range of variables \tilde{X} is in use, matching can be very difficult to implement due to the high dimensionality of the problem. A more feasible alternative based on the results of Rosenbaum and Rubin (1983) is to match on a *balancing score*, that is a function of the observables \tilde{X} , $e(\tilde{X})$, with the property: $\tilde{X} \perp S_1 | e(\tilde{X})$. This is usually carried out on the *propensity score*, the propensity to receive treatment given the set of observed characteristics jointly affecting treatment status and outcomes: $p(\tilde{X}_i) \equiv P(S_{1i} = 1 | \tilde{X}_i)$. By definition, treatment and non-treatment observations with the same value of the propensity score have the same distribution of the full vector of regressors \tilde{X} . Rosenbaum and Rubin have further shown that under MM:A1 and MM:A2, that is when

$$(y^{1}, y^{0}) \perp S_{1} \mid \tilde{X} \text{ and } 0 < p(\tilde{X}) < 1$$

then

$$(y^1, y^0) \perp S_1 \mid p(\tilde{X})$$

In other words, the conditional independence remains valid if $p(\tilde{X})$ – a scalar variable on the unit interval – is used for matching rather than the complete vector of \tilde{X} . Under the two matching assumptions, a matched sample at each propensity score $p(\tilde{X})$ is thus equivalent to a random sample: conditioning on the propensity score, each individual has the same probability of assignment to treatment, as in a randomised experiment, so that individuals with the same value of $p(\tilde{X})$ but a different treatment status can act as controls for each other. At any value of $p(\tilde{X})$, the difference between the treatment and the non-treatment averages is thus an unbiased estimate of the average treatment effect at that value of $p(\tilde{X})$, and the estimate of matching can be thought of as a weighted average of the estimates from a series of mini random experiments at the different values of $p(\tilde{X})$.

It should be noted that in empirical applications the propensity score first needs to be estimated. Since a fully non-parametric estimation of the propensity score would be liable to suffer from the same curse of dimensionality the Rosenbaum and Rubin theorem allows one to circumvent, the estimation task is generally accomplished parametrically (e.g. via a logit or probit specification). Propensity score matching thus becomes a *semi*-parametric approach to the evaluation problem. The estimated propensity score is used in a first step only to (parametrically) correct for the selection bias (on observables) by selecting that subset of the nontreated group to act as control. In fact, all that is required is its ability to balance the relevant \tilde{X} 's in the two groups ($\tilde{X} \perp S_1 \mid \hat{p}(\tilde{X})$), and not to consistently estimate the selection process. Note to this regard that the extent to which the estimated propensity score effectively achieves the balancing in any given finitesample application can indeed be checked (and possibly improved by experimenting with different types of matching) after matching is performed.⁴

The second step, the estimation of the treatment effect, can then be accomplished in a fully non-parametric way, in particular without imposing any functional form restriction on how the treatment effect can vary according to the \tilde{X} 's. The curse of dimensionality is thus sidestepped by parametrically estimating the propensity score only, while the specification of the treatment effect $E(y^1 - y^0 | S_1 = 1, X)$ is left completely unrestricted.

A final remark concerns the estimation of the standard errors of the treatment effects, which should ideally adjust for the additional sources of variability introduced by the estimation of the propensity score as well as by the matching process itself; for kernel-based matching, analytical results have been derived by Heckman, Ichimura and Todd (1998), while for one-to-one matching the common solution is to resort to bootstrapped confidence intervals.

⁴ The dimensionality reduction when using $p(\tilde{X})$ rather than \tilde{X} is achieved because the number of p-values – although in principle infinite on the [0,1] interval – is in practice considerably smaller than the number of \tilde{X} values. For the balancing property of the propensity score to hold, however, enough variation within p-cell is needed. Balancing checks are typically performed by deciles of $\hat{p}(\tilde{X})$.

IV.2.5.3 Implementing propensity score matching estimators

The main idea of matching is to pair to each treated individual *i* some group of 'comparable' non-treated individuals and to then associate to the outcome y_i of treated *i*, a matched outcome \hat{y}_i given by the (weighted) outcomes of his 'neighbours' in the comparison group.

The general form of the matching estimator for the average effect of treatment on the treated (within the common support) is then given by

$$\widehat{\boldsymbol{\beta}}_{MM} = \sum_{i \in \{S_{1i} = 1 \cap C^*\}} \{y_i - \widehat{y}_i\} w_i$$

with w_i typically set equal to $1/N_1^*$ (N_1^* being the number of treated individuals falling within the common support C^*).

The general form for the outcome to be paired to treated *i*'s outcome is

$$\hat{y}_i = \sum_{j \in C^0(p_i)} W_{ij} \cdot y_j \tag{22}$$

where

- $C^0(p_i)$ defines treated *i*'s neighbours in the comparison group (where proximity is in terms of their propensity score to *i*'s propensity score, p_i) and
- $W_{ij} \in [0,1]$ with $\sum_{j \in C^0(p_i)} W_{ij} = 1$ is the weight placed on observation *j* in forming a comparison with tracted observation *i*

forming a comparison with treated observation *i*.

The different matching estimators differ in how they construct the matched outcome \hat{y} , that is in how they define the neighbourhood for the control group for each treated observation. They also differ in how they choose the weights for the control group.

The traditional and most intuitive form of matching is *nearest-neighbour matching*, which associates to the outcome of treated unit *i* a 'matched' outcome given by the outcome of the most observably similar control unit k_i . This amounts to defining $C^0(p_i)$ as a singleton:

$$C^{0}(p_{i}) = \left\{k_{i} \in \{S_{1} = 0\} : |p_{i} - p_{k_{i}}| = \min_{j \in \{S_{1} = 0\}}\{|p_{i} - p_{j}|\}\right\}$$

and setting $W_{ij} = 1(j = k_i)$ (i.e. giving a unity weight to the closest control obser-

vation and zero to any other).

A variant of nearest-neighbour matching is *caliper matching* (see Cochran and Rubin, 1973, and for a recent application, Dehejia and Wahba, 1999). The 'caliper' is used to exclude observations for which there is no close match, thus allowing to better enforce common support on the propensity score. This involves matching treated individual i with its nearest-neighbour non-treated individual j provided that:

$$\delta > |p_i - p_j| = \min_{k \in \{S_i = 0\}} \{|p_i - p_k|\}$$

If none of the non-treated individuals are within a certain predefined absolute distance or caliper δ of the treated individual *i* under consideration, individual *i* is left unmatched.

A different class of matching estimators has been recently proposed by Heckman, Ichimura and Todd (1997 and 1998) and Heckman, Ichimura, Smith and Todd (1998). In *kernel-based matching*, the outcome y_i of treated individual *i* is matched to a weighted average of the outcomes of more (possibly all) nontreated individuals, where the weight given to non-treated individual *j* is in proportion to the closeness of the propensity scores of *i* and *j*. That is, the weight in equation (22) above is set to:

$$W_{ij} = \frac{K\left(\frac{p_i - p_j}{h}\right)}{\sum_{j \in C^0(p_i)} K\left(\frac{p_i - p_j}{h}\right)}$$

With e.g. the Gaussian kernel, $K(u) \propto \exp(-u^2/2)$ and all the non-treated units are used to smooth at p_i , that is $C^0(p_i) = \{j : S_{1j} = 0\}$. By contrast, with the Epanechnikov kernel, $K(u) \propto (1-u^2) \cdot 1(|u|<1)$ and thus only those non-treated units whose propensity score falls within a fixed 'caliper' of *h* from p_i are used to smooth at p_i , that is $C^0(p_i) = \{j \in \{S_{1j} = 0\} : |p_i - p_j| < h\}$.

Before concluding this overview of implementation of matching estimators, we briefly consider how the various types actually implement the common support requirement. Simple nearest neighbour matching does not impose any a priori common support restriction. In fact, the nearest neighbour could at times turn out to be quite apart. By contrast, its caliper variant, provided not too 'tolerant' (as perceived by the researcher), automatically uses the observations within the common support of the propensity score. As to kernel-based matching estimators, two factors automatically affect the imposition of common support: the choice of bandwidth (a small bandwidth amounts to being very strict in terms of the distance between a control and the treated under consideration, de facto using, i.e. placing weight on, only those controls in a close neighbourhood of the treated individual's propensity score) and, to a lesser extent, the choice of kernel (e.g. to smooth for a given treated, the Gaussian kernel uses all the non-treated units, while the Epanechnikov only those non-treated units who fall within a fixed caliper or radius h). Typically with kernel-based matching the common support is additionally imposed on treated individuals at the boundaries, that is those treated whose propensity score is larger than the largest propensity score in the nontreated pool are left unmatched. A more refined procedure is the one suggested by Heckman, Ichimura and Todd (1997), who 'trim' the common support region of those treated falling where the density of the comparison group – albeit strictly positive – is still considered too thin to produce reliable estimates.

IV.2.5.4 The multiple treatment model

Rosenbaum and Rubin's (1983) potential outcome approach for the case of a single treatment has recently been generalised to the case where a whole range of treatments is available by Imbens (2000) and Lechner (2001a).

In our educational context, we consider four treatments (0 = basic education, 1 = O-levels, 2 = A-levels and 3 = higher education) and are thus interested in the causal average effect of treatment *m* relative to another treatment *l*, $E(y^m - y^l | D = j)$ (with $m, l \in \{0,1,2,3\}$ and m > l) for j=m (effect on the treated) and j=l (effect on the non-treated). Each of these pair-wise comparisons represents the average incremental effect of a higher educational level compared to stopping to a lower one, for the group being considered.

With assumptions MM:A1 and MM:A2 appropriately extended, all the required effects are identified. As with the single-treatment case, a one-dimensional (generalised) propensity score can be derived, which ensures the balancing of the observables in the two groups being compared at a time. The details of this extension and of how the propensity score has been derived to take account of the sequential nature of the treatments under examination are presented in the Appendix to this chapter.

IV.2.5.5 Some drawbacks to matching

The most obvious criticism that may be directed to the matching approach is the fact that the identifying conditional independence assumption (MM:A1) on which the method relies is in general a very strong one. Note that despite the fact that compared to OLS, matching is implemented in a more flexible way (in particular not imposing linearity or a homogeneous additive treatment effect), both matching and OLS estimators depend on this same crucial assumption of selection on observables. As mentioned above, its plausibility should always be discussed on a case-by-case basis, with account being taken of the informational richness of the available dataset (\tilde{X}) in relation to the institutional set-up where selection into the treatment takes place.

Furthermore, the common support requirement implicit in MM:A2 may at times prove quite restrictive. In the case of social experiments, randomisation generates a comparison group for each \tilde{X} in the population of the treated, so that the average effect of the treatment can be estimated over the entire support. By contrast, under the conditional independence assumption matching generates a comparison group, but only for those \tilde{X} values that satisfy MM:A2. In some cases, matching may not succeed in finding a non-treated observation with similar propensity score for all of the participants. If MM:A2 fails for some subgroup(s) of the participants, the estimated treatment effect has then to be redefined as the mean treatment effect for those treated falling within the common support.

If the impact of treatment is homogeneous, at least within the treated group, no additional problem arises besides the loss of information. Note, however, that the setting is general enough to include the heterogeneous case. If the impact of participation differs across treated individuals, restricting to the common subset may actually change the parameter being estimated; in other words, it is possible that the estimated impact does not represent the mean treatment effect on the treated. This is certainly a drawback of matching in respect to randomised experiments; when compared to standard parametric methods, though, it can be viewed as the price to pay for not resorting to the specification of a functional relationship allowing to extrapolate outside the common support. In fact, the absence of good overlap may in general cast doubts as to the robustness of traditional methods relying on functional form. Lechner (2001b) derives nonparametric bounds for the treatment effect to check the robustness of the results to the problem of a lack of common support.

IV.2.6 OLS, Matching, Instrumental Variables and Control Function

This final sub-section briefly summarises the relationship between the estimators we have considered.

First of all, in the following assume away the issue of mis-specification, in particular assume that

- 1) the OLS, IV and the control function estimators are properly specified, allowing for flexible $m_0(X_i)$ and $b(X_i)$ term
- 2) the issue of common support can be ignored (either by assuming that there is sufficient overlap in the distribution of X in the treated and non-treated subsamples, or that all estimators condition on observations falling within the common support).

These assumptions are made to avoid the two sources of bias due to observables B^{1} and B^{2} ; it would not be possible to make clear statements as to the direction of this bias, which as highlighted in 2.2 would depend on: the sign of the differential treatment impact for those regressors whose interaction terms with the treatment indicator are omitted, the sign of the effect of omitted interactions among the regressors or of their higher order terms on the outcome, the distribution (in particular the mean) of these regressors in the population for whom we intend to measure the treatment effect, the distinct characteristics of those individuals falling within the common support, etc. Under the no-misspecification assumptions above, note immediately that OLS and matching coincide.

Secondly, once we have thus brought all estimators on an equal footing, matching (= OLS), IV and control function would produce the same estimates in the absence of selection on unobservables.

We consequently now focus on a situation characterised by bias due to unobservables (B^3) .

To focus on the relative performance of matching compared to the regressionbased estimators when the basic conditions for the applicability of the latter are met, let us further assume that the exclusion restriction $E[\alpha_i | X_i, Z_i]=0$ for the instrument used by IV as well as the decomposition required by the control function estimator (including postulated structure between the error terms and exclusion restriction) are verified.

In the presence of ability bias, arising from the correlation between α_i and S_{1i} , both the IV and control function estimators should correctly recover the average effect of treatment on the treated (IV directly, the control function exploiting the assumed structure). The effect of treatment on the treated recovered by matching should however be upward biased (assuming more able individuals are more likely to choose $S_{1i}=1$); the effect of treatment on the non-treated would be similarly upward biased, and thus so would the average treatment effect.

When selection into schooling is driven by individuals' idiosyncratic gain from it, b_i , the control function estimator would directly recover the correct average treatment effect, while IV would pick out an instrument-related margin (LATE), which could be much higher or much lower than the average effect for a random individual in the population. Provided the individual-specific gain is unrelated to ability (α_i)⁵, both the matching and control function estimators⁶ could recover the right average effect on the treated. However, in contrast to the control function estimate, the effect of treatment on the non-treated – and thus the average treatment effect – obtained with matching would be upward biased (assuming that

⁵ If on the other hand the individual gain were related to the no-treatment unobservable, and individuals were to select into treatment on the basis of this gain, then selection (ability) bias would arise.

⁶ Again, the control function would need to exploit the imposed structure to derive β_{ATT} .

those with the higher gains select into education).

IV.3 Appendix: A balancing score for sequential multiple treatments

In the educational context, we can view the sequential treatments (basic education, O-levels, A-levels, higher education) in a dose-response framework (cf. Imbens, 2000). Like a drug which can be applied in different doses, the sequential treatments would thus correspond to ordered levels of a treatment – education (or investment in human capital). We focus on continuous education, where individuals take uninterrupted sequential decisions of an incremental nature: at each point, they can either stop or move on to the next educational level.

We consider four treatments: $D \in \{0, 1, 2, 3\}$

- *D*=0 for stopping at basic (i.e. no qualifications)
- *D*=1 for stopping at O-Levels (i.e. stay on and stop with O-levels)
- *D*=2 for stopping at A-Levels (i.e. stay on, take O-levels and stop with A-levels)
- *D*=3 for stopping at Higher Education (i.e. stay on, take O-levels, take A-levels and stop with HE)

Incidentally, we can link this analysis to the dynamic programme evaluation framework recently suggested by Lechner and Miquel (2001). In our case of an obliged chain of educational choices, we only have a restricted set of possible sequences, four in fact. In addition, in each period there is only one type (and in fact a different type) of treatment available.

Consider four periods: in period 0 everyone achieves basic qualifications, in period 1 the relevant choice is whether to take O-levels or not; in period 2 the only treatment available is A-levels but only provided one has achieved O-levels in the previous period; while the treatment in period 3 is higher education but available only for those with A-levels. Outcome y is then observed after period 3 (at age 33). The four possible sequences, corresponding to the four values of D defined above, are thus:

t	0	1	2	3
D=0	1	0	0	0
D=1	1	1	0	0
D=2	1	1	1	0
D=3	1	1	1	1

Define the sequence of treatments, each received at the beginning of each period *t*, as $S = (S_0, S_1, S_2, S_3)$, in our case, $S_0 = 1$ for all; for t = 1, 2, 3, $S_t \in \{0, 1\}$ and $S_t = 1 \Longrightarrow S_{t-1} = 1$.

The sequential nature of the decision process is captured by the modelling of the choice probabilities as follows:

Define:

• $r(j, x) \equiv P(D=j | X=x)$ for j = 0, 1, 2, 3.

Further define the following probabilities:

- $P^{1}(x) \equiv P((OL \lor AL \lor HE) = 1 | X=x)$ the stay-on probability
- $P^2(x) \equiv P((AL \lor HE) = 1 | OL=1, X=x)$
- $P^{3}(x) \equiv P(HE=1|AL=1, X=x)$

Thus we have:

$$D = \begin{cases} 0 & r(0, x) = 1 - P^{1}(x) \\ 1 & r(1, x) = (1 - P^{2}(x))P^{1}(x) \\ 2 & r(2, x) = (1 - P^{3}(x))P^{2}(x)P^{1}(x) \\ 3 & r(3, x) = P^{3}(x)P^{2}(x)P^{1}(x) \end{cases}$$

Let y^k denote the outcome if the individual were to receive treatment (or education level) *k*.

We are interested in the following 12 pairwise comparisons of the effects of treatments (education levels) *m* and *l* with *m*, $l \in \{0,1,2,3\}$ and *m>l*:

$$\mathrm{E}(y^m - y^l / D = j)$$

for j = m (effect on the treated) and j = l (effect on the non-treated).

Note that in the framework by Lechner and Miquel (2001), evaluating the impact of stopping at one education level versus stopping at another one for those individuals who have stopped at one of such levels amounts to evaluating the effect of one sequence of treatments compared to another sequence of treatments (each of length four) for those individuals who have followed a given sequence.

In our case, we need to identify all the counterfactuals of the type:

$$E(y^m | D = l)$$
 and $E(y^l | D = m)$

(More precisely, we need to identify $E(y^k | D = j)$ for

(k, j) = (0,1), (0,2), (0,3), (1,2), (1,3), (2,3) (effect on the treated)

as well as for

(k, j) = (1, 0), (2, 0), (3, 0), (2, 1), (3, 1), (3, 2) (effect on the non-treated).)

An extension of the conditional independence assumption MM: A1 that would allow to identify them is what Imbens (2000) termed 'strong unconfoundedness':

$$\{y^0, y^1, y^2, y^3\} \perp D \mid X_0 = x_0$$

In words, conditional on the information observed prior to period 0, X_0 , assignment to treatment in each period is independent of potential outcomes, in particular it is not affected by any new information related to the outcomes that may arrive in between schooling choices.

This implies that the complete treatment sequence, in our case the maximum level of education attained, is chosen at the beginning of period 0 – just as the dose of a drug is decided at the start – based on the information contained in X_0 .

The assumption that subsequent schooling choices are not affected by the outcomes of the schooling decisions in the previous periods hinges on the absence of intermediate outcomes on which to possibly base future *S* decisions. This amounts to ruling out 'intermittent' educational choices – where an individual achieves a level of education, drops out of the education system, observes the corresponding outcomes (both in terms of *y* and of possibly endogenous *X*'s) and then possibly decides on re-entering the schooling system for investment in the next level of education.

Note however that the weaker form (implied by strong unconfoundedness) would suffice to our purposes:

$$\{y^{l}, y^{m}\} \perp D \mid X_{0} = x_{0}, D \in \{l, m\}$$

for $(l, m) = (0, 1), (0, 2), (0, 3), (1, 2), (1, 3), (2, 3)$

This relaxes the CIA above by requiring conditional independence to hold only for the subpopulations receiving treatment m or treatment l (see Lechner, 2001a).
The common support assumption corresponding to MM: A2 is:

$$0 < r(j, x) \equiv P(D=j | X_0=x) < 1$$
 for $x \in C^*$ and $j=0,1,2,3$

which for the P^{j} 's implies the requirements:

$$0 < P'(x) < 1$$
 for $x \in C^*$ and $j=1,2,3$.

IV.3.1 Looking for a balancing score

If we wanted the X's (in what follows we drop the time-0 subscript from X) to be simultaneously balanced *in the four groups* defined by the highest level of education attained, i.e. if we required the same distribution in four selected (matched) subgroups of the four types of treated, we would need to look for a function of the X's, e(X), such that (cf. Theorem 2 by Rosenbaum and Rubin, 1983 and Proposition 1 in Lechner 2001a):

$$X \perp D \mid e(X)$$

$$\Leftrightarrow \mathsf{E}(P(D=m \mid X) \mid e(X)) = P(D=m \mid X) \quad \forall m=0,1,2,3.$$

Setting up the corresponding system:

$$\begin{cases} E(P(D=0|X)|e(X)) = r(0,x) = 1 - P^{1}(x) \\ E(P(D=1|X)|e(X)) = r(1,x) = (1 - P^{2}(x))P^{1}(x) \\ E(P(D=2|X)|e(X)) = r(2,x) = (1 - P^{3}(x))P^{2}(x)P^{1}(x) \\ E(P(D=3|X)|e(X)) = r(3,x) = P^{3}(x)P^{2}(x)P^{1}(x) \end{cases}$$

Choosing either $e(X) = \{r(1, X), r(2, X), r(3, X)\}$ or $e(X) = \{P^{I}(X), P^{2}(X), P^{3}(X)\}$ would solve the system. (Note that the dimensionality has been reduced by one; this is allowed by the adding up of the treatment probabilities).

We are however just interested in the *pairwise* comparison of the various levels of the treatment, so that the above balancing score may actually be more restrictive than required for some type of comparison.

IV.3.2 A balancing score for the pairwise comparisons

For the generic effects of treatment on the treated and treatment on the non-treated $E(y^m - y^l | D = m)$ and $E(y^m - y^l | D = l)$, with $m, l \in \{0, 1, 2, 3\}$ and m > l, we

just need

$$X \perp D \mid e(X), \ D \in \{m, l\}$$

which is verified if

 $E(P(D=m | X, D \in \{m, l\}) | e(X)) = P(D=m | X, D \in \{m, l\})$

Once could use the propensity score:

$$P^{m|m|}(X) \equiv P(D = m \mid X, D \in \{m, l\}) = \frac{r(m, X)}{r(m, X) + r(l, X)}$$

or, alternatively, a finer balancing score e(X) made up of the elements $P^{m|m|}(X)$ is a function of.

Thus an alternative balancing score for:

- 1. $E(y^{1} y^{0} / D = j)$ with j=0,1 is $e(X) = \{P^{1}(X), P^{2}(X)\};$
- 2. $E(y^2 y^0 / D = j)$ with j=0,2 and for $E(y^3 y^0 / D = j)$ with j=0,3 is $e(X) = \{P^1(X), P^2(X), P^3(X)\};$
- 3. $E(y^2 y^1 / D = j)$ with j=1,2 and for $E(y^3 y^1 / D = j)$ with j=1,3 is $e(X) = \{P^2(X), P^3(X)\}$; while
- 4. $E(y^3 y^2 / D = j)$ with j=2,3 is simply $e(X) = P^3(X)$.

When matching on the one-dimensional propensity score, imposing common support can be done in terms of this scalar; when matching on finer balancing scores, it should be imposed on each element.

CHAPTER V

An application to the British NCDS

V.1 Introduction

In this section we apply the different estimation approaches outlined in the previous chapter in the single and in the multiple treatment framework. The relative magnitude of the different estimates is compared and contrasted to see what can be learnt about the selection and outcome models.

As to the single treatment, we follow the traditional choice in the US literature (college *versus* no college) and consider the return from undertaking some form of higher education (HE). We subsequently consider the sequence of multiple treatments (no qualifications, O levels or equivalent, A levels or equivalent, higher education). The outcome of interest are individual wages at age 33. In order to fully focus on the returns to education and avoid selection into employment issues, we restrict our attention to males only.

V.2 Single treatment models: Higher education

The estimated returns to undertaking some form of higher education are shown in Table (5.1). In this model, the 'non-treated' are a heterogeneous group made up of those leaving school with no formal qualifications, those stopping at O-levels and those finishing with A-level qualifications.

We begin by comparing the two methods which rely on the selection on observables assumptions, OLS and matching.

The form of OLS we consider is the linear and common coefficient standard specification, of which non- (or semi)-parametric matching represents a flexible version.

Our comparison of the two methods also includes an assessment of their sensitivity to the richness of the conditioning data. Given their common identifying assumption, the nature of the available observables is crucial for the credibility of the estimates. In particular, we compare estimates based on the detailed information in the NCDS to those obtained from the standard pre-treatment information in commonly available datasets.

Some comments on the choice and interpretation of these X variables may be useful at this stage. As described in Section 1 of the Chapter 4, our control variables X need to be 'attributes' unaffected by the treatment itself. Candidate regressors are thus pre-treatment variables, as well as all time-invariant characteristics of the individuals.¹

In addition, all the variables that are thought to influence *both* the educational decision of interest *and* wage outcomes should ideally be included as regressors. Instrumental variable would thus not make good conditioning regressors.

Finally, note that since our conditioning *X* variables, say X_0 , are measured before (or at the time of) the educational choice, the treatment effects we estimate will include the effect of schooling on some subsequent *X* which would also affect measured outcomes (examples include tenure, experience, or type of occupation found). The treatment effect will thus consist of all channels through which education affects wages, both directly (e.g. through productivity) and indirectly (via some of the *X*'s).²

With these considerations in mind, we now turn to the estimates presented in Table 5.1.

Specification (i) gives the OLS estimate when we only use minimal controls (region and ethnicity). The corresponding matching estimate is shown in row (iv). We see that the estimated return to HE for men is around 40% for both estimators, with the matching point estimate³ basically coinciding with the one from OLS.

¹ Pre-treatment values of the outcome Y (provided not influenced by future treatment take-up through e.g. anticipatory effects), would provide valuable conditioning information. In contrast to the evaluation of training, for educational choices such information is generally not available since very few individuals would have pre-education labour market experience.

 $^{^2}$ So for instance the effect on wages of HE *versus* no-HE is net of the effect on experience of HE *versus* no-HE: individuals with HE have less experience at any given age than otherwise identical individuals who have stopped before. If experience is rewarded on the labour market, the effect on wages of HE visà-vis less-than-HE is net of this.

³ Various types of matching estimators have been tried. Based on indicators of the resulting balancing of X in the two matched sub-samples presented in summary form in the Appendix, the results shown follow from kernel matching, with an Epanechnikov kernel, 0.1 bandwidth and the imposition of common support. The results from the various methods were in any case surprisingly close.

		۸ T T	ATE	
01.6		AII	AIL	AINI
OLS				
(i)	basic specification	39.8	39.8	39.8
		(37.1; 42.5)	(37.1; 42.5)	(37.1; 42.5)
(ii)	full specification	27.1	27.1	27.1
		(23.8; 30.4)	(23.8; 30.4)	(23.8; 30.4)
(iii)	fully interacted	25.6	29.5	31.0
~ /	2	(21.8; 29.6)	(24.9; 33.6)	(25.4; 36.3)
MATCH	IING			
(iv)	basic specification	40.1	40.1	40.2
	1	(37.5; 43.1)	(37.5; 42.8)	(37.5; 42.8)
(v)	full specification	25.4	28.2	29.3
		(21.1; 29.7)	(24.9; 31.6)	(24.2; 33.4)
CONTR	OL FUNCTION			
Homog	geneous returns			
(vi)	full specification	28.1	28.1	28.1
(11)	full speenleuton	(15.0; 41.3)	(15.0; 41.3)	(15.0; 41.3)
(vii)	fully interacted	26.0	29.9	31.5
		(5.3; 63.2)	(7.0; 70.4)	(7.2; 73.7)
Hetero	ogeneous returns			
(viii)	full specification	43.6	36.5	33.7
	-	(18.8; 83.7)	(12.7; 66.6)	(11.1; 61.7)
(ix)	fully interacted	26.8	28.9	29.8
	5	(1.4; 55.9)	(9.6; 59.4)	(2.3; 67.2)
INSTRUMENTAL VARIABLES				
(x)	full specification	-16.7		
	•	(-101; 68)		
(xi)	fully interacted (A)	65.3		
()	,	(-587; 6011)		
(xii)	fully interacted (B)	-36.4		
()		(-160; 32)		

Table 5.1 The returns to higher education compared to less-than higher education (% wage gain): Average treatment effect (ATE), average effect of treatment on the treated (ATT) and average effect of treatment on the non-treated (ATNT)

Notes: Basic specification: ethnicity and region. Full specification: ethnicity, region, standard parental background information, tests at 7 and 16, school variables.

IV specification (xi): method A, i.e. Z_i and $Z_i X_i$ to instrument S_{Ii} and $X_i S_{Ii}$.

IV specification (xii): method B, i.e. \hat{S}_{1i} replaces S_{1i} and $\hat{S}_{1i}X_i$ replaces X_iS_{1i} .

Sample size N = 3,639, except for matching: ATE (3,414), ATT (1,019) and ATNT (2,395).

Bootstrapped 95% bias-corrected percentile confidence intervals for all specifications (500 repetitions) except for (i), (ii), (vi) and (x), for which the confidence intervals are based on the White-corrected standard errors.

When we include a richer set of controls – ability measures both at 7 and 16, school type variables and standard family background variables⁴ (specifications (ii) and (v)) – both these estimates fall to between 25 and 29%. In particular, the OLS coefficient – constrained to be homogeneous – shows a 27.1% average wage gain from taking some form of higher education. Matching estimates are more informative, showing that the higher-educated enjoy a 25.4% average gain from having taken HE (treatment on the treated), while the estimated return for those who stopped (at any stage) before HE would have been 29.3% (treatment on the non-treated). These results suggest that if those who have not continued to HE were to undertake it, they would enjoy a higher benefit than the group who has effectively gone on HE.

We showed in Section 2.2 of Chapter 4 that if there are X-heterogeneous returns to HE, standard OLS regressions would produced biased estimates. To check this issue, in specification (iii) we allow the OLS estimate to model the (observably) heterogeneous returns $b(X_i)$ in a more flexible way by allowing all interactions⁵ between the X's and the treatment indicator S_1 . These interactions $X \cdot S_1$ are in fact significant (*F*=1.54, with p=0.008), and allowing for them brings the OLS estimate of the ATT extremely close to the matching one (cf. (iii) with (v)).

This first set of results usefully highlights several issues.

- At least in our application, the basic pre-education information available in common datasets would not have been enough to identify gains in a reliable way; in our case, generally unobserved ability and family background variables would have led to an upward bias of around 60%.
- 2) By allowing for an (observably) heterogeneous gain from HE, matching estimates provide additional interesting information as to the average gains for the subgroups of treated and non-treated.
- 3) Even though the interactions terms are statistically significant, thus providing evidence of the presence of heterogeneous returns $b(X_i)$, once we average the conditional effect over the distribution of X in the treated sample and in the full sample,

 $^{^4}$ The family background variables that we include are parent's education, age, education×age, father's social class when the child was 16 (six dummies), mother's employment status when the child was 16, and the number of siblings the child had at 16.

⁵ Although we allow only for all linear $X S_{1i}$ interactions, most of the X's (45 out of 52) are dummy variables.

the resulting ATT and ATE estimates turn out to be relatively close. The point estimate of the ATNT does however lie outside of the confidence interval for the ATT.

Both the least squares and matching methods rely on the assumption of selection on observables; however rich our dataset may seem, it is important to check this assumption against the performance of the two estimation methods reviewed in the previous chapter – instrumental variables and control function – that try to control for selection on unobservables.

Both of these estimator try to control for the endogeneity in the education variable by exploiting some 'exogenous' variation in schooling by way of an excluded instrument.

The choice of an appropriate instrument *Z*, like the choice of the appropriate conditioning set *X* for matching or OLS, eventually boils down to untestable judgement. In fact, although there might be widespread consensus in including test score variables as ability measures among the *X*'s or in viewing an exogenous change in some educational rule or qualification level for one group but not another as an appropriate instrument, for several variables one can always make a case pro or against their use as instruments. 'Intermediate' variables of this latter kind are for instance standard family background variables – parental education, income or social class – , although probably most researchers would now regard them as correlated with the unobservables in the earnings equation and thus as appropriate matching variables as opposed to instruments. Some more borderline cases would for instance be information on whether the family was experiencing financial difficulties at the time of the child's educational choice, parental interest in the child's education as assessed by the child's teacher, or the number of older siblings the child has, which, controlling for the total number of children should be exogenous (and equivalent to birth order).

In our application, in order to highlight the interpretation of the results, we focus on one instrument, birth order – a highly significant determinant of the choice to undertake higher education (conditional on the full set of controls X).⁶ The IV assumption IV:A4 of the instrument affecting S_I is thus verified; however as to the identifying exclusion restriction IV:A1 one could of course still argue that in addition to educational attainment, birth order could affect other individual traits (e.g. motivation, parents' invest-

⁶ Its coefficient in the linear probability model is -3%, with a p-value of 0.001.

ment in the child's human capital) that could in turn affect earnings.⁷ To this regard note however that the credibility of an instrument depends on what *X* variables one is already controlling for, as highlighted in IV:A1 $E(\alpha_i | X_i, Z_i) = 0$. In our *X* set we do include all the ability (measured at 7), tests at 16 and standard family background controls, thus requiring the instrument to be unrelated to the unobserved component of potential earnings for given ability, early school performance and parental background.

In the IV and CF specifications in Table 5.1, we thus always include the full controls for region, ethnicity, standard family background variables together with ability and school type variables (as in specifications (iii) and (v)) and use the number of older siblings the child has (controlling for total number of siblings) as instrument.

We start by presenting and discussing the control function estimates, and in particular the ones obtained in the homogenous returns model. This endogenous treatment model produces an estimated gain of 28.1% (ATT being equal to ATE and to ATNT by construction), and, under the assumed distributional assumptions, rejects the presence of remaining selection bias on unobservables.⁸

Furthermore, once we make our estimate fully comparable to the matching ones by allowing for full interactions $X \cdot S_I$ in the earnings equation and $X \cdot Z$ in the participation equation, the estimated ATT decreases to 26%, basically the same return estimated by matching and OLS. And as for matching and OLS, the ATNT parameter at 31.5% is estimated to be slightly higher than the ATT. Note however how the precision of the estimates – already low compared to OLS and matching – drastically decreases as we allow for a fully interacted model.

These results are further confirmed by the results from the control function model which specifically allows for the possibility that individuals select into higher education on the basis of their unobserved heterogeneous returns. This model is simply a more general version of the endogenous treatment effects model, with the inverse mills ratio fully interacted with the HE dummy.

⁷ As pointed out by one referee.

 $^{^{8}\}rho_{av}$ is estimated at -0.017 with a robust standard error of 0.10.

Interestingly, if we omit ability from the regression, the coefficient on the control function is positive and significant

A first interesting result is that the coefficients on both correction terms are insignificant⁹, confirming the homogenous control function result of no selection on (remaining) unobservables and rejecting the hypothesis of individuals selecting into higher education on the basis of their idiosyncratic gains from it. The heterogeneous-returns CF estimate of the return to HE for those taking HE is increased to 43.6% from the homogenous returns model. It would also seem that the point estimate of the return for the nontreated is lower (33.7%) than for the treated, in contradiction to the matching, fully interacted OLS and fully interacted homogenous control function estimates. However, once we allow for full interactions in the heterogeneous control function too, the point estimates fall in line with the previous ones, both in magnitude and in ranking: ATT of 26.8%, a slightly higher ATNT of 29.8% and an intermediate ATE of 28.9%. This finding suggests that the control function estimator accounting for heterogeneous effects is more sensitive to the specification of the interaction terms.

Just as in the homogenous control function case, however, the precision of the estimates is rather poor, and would not allow rejection of the fully interacted 'matching' specification.

Thus again, we find some evidence that we need to allow for full interactions in order to get the point estimates close to matching; but once we do so, there is no large point evidence of heterogeneity in impacts, and the precision of the estimates severely deteriorates.

An alternative check on the matching specification with its underlying conditional independence assumption is to use instrumental variables on a fully interacted model. If we believe the above results of no selection on unobserved individual gains, IV should recover the average effect of treatment on the treated.

We start with the standard IV specification, where we instrument schooling with birth order while controlling for ethnicity, standard parental background information, tests at 7 and 16, school and regional variables. The estimate we obtain (-16.7%) is nowhere near statistical significance.

The previous findings that the return to HE does indeed depend on the *X*'s, and that these *X*'s also impact on the schooling decision (one example in the present application

⁹ $\rho_{\alpha\nu}$ is estimated at -0.107 with a robust standard error of 0.08, while $(\rho_{\alpha\nu} + \rho_{\beta\nu})$ is estimated at -0.024 with robust standard error of 0.07.

is parental education), cast however doubts on this IV result. In fact, our IV estimation would need to control for these endogenous interaction terms as well.

By allowing observable heterogeneity in returns in a fully interacted model with all the $X \cdot S_I$ terms, instrumenting it appropriately with a corresponding set of instruments $X \cdot Z$ (method A in Chapter 4, Section 2.3), and averaging it over the X distribution of the treated, the IV estimate does in fact turn positive (65%), though still clearly extremely imprecise (specification (xi)). This severe lack of precision points to the fact that while this IV estimation requires us to instrument every one of the endogenous $X \cdot S_I$ terms, our corresponding instruments do not have enough power to predict all the interactions well, resulting in a poor performance of our interacted IV model.

We also experimented controlling for the interactions using IV method B, i.e. fully exploiting the conditional mean independence assumption and using the schooling prediction to replace both schooling and its interaction terms in the outcome equation (specification (xii)). However, although this does shrink the confidence interval, there still remains insufficient variation in \hat{S}_{1i} and $X_i \hat{S}_{1i}$ to recover a precise and statistically significant estimate of the average treatment effect on the treated the estimate.

More in general, we can envisage three cases concerning a variable M, say parental education, which (1) affects the outcome y_i directly, conditional on X; (2) affects returns from HE in terms of y; and (3) helps determine HE participation.

- a) <u>M is unobserved.</u> In this case, provided the instrument is uncorrelated with M, IV identifies an instrument-determined local effect, LATE. (Note that it is irrelevant whether individuals select on this heterogeneous return or not; even if they did not, in this situation IV would not identify the ATT parameter.)
 If however the instrument is correlated with M, even LATE would break down.
- b) <u>*M* is observed and we condition on it linearly in both the participation and the outcome equations</u> (as done in specification (x)). Since we do not control for the interaction, IV is inconsistent: $M \cdot S_I$ is omitted from the outcome equation, giving rise to an omitted endogenous variable bias.
- c) <u>*M* is observed, we control for it linearly in the participation equation and interacted</u> with S_{1} in the outcome equation (as done in specification (xi)). Since there is now an additional $M \cdot S_{1}$ endogenous term in the outcome equation, we would need the

additional instrument $M \cdot Z$. The potential problem in this case is that our instruments Z and $M \cdot Z$ may not predict the interactions S_1 and $M \cdot S_1$ very well, resulting in a loss of precision. The alternative of exploiting the schooling prediction both linearly and interacted with M may buy us some efficiency, but would still require the instrument Z to provide sufficient variation in \hat{S}_1 and $M \cdot \hat{S}_1$ so as to allow b_0 and b_M to be independently identified. Either IV method would place strong demands on the instrument, particularly when there are many interaction terms.

Using the instrument we have, we find that, just as in the control function case, we cannot reject the matching specification. However, and unlike control function, we do not get any precision for our IV estimates.

In fact, our control function with interactions model allows us to settle on the intermediate case, where all the $X \cdot S_I$ interactions are included in the outcome equation and the $X \cdot Z$ terms are exploited in the first-step probit, from which however still *only two* predictions (λ_I and λ_0) need to be computed. All this is obviously possible only by making stronger assumptions, in particular the heterogeneity b_i to be additive in the observables and unobservables. Placing a much heavier structure on the problem than does IV, the control function method thus allows it to recover the ATE, ATT and ATNT parameters directly (as opposed to the more local parameters arising from IV) and to do so in a considerably more efficient way – at the obvious price of being much less robust than IV.

To conclude, the following lessons have emerged from the contraposition and discussion of the various estimates.

In the NCDS there seem to be enough variables to be able to control *directly* for selection on unobservables – both unobservable individual traits and unobservable returns. In other words, OLS and matching with the available set of X's do not seem to be subject to selection bias, and individuals do not seem to select into higher education on the basis of returns still unobserved from the econometrician.

Connected to this latter point, we have found limited evidence that interactions matter. More precisely, interactions do matter in terms of obtaining unbiased point estimates (thus allowing to avoid omitted relevant terms bias, especially from incremental returns to HE based on parental education or region). In practice, though, once the effects are averaged by groups, it turned out that the distribution of these characteristics is reasonably balanced among them, resulting in the point estimate of the average effect on a treated individual being quite close to the expected effect on an individual randomly picked from the population. Comparing the effect for the treated and for the non-treated, however, they appear to be statistically different, though by a small margin (based on OLS and matching).

In terms of efficiency, in our application matching performs extremely well, providing point estimates which are always just as precise as the corresponding OLS ones. Control function estimates are by contrast considerably less precise, while those arising from IV fall nowhere near significance. As we discussed above, the argument for IV estimation in our case and with the instrument we have is very weak.

V.3 Multiple treatment models

We now turn to a more disaggregated analysis that focuses on the sequential nature of educational qualifications. To this end we separate the qualifications variable into those who dropped out of school with no qualifications, those who stopped education after completing O levels or equivalent, those who stopped after completing A levels or equivalent, and those who completed O levels, A levels *and* higher education.

Since now we have four treatments, IV estimation would require at least three credible instruments (assuming no X-heterogeneous effects need to be instrumented). As to the control function approach, in the first stage one could exploit the sequential nature of the treatments and estimate an ordered probit model for the various levels of education based on one instrument only. This would however unduly rely on the structure (arbitrarily) imposed of the problem, since the model would be purely identified from the postulated treatment choice model. In fact, the literature has highlighted how poorly the model can perform in the absence of exclusion restrictions. In order to allow for 'non-parametric identification', multiple instruments would be needed.

In our data, we do not however feel to have defensible instruments, and the poor performance of IV in the single-treatment model would discourage further explorations on this front.

Consequently, we assume that the conclusion obtained for the single treatment case (i.e. that in the NCDS there seem to be enough variables to be able to control directly for selection) carries over to the multiple treatment case, so that we concentrate on matching (and linear homogenous OLS as a comparison).

Our approach involves estimating the incremental return to each of the three qualifications by actual qualification. For those with no qualifications, we estimate the returns they would have got if they had undertaken each of the three qualifications (treatment on the non-treated). For those with O-level qualifications we estimate the return they obtained for taking that qualification (treatment on the treated) and the returns they would have obtained if they had progressed to A levels or HE (treatment on the non-treated). For those with A-levels we estimate the returns they obtained for undertaking O- and A-level qualifications (treatment on the treated) and the returns they would have obtained if they had progressed to HE (treatment on the non-treated). For

obtained if they had progressed to HE (treatment on the non-treated). For those with HE all estimates are treatment on the treated.

Our matching estimator uses an adaptation of the propensity score matching method for multiple sequential treatments, which is described in detail in the Appendix to the previous chapter, Section 3. Outcomes across each of the four groups, matched on the appropriate *one* dimensional propensity score for the particular transition in question, are then compared.

The set of results we present was obtained on the basis of the 'best' specification (i.e. the one resulting in the 'best' balancing of the X's in the matched sub-samples¹⁰) and where common support was imposed. In fact, in estimating the effects and in calculating the probabilities for the average treatment effects, common support was imposed also in terms of only including individuals who are matched for every possible transition (so that we can make comparisons across the same sets of individuals).

The multiple treatment results are shown in Table 5.2.

Overall the results have a distinct pattern.

Although to save space we only present results for the full set of controls X (now excluding ability tests at 16), we note that, as in the single treatment case, controlling for ability and school type is important and reduces the return to education at all levels.

Secondly, the results show significant overall returns to educational qualifications at each stage of the educational process, even after correcting for detailed background variables and ability differences, as well as allowing for (observed) heterogeneity in the education response parameters.

As in the single treatment case, OLS and the average treatment effect obtained via matching are extremely close. They show an average wage return of 16-18% from obtaining O-levels compared to leaving schools with no qualifications, a further 7% return from completing A-levels and a further 24-25% wage premium for then achieving higher education. Compared to no qualifications then, the average return to O-levels is roughly 17%, to A-levels is 25%, which doubles to 50% for HE.

¹⁰ Specification search included, *for a given set of X's*, using a balancing score constructed following the steps outlined in the Appendix or estimated from a simple probit, as well as using nearest neighbour or kernel-based matching methods. Simple probits and kernel-based matching often actually turned out to be 'better' at balancing the X's.

	O-level	A-level		HE			Ν
	versus None	versus O-level	versus None	versus O-level	versus A-level	versus None	
None	14.9 (11.1; 18.7)	6.8 (2.7; 11.8)	21.7 (17.3; 26.5)	33.4 (27.6; 40.9)	26.6 (19.6; 34.20)	48.3 (41.8; 55.4)	624
O-level	17.0 (12.0; 20.4)	7.0 (3.7; 10.7)	24.1 (18.6; 27.9)	32.5 (28.7; 36.9)	25.5 (20.8; 29.6)	49.5 (44.0; 54.1)	963
A-level	18.3 (13.5; 23.2)	6.6 (3.0; 10.2)	24.8 (20.1; 30.3)	32.4 (28.3; 36.5)	25.8 (21.3; 29.5)	50.7 (45.5; 56.0)	911
HE	22.3 (16.8; 30.6)	7.7 (1.7; 13.8)	30.0 (24.6; 36.1)	30.4 (25.0; 35.2)	22.7 (18.0; 26.7)	52.7 (48.0; 58.1)	871
any: ATE	18.3 (14.2; 22.1)	7.0 (3.5; 10.3)	25.3 (21.7; 29.5)	32.1 (28.2; 36.1)	25.1 (20.8; 29.2)	50.4 (45.7; 54.8)	3,251
OLS	16.3 (12.7; 19.9)	7.1 (3.7; 10.4)	23.3 (19.5; 27.2)	31.4 (28.0; 34.8)	24.3 (20.8; 27.9)	47.6 (43.6; 51.8)	3,639

Incremental treatment effects (% wage gain) - Matching and OLS estimates Table 5.2

Notes:

Controlling for ethnicity, region, standard parental background information, tests at 7, school variables.

Matching estimates: based on 'best' specification, always imposing common support. 95% bias-corrected percentile confidence intervals obtained by bootstrapping for the matching estimates (700 repetitions);

for OLS, 95% confidence intervals based on robust standard errors.

In **bold**: average effect of treatment on the treated (*versus* no qualifications)

These average treatment effects for O-levels and A-levels basically coincide with the corresponding effects of treatment on the treated, while for HE, the return for those actually obtaining it (ATT) is slightly higher (52.7%) than for those who did not obtain it (49-50%). The disaggregated results in the table interestingly show that this higher average effect of HE (vis-à-vis no qualifications) on those with HE actually stems from individuals with HE enjoying a higher return from their O-level investment than other individuals (22.3% compared to 15-18%), In fact, the incremental return from A-levels to HE is lowest (though not by a wide margin) exactly for those individuals with HE.

From the results in the table, then, two educational groups appear particularly interesting to discuss: the individuals who achieved HE and the (baseline) group of individuals who left school without any qualifications. On the other hand, for the other two groups with O-levels and with A-levels, (observable) heterogeneity in impacts wouldn't seem to be a particularly important feature of these data; average treatment effects are quite close to those on the treated and on the non-treated, and similarly close to standard OLS estimates.

We thus turn to comparing the results from the single and the multiple treatment models that concern HE. Note that these latter results contain more disaggregated information than that arising from the single treatment model, where the average return to HE was estimated vis-à-vis a comparison state implicitly given by an average of the other three educational outcomes (no qualifications, O-levels and A-levels). Similarly, the results from the single treatment model concerning the average return to HE for the non-treated effectively pooled individuals with quite different achieved educational levels into one group. Several dimensions highlighted in the multiple treatment model were thus missed in the single treatment one.

Comparing the two sets of results for the HE group¹¹, we can infer that their 25.4% return to HE compared to anything less than HE is mainly driven by their return compared to A-levels (22.7%).¹² It would seem in effect likely that on the basis of their

¹¹ Note however that we cannot control for the same set of variables across the two models, in particular only the single treatment model controls for tests at 16.

 $^{^{12}}$ To a lesser extend to their return compared to O-levels (30.4%) but most likely not by their return compared to no qualifications (52.7%).

characteristics, if for random reasons (i.e. unrelated to wages) they had not gone on until HE, this group would have stopped at A-levels.¹³

On the other hand, the single-treatment results show that individuals without HE would have had an additional 3% gain from HE than those individuals effectively achieving this highest level, compared to less than HE.

The multiple treatment model shows how this treatment on the non-treated result is effectively driven by the relatively large (counterfactual) returns to HE that the group without any qualifications would have reaped from HE compared to O-levels (33.4%) and to A-levels (26.6%). In fact, a closer look at this group reveals that they would have obtained the lowest returns to going on to O-levels, which could in fact justify their observed behaviour of stopping before. Nevertheless, as mentioned above, had they persevered until HE, they would have enjoyed the highest incremental return from it among the various educational groups, while those individuals effectively going on to HE appear to have reaped the smallest incremental returns.

Several caveats do however apply when looking at these results.

As to the average effect of HE on the non-treated arising from the single treatment model, matching does not seem to perform as good a job in balancing the *X*'s as it does for the average effect on the treated.¹⁴ In this case, not only do we start with two potentially highly diverse groups, but to find adequate matches from within the pool of potential HE controls for the no-HE individuals, at less than 40% of the no-HE group, is stretching the common support requirement considerably. In fact, 8.2% of the no-HE individuals fall out of the overlapping region and are dropped from the estimation of the average effect of HE on the non-treated. Furthermore, those few HE individuals who most closely resemble individuals without any qualifications will (rightly) receive more weight in estimation; any outlier problem with these comparably few observations will be reflected and magnified in the ATNT estimate.

More in general, identifying the average effect on the non-treated parameter requires more restrictive assumptions; as highlighted in Section 2.5.1 of Chapter 4, the condi-

¹³ A further heuristic confirmation is obtained if we perform one-to-one matching; in this case, individuals with A-levels make up 56% of the nearest-neighbour matched controls, individuals with O-levels 37% and individuals with no qualifications the remaining 7%.

¹⁴ For ATT: 1.4% mean bias and 1.7% median bias left; for ATNT: 3% mean bias and 3.6% median bias left.

tional independence assumption needs to be strengthened to include both counterfactual outcomes (thus excluding selection based on unobserved returns), and the common support assumption (or restriction) similarly needs to be extended. Furthermore, to recover a reliable estimate of the ATNT more stringent demands are placed on the data if, as in our model and as often is the case, the non-treated represent a larger group than the treated; in this latter case, considerable initial differences in the two groups would make it hard to obtain close matches which are not outlier-dependent. Note incidentally that since the average treatment effect is an average of the treatment on the treated and on the non-treated parameters, it will be affected by a poorly estimated effect on the non-treated.

As to the reliability of the gain individuals without qualifications would have enjoyed had they achieved a HE qualification instead, the same considerations apply. The HE and no-qualifications group are extremely different groups, way apart to believe that all the relevant observables characterising them could be adequately controlled for (by matching or OLS), due to the comparability and common support issues highlighted above. Finally, one might additionally argue that considerable unobservable bias may still characterise the two matched groups (note that in the multiple treatment model we can no longer condition on ability measures at 16).

The discussion of these results has highlighted (cf. also Heckman, LaLonde and Smith, 1999) how in contrast to standard parametric methods like OLS, matching estimators can appropriately highlight the problem of common support and thus of the actual comparability of groups of individuals. As emphasised also by Dehejia and Wahba (1999), matching represents an informative starting point, in that it easily reveals the extent to which the treatment and non-treatment groups overlap in terms of preintervention variables. The matching approach also offers simple and effective ways of *ex post* assessing the quality of a matched comparison group. Non-parametric methods such as matching thus force the researcher to compare only comparable individuals; if on the other hand, treated and non-treated are too different in terms of the observables, the researcher needs to accept the fact that there simply is not enough information in the available data to achieve sufficiently close – and thus reliable – matches.¹⁵

¹⁵ Although various diagnostic tools are available to assess 'matching quality', the question of how close the matches should be in practice, or equivalently how well the observables should be balanced in the two matched samples, is left to the sensitivity of the individual researcher.

By contrast, OLS estimators might hide from the analyst the possibility that observationally different individuals may *de facto* end up being compared on the basis of extrapolations purely based on the imposed functional form.

In conclusion, although it would seem from the discussion of the single-treatment results that fully interacted OLS represents a reasonable way to estimate educational returns in the NCDS data, matching presents several advantages (while not suffering a loss in precision compared to even standard OLS). On a practical front, it automatically both allows for all interactions and averages over the appropriate distributions. On a more substantive and 'good practice' front, even if in the present application it reproduced the results obtained by the regression-based methods, it allows to highlight the issue of the true comparability of the groups (common support, balancing of the X's, ATT *versus* ATNT) which would have been missed in the other approaches. It thus helped in determining which results could be viewed as most reliable.

V.4 Appendix

Variable	Mean	Std Dev
Real log hourly wage 1991	2.040	(0.433)
Qualifications:		
O levels or equivalent	0.821	(0.383)
A levels or equivalent	0.548	(0.498)
Higher Education	0.283	(0.451)
White	0.969	(0.173)
Maths ability at 7:		
5th quintile (highest)	0.212	(0.408)
4th quintile	0.190	(0.392)
3rd quintile	0.185	(0.389)
2nd quintile	0.158	(0.365)
1st quintile (lowest)	0.141	(0.348)
Reading ability at 7:		
5th quintile (highest)	0.165	(0.371)
4th quintile	0.187	(0.390)
3rd quintile	0.188	(0.391)
2nd quintile	0.179	(0.383)
Ist quintile (lowest)	0.166	(0.372)
Ability at / missing	0.115	(0.319)
Maths ability at 16:	0.222	(0.417)
Sth quintile (highest)	0.223	(0.417)
4th quintile	0.175	(0.380)
3rd quintile	0.152	(0.359)
2nd quintile	0.124	(0.330)
Prading ability at 16.	0.225	(0.417)
Sth quintile (highest)	0.202	(0.402)
Ath quintile	0.202	(0.402) (0.380)
3rd quintile	0.173	(0.360)
2nd quintile	0.133	(0.300) (0.329)
1st quintile (lowest)	0.234	(0.32)
Ability at 16 missing	0.202	(0.423) (0.402)
Comprehensive school 1974	0.468	(0.102) (0.499)
Secondary modern school 1974	0.162	(0.368)
Grammar school 1974	0.099	(0.299)
Private school 1974	0.052	(0.222)
Other school 1974	0.018	(0.134)
Father's years of education	7.270	(4.827)
Father's education missing	0.172	(0.377)
Mother's years of education	7.342	(4.606)
Mother's education missing	0.159	(0.366)
Father's age 1974	43.17	(13.74)
Father's age missing	0.075	(0.263)
Mother's age 1974	41.48	(10.86)
Mother's age missing	0.049	(0.216)
Father's social class 1974:		
Professional	0.044	(0.205)
Intermediate	0.145	(0.352)
Skilled non-manual	0.076	(0.265)
Skilled manual	0.297	(0.457)
Semi-skilled non-manual	0.010	(0.098)
Semi-skilled manual	0.095	(0.293)
Missing	0.106	(0.308)
Mother employed 1974	0.513	(0.500)
Number of siblings	1.692	(1.789)
Number of siblings missing	0.106	(0.308)
Number of older siblings	0.821	(1.275)
Father's interest in education:	0.010	(0.11.0)
Expects too much	0.013	(0.114)
Very interested	0.252	(0.434)
Some interest	0.215	(0.411)

Table 5A.1 Summary statistics, NCDS Men (N = 3,639)

Mother's interest in education:		
Expects too much	0.032	(0.175)
Very interested	0.344	(0.475)
Some interest	0.354	(0.478)
Bad finances 1969 or 1974	0.159	(0.365)
Region 1974:		
North Western	0.100	(0.300)
North	0.070	(0.256)
East and West Riding	0.079	(0.270)
North Midlands	0.072	(0.258)
Eastern	0.073	(0.261)
London and South East	0.143	(0.350)
Southern	0.057	(0.232)
South Western	0.061	(0.240)
Midlands	0.088	(0.283)
Wales	0.054	(0.227)
Scotland	0.096	(0.295)

 Table 5A.2 Matching quality indicators (best specification)

Treatment	N	Comparison	Ν	Mean bias	Median bias
HE	1,030	no-HE	2,609	1.4	1.7
no-HE	2,609	HE *	1,030	3.0	3.6
none	651	O-level	<i>993</i>	1.0	1.3
		A-level	965	1.5	2.1
		HE	1,030	3.7	5.1
O-level	993	none	651	2.1	2.7
		A-level	965	0.7	0.8
		HE	1,030	1.8	2.2
A-level	965	none	651	2.6	3.4
		O-level	<i>993</i>	0.8	1.1
		HE	1,030	1.6	1.8
HE	1,030	none **	651	5.0	11.7
		O-level	<i>993</i>	1.7	2.3
		A-level	965	1.4	1.8

Notes: Mean and median bias: mean and median overall absolute percentage bias, where the mean or median is taken over the post-matching absolute standardised differences of all variables in estimation of the choice model. For a given regressor, the standardised difference after matching is defined as the difference of the sample means in the treated and matched comparison sub-samples as a percentage of the square root of the average of the sample variances in the treated and comparison groups (cf. Rosenbaum and Rubin, 1985).

* 214 individuals (8.2%) are lost due to common support

** Full Mahalanobis-metric matching on $\{P(X), X\}$

Summary and conclusions to Part Two

The aim of this paper has been to review alternative methods and models for the estimation of the effect of education on earnings, and to apply these to a high quality common data source.

In the methodological Chapter 4, we have highlighted the importance of the model specification – in particular the distinction between single treatment and multiple treatment models – as well as the importance of allowing for heterogeneous returns – that is returns that vary across individuals for the same educational qualification. We have considered four main estimation methods which rely on different identifying assumptions – least squares, instrumental variable methods, control function methods and propensity score matching methods. The properties of the estimators were analysed distinguishing between a single treatment model and a model where there are a sequence of possible treatments. We argued that the sequential multiple treatment model is well suited to the education returns formulation, since educational qualification levels in formal schooling tend to be cumulative.

With heterogeneous returns, defining the 'parameter of interest' is central. We distinguished four possible parameters of interest: the treatment on the treated, the average treatment effect, the impact of treatment on the non-treated and the local average treatment effect. In the homogeneous effects model these would all be equal, but in the heterogeneous effects model they can differ substantially. Which one is of interest will depend on the policy question. Moreover, different estimation methods were shown to identify different parameters of interest.

The choice of an appropriate estimator should thus be guided by the postulated model generating outcomes and participation, by the parameter of policy interest, as well as by the nature and richness of the data available in the specific application.

Our application in Chapter 5 aimed at estimating the wage returns to different educational investments using the NCDS 1958 birth cohort study for Britain. This data set is in fact ideally suited for evaluating the impact of education on earnings using nonexperimental methods and is sufficiently rich to allow the comparison of least squares, matching, control function and instrumental variable methods. On the one hand, there are extensive and commonly administered ability tests at early ages, as well as accurately measured family background and school type variables, all ideal for methods relying on the assumption of selection on observables, notably least squares and matching. On the other hand, there are also variables likely to influence schooling but not wage outcomes, such us the age composition of siblings. These could represent reasonable choices for excluded instruments in the application of instrumental variables or control function methods.

This application has highlighted the following key points:

1) Correcting for detailed background variables and ability differences is important and reduces the return to education at all levels; the basic pre-education information available in common datasets would not have been enough to identify gains in an unbiased way.

2) The overall returns to educational qualifications at each stage of the educational process remain however sizeable and significant, even after allowing for (observed) heterogeneity in the education response parameters.

3) At times it may in fact be important to allow for a flexible specification in the *X*'s – in particular to allow for (observably) heterogeneous returns – in order to obtain unbiased point estimates. It may happen, though, that once the effects are averaged by groups, the resulting estimate of the average effect on the treated is not too different from the average treatment effect, although in our application the effect for a treated person tended to be statistically different from the effect on an untreated individual.

4) In terms of efficiency, in our application matching performs extremely well, providing point estimates which are always just as precise as the corresponding OLS ones. Control function estimates are by contrast considerably less precise, whilst those arising from IV fall nowhere near significance – although the argument for IV estimation in our case and with the instrument we have is very weak.

5) Overall, matching has been found to perform very well in our application:

First of all, and although obvious, it may be worth reiterating that matching is as good as the X's it uses (cf. also Smith and Todd, 2000, and more generally the "better

data help a lot" lesson highlighted in Heckman, LaLonde and Smith, 1999, p.1868). Matching (like OLS) thus requires very rich data to make its identification strategy credible. The NCDS dataset seems however to be informative enough to allow the researcher to control *directly* for selection on unobservables – both unobservable individual traits and unobservable returns.

We also noted in point (3) above how estimators which allow for an (observably) heterogeneous gain from educational investments not only avoid potential omitted variable bias, but also provide additional interesting information as to the average gains for the various subgroups in the population. From a practical point of view, propensity score matching is a very convenient way of implementing a flexible specification compared to fully interacted OLS, since it automatically allows for all interactions and by design averages over the appropriate distributions. It may also be worth mentioning that in our application propensity score matching estimates did not suffer a loss in precision compared to even standard OLS.

Finally, in contrast to standard parametric methods like OLS, matching forces the researcher to effectively compare only comparable individuals, thus helping to discriminate between more reliable results and those which should be viewed with particular caution.

References

- Angrist, J.D., Imbens, G. and Rubin, D.B. (1996), "Identification of causal effects using instrumental variables", *Journal of the American Statistical Association*, 91, 444-472.
- Angrist, J.D. and Krueger, A. B. (1991), "Does compulsory schooling attendance affect schooling decisions", *Quarterly Journal of Economics*, 106, 4, 970-1014.
- Angrist, J.D. and Krueger, A. B. (1992), "Estimating the payoff to schooling using the Vietnam-era draft lottery", National Bureau of Economic Research, Working Paper no. 4067.
- Blundell, R., Dearden, L., Goodman, A. and Reed, H. (2000), "The returns to higher education in Britain: Evidence from a British cohort", *Economic Journal*, 110, F82-F99.
- Bound, J., Jaeger, D. and Baker, R. (1995), "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak", *Journal of the American Statistical Association*, 90, 443-50.
- Butcher, K. and Case, A. (1994), "The effect of sibling sex composition on women's education and earnings", *Quarterly Journal of Economics*, 109, 531-63.
- Card, D. (1999), "The causal effect of education on earnings", in O. Ashenfelter and D. Card, *Handbook of Labor Economics*, Vol 3, Elsevier-North Holland.
- Cochran, W. and Rubin, D.B. (1973), 'Controlling bias in observational studies', *Sankyha*, 35, 417-446.
- Dearden, L. (1999a), "The effects of families and ability on men's education and earnings in Britain", *Labour Economics*, 6, 551-567.
- Dearden, L. (1999b), "Qualifications and earnings in Britain: How reliable are conventional OLS estimates of the returns to education?", IFS Working Paper No. 99/7.
- Dehejia, R.H. and Wahba, S. (1999), 'Causal effects in non-experimental studies: reevaluating the evaluation of training programmes', *Journal of the American Statistical Association*, 94, 1053-1062.
- Garen, J. (1984), "The returns to Schooling: A selectivity bias approach with a continuous choice variable", *Econometrica*, 52, 1199-1218.
- Griliches, Z. (1977), "Estimating the returns to schooling: some econometric problems", *Econometrica*, 45, 1-22.
- Harmon, C. and Walker, I. (1995), "Estimates of the economic return to schooling for the UK", *American Economic Review*, 85, 1278-86.
- Heckman, J.J. (1979), "Sample selection bias as a specification error", *Econometrica*, 47, 153-61.

- Heckman, J.J and Robb, R. (1985), "Alternative methods for evaluating the impact of interventions", in *Longitudinal Analysis of Labour market Data*, New York: Wiley.
- Heckman, J.J., Ichimura, H. and Todd, P. (1997), "Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme", *Review of Economic Studies*, 64, 605-654.
- Heckman, J.J., Ichimura, H., and Todd, P (1998), "Matching as an econometric evaluation estimator", *Review of Economic Studies*, 65, 261-294.
- Heckman, J.J., Ichimura, H., Smith, J. and Todd, P. (1998), "Characterizing selection bias using experimental data", *Econometrica*, 66, 1017-98.
- Heckman, J.J., R. LaLonde and J. Smith (1999), "The economics and econometrics of Active Labor Market Programs" in O. Ashenfelter and D. Card, *Handbook of Labor Economics*, Vol 3, Elsevier-North Holland.
- Heckman, J.J., Smith, J. and Clements, N., (1997), "Making the most out of program evaluations and social experiments: accounting for heterogeneity in program impacts", *Review of Economic Studies*, 64, 487-536.
- Holland, P.W. (1986), "Rejoinder", *Journal of the American Statistical Association*, 81, 396, 968-970.
- Imbens, G. (2000), "The role of propensity score in estimating dose-response functions", *Biometrika*, 87, 3, 706-710.
- Imbens, G. and Angrist, J. (1994), "Identification and estimation of local average treatment effects", *Econometrica*, 62, 2, 467-476.
- Lechner, M. (2001a), "Identification and estimation of causal effects of multiple treatments under the conditional independence assumption", in Lechner, M., and Pfeiffer, F., *Econometric Evaluation of Labour Market Policies*, Heidelberg: Physica/Springer, 43-58.
- Lechner, M. (2001b), "A note on the common support problem in applied evaluation studies", Discussion Paper 2001-01, Department of Economics, University of St. Gallen.
- Lechner, M. and Miquel, R. (2001), "A potential outcome approach to dynamic programme evaluation – Part I: Identification", discussion paper, SIAW, University of St. Gallen.
- Meghir, C. and Palme, M. (2000), "Estimating the effect of schooling on earnings using a social experiment", IFS Working Paper 99/12.
- Rosenbaum, P.R. and Rubin, D.B. (1983), "The central role of the propensity score in observational studies for causal effects", *Biometrika*, 70, 41-55.
- Rosenbaum, P.R. and Rubin, D.B. (1985), "Constructing a comparison group using multivariate matched sampling methods that incorporate the propensity score", *The American Statistician*, 39, 33-38.
- Rubin, D.B. (1979), "Using multivariate matched sampling and regression adjustment to control bias in observational studies", *Journal of the American Statistical Association*, 74, 318-329.

- Rubin, D.B. (1980), "Discussion of 'Randomisation analysis of experimental data in the fisher randomisation test" by Basu, *Journal of the American Statistical Association*, 75, 591-593.
- Rubin, D.B. (1986), "Discussion of 'Statistics and causal inference" by Holland, *Journal of the American Statistical Association*, 81, 396, 61-962.
- Staiger, D. and J. H. Stock (1997), "Instrumental variables regressions with weak instruments", *Econometrica*, 65, 557-86.
- Smith, R. and Blundell, R.W. (1986), "An exogeneity test for a simultaneous tobit model with an application to labour supply", *Econometrica*, 54, 679–685.
- Smith, J. and Todd, P. (2000), "Does matching overcome LaLonde's critique of nonexperimental estimators?", November, mimeo.
- Willis, R. J. and Rosen, S. (1979), "Education and self-selection", *Journal of Political Economy*, 87, S1-S36.
- Wooldridge, J. (1997), "On two stage least squares estimation the of average treatment effect in a random coefficient model", *Economic Letters*, 56, 129-133.

Publication series published by the Institute for Labour Market Policy Evaluation (IFAU) – latest issues

Rapport (some of the reports are written in English)

- **2002:1** Hemström Maria & Sara Martinson "Att följa upp och utvärdera arbetsmarknadspolitiska program"
- 2002:2 Fröberg Daniela & Kristian Persson "Genomförandet av aktivitetsgarantin"
- 2002:3 Ackum Agell Susanne, Anders Forslund, Maria Hemström, Oskar Nordström Skans, Caroline Runeson & Björn Öckert "Follow-up of EU's recommendations on labour market policies"
- **2002:4** Åslund Olof & Caroline Runeson "Follow-up of EU's recommendations for integrating immigrants into the labour market"
- **2002:5** Fredriksson Peter & Caroline Runeson "Follow-up of EU's recommendations on the tax and benefit systems"
- **2002:6** Sundström Marianne & Caroline Runeson "Follow-up of EU's recommendations on equal opportunities"
- **2002:7** Ericson Thomas "Individuellt kompetenssparande: undanträngning eller komplement?"
- **2002:8** Calmfors Lars, Anders Forslund & Maria Hemström "Vad vet vi om den svenska arbetsmarknadspolitikens sysselsättningseffekter?"
- **2002:9** Harkman Anders "Vilka motiv styr deltagandet i arbetsmarknadspolitiska program?"
- 2002:10 Hallsten Lennart, Kerstin Isaksson & Helene Andersson "Rinkeby Arbetscentrum – verksamhetsidéer, genomförande och sysselsättningseffekter av ett projekt för långtidsarbetslösa invandrare"
- 2002:11 Fröberg Daniela & Linus Lindqvist "Deltagarna i aktivitetsgarantin"

Working Paper

- **2002:1** Blundell Richard & Costas Meghir "Active labour market policy vs employment tax credits: lessons from recent UK reforms"
- **2002:2** Carneiro Pedro, Karsten T Hansen & James J Heckman "Removing the veil of ignorance in assessing the distributional impacts of social policies"
- **2002:3** Johansson Kerstin "Do labor market programs affect labor force participation?"
- **2002:4** Calmfors Lars, Anders Forslund & Maria Hemström "Does active labour market policy work? Lessons from the Swedish experiences"

- **2002:5** Sianesi Barbara "Differential effects of Swedish active labour market programmes for unemployed adults during the 1990s"
- **2002:6** Larsson Laura "Sick of being unemployed? Interactions between unemployment and sickness insurance in Sweden"
- 2002:7 Sacklén Hans "An evaluation of the Swedish trainee replacement schemes"
- **2002:8** Richardson Katarina & Gerard J van den Berg "The effect of vocational employment training on the individual transition rate from unemployment to work"
- **2002:9** Johansson Kerstin "Labor market programs, the discouraged-worker effect, and labor force participation"
- **2002:10** Carling Kenneth & Laura Larsson "Does early intervention help the unemployed youth?"
- 2002:11 Nordström Skans Oskar "Age effects in Swedish local labour markets"
- **2002:12** Agell Jonas & Helge Bennmarker "Wage policy and endogenous wage rigidity: a representative view from the inside"
- **2002:13** Johansson Per & Mårten Palme "Assessing the effect of public policy on worker absenteeism"
- **2002:14** Broström Göran, Per Johansson & Mårten Palme "Economic incentives and gender differences in work absence behavior"
- **2002:15** Andrén Thomas & Björn Gustafsson "Income effects from market training programs in Sweden during the 80's and 90's"
- **2002:16** Öckert Björn "Do university enrollment constrainsts affect education and earnings?"
- **2002:17** Eriksson Stefan "Imperfect information, wage formation, and the employability of the unemployed"
- **2002:18** Skedinger Per "Minimum wages and employment in Swedish hotels and restaurants"
- **2002:19** Lindeboom Maarten, France Portrait & Gerard J. van den Berg "An econometric analysis of the mental-health effects of major events in the life of older individuals"
- **2002:20** Abbring Jaap H., Gerard J. van den Berg "Dynamically assigned treatments: duration models, binary treatment models, and panel data models"

Dissertation Series

2002:1 Larsson Laura "Evaluating social programs: active labor market policies and social insurance"

- **2002:2** Nordström Skans Oskar "Labour market effects of working time reductions and demographic changes"
- **2002:3** Sianesi Barbara "Essays on the evaluation of social programmes and educational qualifications"