EUROPEAN UNIVERSITY INSTITUTE
Department of Economics

Contributions to causal inference with an application to the estimation of returns to job mobility during transition and returns to education in Germany

Frank Siebern-Thomas

Thesis submitted for assessment with a view to obtaining the degree of Doctor of the European University Institute

Florence June 2002

13 | C-D

EUROPEAN UNIVERSITY INSTITUTE
Department of Economics

Contributions to causal inference with an application to the estimation of returns to job mobility during transition and returns to education in Germany

Frank Siebern-Thomas

The Thesis Committee consists of:

Prof. Bart Cockx, Université Catholique de Louvain
" Jennifer Hunt, University of Montreal
" Andrea Ichino, EUI, Supervisor
" Grayham Mizion, University of Southampton, Supervisor

Contents

[1]joint work with Sascha O. Becker

List of Tables

List of Figures

"Conhecia muita gente por aquelas bandas - e por isso gostava de viajar. A gente acaba sempre por fazer amigos novos, e sem a necessidade de ficar com eles dia após dia. Quando vemos sempre as mesmas pessoas - e isto acontecia no seminário - acabamos por considerar que elas fazem parte das nossas vidas. E como elas fazem parte das nossas vidas, passam também a querer modificar as nossas vidas. E se nós não formos como elas esperam, ficam chateadas."

Paulo Coelho, O Alquimista

Acknowledgement

I am indebted in particular to Professors Ichino, Mizon and Phlips for supervision and advice throughout my postgraduate studies at the European University Institute; I would further like to thank Professors Hasenkamp and Hübner (Universität Hamburg) and Professors Bauwens, Cockx, Mouchart and Vanderlinden (Université Catholique de Louvain) for their support and supervision throughout my undergraduate studies; Fragiskos Archontakis, Enrico Bernini, Asuncion Soro Bonmati, Rainer Kiefer, Alvaro Pina, Daniela Vuri, and in particular the co-author of chapter 5, Sascha Becker, for their comments, their encouragement and their friendship; Jennifer Hunt, Viktor Steiner and Iris Zöllner for their detailed comments on the paper version of chapter 4 which have led to significant improvements of this central chapter of the thesis; and seminar participants in Berlin, Florence, Hamburg, Lausanne, Louvain-la-Neuve, Mannheim, München, Rydzyna and Salerno whose comments also contributed to improving previous versions of this thesis. It goes without saying that none of them is responsible for the approach taken, for the errors present or for any omissions made in this thesis.

I am further truly grateful to the *Deutsches Institut für Wirtschaftsforschung* for providing the microdata underlying the empirical analyses in this thesis; to the *Verein der Freunde des DIW* and in particular to Professors Hamermesh, Trommsdorf and Wagner for selecting the paper version of chapter 4 as best junior publication 1999-2001 using microdata from the German Socio-Economic Panel (GSOEP); to the *Studienstiftung des deutschen Volkes* for material and immaterial support throughout my undergraduate studies; to the *Deutscher Akademischer Austauschdienst* and to the *European University Institute* for financial support throughout my postgraduate studies; and, last but not least, to the Chirici and Kübli families for their warm welcome, and to Stefanie and my family for all their love and patience.

Chapter 1

Introduction

1.1 Context and Outline

More than ten years after the opening of the Berlin wall in November 1989 and the subsequent unification of the two German states in October 1990, the economic situation in East Germany continues to be a topic for controversial debates. In particular questions about the potential individual benefits and losses and about the likely 'winners' and 'losers' of the transition process remain at the center of these debates.

Both survey results on self-reported gains and losses experienced by East Germans throughout transition and results from empirical studies on the distribution and evolution of income and income changes in the transition process show an ambiguous picture: while a large majority of Eastern Germans seems to have benefitted from the transition to a market economy, a substantial - and recently increasing - share of around one fifth of the East Germans seems to have experienced income losses or a deterioration of their economic situation throughout the transition process.[1]

Other - indeed numerous - studies have contributed to this picture by showing how the returns to individual characteristics have changed during transition. These studies helped to disentangle those individuals who seemed to benefit most from the transition - the highly educated and the young - from those who had difficulties in adapting to the needs of a rapidly

[1] Cf. e.g. Beblo et al. (2001a,b).

changing economic, political and social environment.[2]

While these studies provide a picture of the overall evolution and further allow in some cases a characterisation of the groups experiencing income gains or losses throughout transition, they do generally not relate these results to individual behaviour but only to individual endowments and the job and individual characteristics in 1990. The fact that certain groups have experienced income gains or losses throughout transition, however, does not entail any *causality* between individual behaviour, on the one hand, and well-being, income or employment status, on the other.

As regards job mobility, there seems to be a general understanding in the economics literature that there are positive individual returns to job changes, both in the short and in the long run.[3] In his analysis of short- and long-run wage changes in job changes, Mincer (1986) e.g. shows that job movers generally experience positive wage gains and that these gains are higher after quits than after layoffs.[4] Surprisingly, many analysts seem tempted to apply such "stylised facts" in a relatively straightforward way to the particular case of job mobility in the transition from a planned economy to a market economy. These analysts seem to disregard that even in market economies that are not subject to dramatic structural changes there is in general a large dispersion of wage gains experienced by job movers. As also mentioned in Mincer (1986), "significant numbers of movers incur losses" and in particular experienced movers, "despite average gains in moving, do not catch up with wage levels of stayers." This applies in particular to older workers and to workers who were laid off. Like Bartel and Borjas (1981), Mincer (1986) provides evidence that even quits might result in wage losses, in particular quits which are either due to exogenous reasons such as family or health reasons or which represent trade-offs of wages for other working conditions. Given this heterogeneity of wage responses to job mobility even in the case of a "stationary" economic environment, it seems largely unclear

[2]Cf. e.g. Bird et al. (1994), Franz and Steiner (1999), Hunt (1997,1999a,b), Krueger and Pischke (1995) and Steiner and Puhani (1997), in the case of Germany, and Brainerd (1998), Chase (1998), Flanagan (1990), Munich et al. (2000), Orazem and Vodopivec (1995,1998), Rutkowski (1994a,b,1996,1997), Sveijnar (1998) and Vecernik (1995), in the case of other Central and Eastern European transition economies.

[3]Similarly, 'macroeconomic' effects of labour mobility throughout transition in terms of employment and the speed of convergence in wages and living standards are often shown theoretically to be positive. Such 'macroeconomic' effects, however, are not subject of this thesis and will not be discussed any further.

[4]Mincer (1986) defines "short-run wage changes" as the difference between the starting wage on the new job and the last wage on the old job and "long-run wage changes" as the difference in wages between the two jobs at the same tenure levels net of experience.

whether, under what conditions and for which groups on the labour market job mobility pays off in a "nonstationary" transitional labour market. In both cases it is further questionable whether the above wage responses can be interpreted as *causal effects* of job mobility on wages.

This dissertation contributes to the analysis and evaluation of individual behaviour in an uncertain transitional environment both from a methodological and empirical point of view. It addresses in particular two problems: first, the underlying economic problem of optimal individual job mobility decision making in an uncertain environment, and second, the identification and estimation of the *causal* effects of job mobility early in transition on employment and income. To this aim, a new quasi-experimental approach to the analysis of (individual) 'gains' and 'losses' following job mobility during transition is proposed. While drawing on the statistical literature on causal effects and its adaptations to socioeconomic problems,[5] this approach takes into account the particularity of individual decision making and self-selection in a transitional environment. In this latter, the driving forces of self-selection into job mobility are indeed less obvious than in a non-transitional framework. While individuals in both self-select into mobility on the basis of their expected gains, as described in the so-called Roy (1951) model, the relatively high uncertainty at the beginning of transition led to particular uncertainties regarding the potential benefits of either changing job immediately or waiting at least for a partial resolution of uncertainty. In this context, the role of unobservable individual characteristics such as ability and access to either old or new networks is particularly difficult to predict.

Before developing a theoretical model of individual job mobility decision making which clarifies both the endogeneity of job mobility and the need for appropriate empirical approaches for the analysis of returns to job mobility in chapter 2, the following sections, first, provide the necessary overview of job mobility during transition in the context of the overall labour market evolution, and, second, discuss the theoretical predictions and empirical results of returns to job mobility in an uncertain transitional environment. After reviewing the identification and estimation of causal effects in a quasi-experimental framework and introducing the main notation in chapter 3, chapter 4 - the central chapter of this thesis - provides an empirical evaluation of the causal effects of early job mobility in transition. Chapter 5 applies the same

[5] Cf. chapter 3 and the references cited therein.

identification and estimation methods to the estimation of the returns to education in West Germany. Chapter 6 concludes.

1.2 Incidence and timing of job mobility during transition

The substantial reallocation of labour is among the most pervasive features of any transition economy and has attracted considerable attention in economic research, both theoretical and empirical. The incidence of both job endings and job mobility in the East German transition process was dramatic:[6] More than 70% (West: 43%) of the individuals had ended their 1990 employment relationship by 1996, and more than 40% (West: 14%) had been displaced at some point since 1990. Two thirds (West: 34%) of the sample population, moreover, had a job change at least once between 1990 and 1996, and more than 40% (West: 21%) performed at least one job-to-job shift in the same period. Less than 20% (West: 50%) of the sample actually stayed with their 1990 job throughout the full period 1990-96. The massive job loss especially at the beginning of the transition period, when displacement rates in the East exceeded those in the West by a factor of 5 to 10, was thus accompanied by a considerable reallocation of labour through job mobility, with job mobility rates in the East generally doubling those in the West (cf. table 1.1)

As to the time pattern of job mobility during transition, while starting from high levels in 1990, the rates of both job endings and job changes peaked in 1991 and declined considerably afterwards. Throughout the first three years of transition, 1990-92, 52% of the individuals employed in 1990 had ended their 1990 employment relationship, more than half of them due to displacement. In the same time, more than 26% of the sample had changed employer. Whereas job mobility rates have come close to Western levels by 1996, displacement rates in 1996 were still considerably higher in the East.

As a result, employment rates in East Germany have dropped by 10 percentage points in

[6]The evidence presented here is based on data from the East and West German subsamples of the German Socio-Economic Panel (GSOEP) for the years 1990-97 (waves 7-14). "Full sample" denotes the balanced sample of all individuals with observations for the first eight years of the transition process. "Restricted sample" denotes the subsample of all those individuals who were employed in 1990. Tables in this section are based on the "restricted sample". For reasons of completeness, tables for the "full sample" are contained in the appendix. For a detailed description of the database and the sample definition, cf. chapter 4.

Table 1.1: Incidence of job mobility 1990-96 in East and West Germany

	East 1990	1991	1992	1993	1994	1995	1996
job ending							
yearly	17.9	22.43	21.50	15.34	12.31	15.11	12.62
cumulative	17.9	38.01	52.18	58.72	62.85	67.68	70.79
displacement							
yearly	8.4	13.4	10.3	7.4	5.5	6.7	5.4
cumulative	8.4	20.6	29.2	34.1	37.7	41.0	43.7
quits							
yearly	5.9	5.0	3.8	2.7	2.7	2.7	1.6
cumulative	5.9	10.8	13.7	15.6	17.1	19.1	19.8
job change							
yearly	17.5	24.9	17.3	14.1	12.9	10.8	10.4
cumulative	17.5	38.9	49.9	56.6	61.1	64.3	67.0
job-to-job shift							
yearly	8.3	12.2	7.9	8.0	7.9	6.8	5.8
cumulative	8.3	20.2	26.2	31.5	36.1	39.6	42.1

	West 1990	1991	1992	1993	1994	1995	1996
job ending							
yearly	9.5	7.8	8.0	8.19	9.31	8.75	9.54
cumulative	9.5	16.2	21.9	27.08	32.82	37.41	42.92
displacement							
yearly	1.3	1.2	1.9	3.6	2.8	2.5	2.7
cumulative	1.3	2.5	4.3	7.6	9.9	11.7	13.6
quits							
yearly	4.4	3.8	3.4	1.9	2.4	2.1	1.7
cumulative	4.4	8.0	10.5	12.0	13.9	15.3	16.3
job change							
yearly	8.8	8.8	8.1	5.1	6.3	6.3	6.3
cumulative	8.8	16.3	22.1	25.4	28.7	32.2	34.4
job-to-job shift							
yearly	4.6	4.7	4.4	3.2	3.4	3.2	3.2
cumulative	4.6	9.0	12.5	14.8	17.0	19.1	20.7

NOTES: Yearly rates show the fraction of individuals who have experienced the form of job mobility in the calendar year. Cumulative rates show the fraction of individuals who have experienced the form of job mobility since 1990; SOURCE: GSOEP, waves 7-14, subsamples A and B (West Germany, N=2160) and subsample C (East Germany, N=1284)

Table 1.2: Labour market status 1990-97: East Germany (restricted sample)

LM status	1990	1991	1992	1993	1994	1995	1996	1997
employed								
total	100.0	92.2	85.1	79.1	77.8	77.4	74.0	73.6
full-time	88.2	83.9	78.9	72.6	70.5	69.3	65.2	65.1
part-time	11.8	7.8	5.7	5.4	6.2	7.1	7.6	7.2
(re)training		0.4	0.2	1.0	0.8	0.6	0.8	0.7
marginally		0.2	0.3	0.2	0.4	0.5	0.4	0.6
not employed								
total		7.8	14.9	20.9	22.2	22.6	26.0	26.4
unemployed		5.1	10.3	14.0	14.5	14.2	15.8	15.0
out of the LF		2.7	4.6	6.9	7.7	8.4	10.2	11.5

NOTES: in % of restricted sample (N=1284); labour market status as of time of survey for all categories with the exception of the calendar information which reports the percentage of individuals who were in the respective labour market state for at least one month up to the respective year and which is based on monthly calendar information; SOURCE: GSOEP public-use file, waves 7-14, subsample C, all individuals employed at the time of the 1990 survey

1990/91 alone, and by almost 25 percentage points between 1990 and 1997 - according to Hunt (1997) "possibly the worst [fall in employment rates] of any European transition economy." Levels of employment rates in East Germany have roughly converged to West German levels since 1992, although remaining slightly higher than in West Germany. The share of the employed in East Germany in full-time jobs also remains higher than in the West, while the share of individuals in part-time jobs remains almost 4 percentage points below that in the West. At the same time, both participation rates and unemployment rates remain considerably higher in East Germany, with unemployment rates in 1997 more than twice as high as in the West and participation rates more than 10 percentage points above those in the West. Of all those employed in 1990, more than one quarter were in non-employment in East Germany in 1997, compared to one fifth in West Germany. In East Germany almost two thirds of those not employed anymore in 1997 were in unemployment compared to only one third in West Germany (cf. tables 1.2 and 1.3 and tables 1.7 and 1.8 in the appendix).

Furthermore, by the end of 1996, more than 40% of all East Germans employed in 1990 had been unemployed and almost one quarter had left the labour force at least temporarily for some time. More than half of those experiencing unemployment did so during the first three

Table 1.3: Labour market status 1990-97: West Germany (restricted sample)

LM status	1990	1991	1992	1993	1994	1995	1996	1997
employed								
total	100.0	95.8	94.2	91.7	88.1	85.7	83.3	80.3
full-time	84.6	81.6	80.3	78.7	75.1	72.8	69.7	66.9
part-time	14.0	12.9	12.7	12.0	11.8	11.7	12.1	11.8
(re)training	0.2	0.5	0.3	0.4	0.4	0.2	0.2	0.2
marginally	1.1	0.9	0.8	0.7	0.9	1.0	1.2	1.4
not employed								
total		4.2	5.8	8.3	11.9	14.3	16.7	19.7
unemployed		1.1	1.5	3.3	5.6	6.1	6.4	7.2
out of the LF		3.1	4.3	5.0	6.4	8.2	10.3	12.6

NOTES: in % of restricted sample (N=2160); labour market status as of time of survey for all categories; SOURCE: GSOEP public-use file, waves 7-14, subsamples A and B, all individuals employed at the time of the 1990 survey

years of transition, 1990-92. The average time spent in either unemployment or inactivity has further increased throughout transition. While those unemployed in the first years of transition remained on average around 4 months in unemployment, those unemployed in 1994 or later remained on average unemployed for more than half a year. The same observation holds for the average duration of spells of inactivity (cf. table 1.4).

In the early 1990s, labour market dynamics thus were much more pronounced in East Germany than in the West. This applies not only to transitions into unemployment but also to transitions out of unemployment or inactivity into employment. After almost ten years of transition, the patterns of labour turnover in East and West Germany still remain significantly different. While in particular transition rates out of employment and unemployment remain significantly higher and transition rates into inactivity considerably smaller in East Germany, retention rates in employment, unemployment or inactivity have come close to Western levels: around 90% stay in either employment or inactivity, and more than half of all unemployed stay in unemployment between two consecutive years.[7]

Most interestingly, according to GSOEP data, in East Germany retention rates in employ-

[7]This section restricts attention to descriptive evidence on employment durations and transitions. For a detailed analysis of the determinants of such durations or transitions in general, and of transitions between emloyment and non-employment as well as of the duration in unemployment, cf. Hunt (1999b).

Table 1.4: Non-employment experience in East Germany 1990-96

	1990	1991	1992	1993	1994	1995	1996	1997
not employed								
total		7.8	17.9	28.3	33.3	37.9	43.0	45.8
unemployed		5.1	13.2	21.3	26.4	30.5	35.7	38.7
out of the LF		2.7	5.6	9.7	11.6	13.8	16.4	18.9
calendar information								
unemployed	7.3	12.7	23.1	30.1	33.3	37.9	41.6	
out of the LF	7.3	11.2	15.3	18.2	20.9	23.1	24.7	
average duration in ...								
employment		11.2	10.6	11.0	10.9	11.1	11.2	11.2
unemployment		2.8	4.8	5.8	6.5	6.5	6.5	6.4
out of the LF		3.5	5.3	6.1	5.9	5.8	6.1	6.7

NOTES: in % of restricted sample (N=1284); percentage of individuals who were in the respective non-employment state at least at one survey date up to the respective year; calendar information reports the percentage of individuals who were either in unemployment or out of the labour force (vocational training) for at least one month up to the respective year and is based on monthly calendar information; average duration (in months) in the respective labour market state (out of the labour force restricted to vocational training) SOURCE: GSOEP public-use file, waves 7-14, subsample C

ment have not increased and transition rates into unemployment have not decreased during transition. At the same time, transition rates into employment for those who lost or left their job in between have dropped sharply from levels above 40% in 1991/92 to disproportionately low levels in 1996/97, below 30% for transitions out of unemployment and below 10% for transitions out of inactivity. While the fall in these latter transition rates also occurred in West Germany, retention rates in employment remained significantly higher and transition rates into unemployment significantly lower in the West (cf. table 1.5 and table 1.9 in the appendix).

To sum up, the following observations seem particularly noteworthy. First, after initially dramatic increases the unemployment rate stabilised at a high level of around 15%, with almost 20% of the population and more than 40% of those employed in 1990 experiencing unemployment at least for some time. Second, the increase in unemployment was paralleled by a strong decline of the labour force. By 1996, more than 25% of those employed in 1990 were in non-employment, of which roughly 40% out of the labour force. Third, persistence of non-employment considerably increased over time due to decreasing outflow rates from non-employment to employment. At the same time, persistence of employment remained relatively

Table 1.5: Labour market transitions: East Germany (restricted sample)

LM transitions from ...	East 90/91	91/92	92/93	93/94	94/95	95/96	96/97
employment to:							
employment	92.2	88.6	87.3	91.5	91.3	89.2	92.2
unemployment	5.1	8.4	9.3	7.2	6.9	8.9	6.7
out of the labour force	2.7	3.0	3.4	1.3	1.8	1.9	1.1
unemployment to:							
employment		45.5	38.6	35.6	38.2	29.1	27.6
unemployment		43.9	50.0	52.2	51.6	53.9	54.2
out of the labour force		10.6	11.4	12.2	10.2	17.0	18.2
out of the LF to:							
employment		41.2	18.6	5.7	11.1	9.3	9.9
unemployment		11.8	20.3	21.6	17.2	15.7	13.7
out of the labour force		47.1	61.0	72.7	71.7	75.0	76.3
LM transitions from ...	West 90/91	91/92	92/93	93/94	94/95	95/96	96/97
employment to:							
employment	95.8	97.1	95.5	94.0	94.4	94.3	94.0
unemployment	1.1	0.9	2.1	3.9	2.8	2.6	3.0
out of the labour force	3.1	2.0	2.4	2.0	2.8	3.1	3.0
unemployment.to:							
employment		37.5	36.4	26.8	31.7	18.9	13.8
unemployment		41.7	45.5	52.1	58.3	62.9	63.0
out of the labour force		20.8	18.2	21.1	10.0	18.2	23.2
out of the LF to:							
employment		22.7	29.0	18.5	12.3	16.4	10.8
unemployment		7.6	14.0	4.6	6.5	4.0	6.3
out of the labour force		69.7	57.0	76.9	81.2	79.7	83.0

NOTES: one-year transition probabilities between labour market states at the time of the survey; SOURCE: GSOEP public-use file, waves 7-14, subsamples A and B (West Germany; N=2160) and subsample C (East Germany; N=1284); all individuals employed at the time of the 1990 survey

Table 1.6: Employed's expectations and mobility intentions 1990

subjective probability of	very likely	probably	unlikely	very unlikely
job loss	6.2	36.3	45.5	11.2
searching new job	16.2	17.2	21.3	44.6
quitting job	4.0	18.9	27.4	49.0
self-employment	1.6	6.2	15.4	75.9
promotion	1.6	13.1	37.4	47.2
demotion	3.0	9.8	36.9	48.8
leaving the labour force	0.9	4.9	13.8	79.3

NOTES: responses in % to 1990 survey question on subjective probability of the respective event; response categories are: very likely (1), probably (2), unlikely (3), very unlikely (4); differences to 100% are missing; SOURCE: GSOEP public-use file, wave 7, subsample C, N=1284

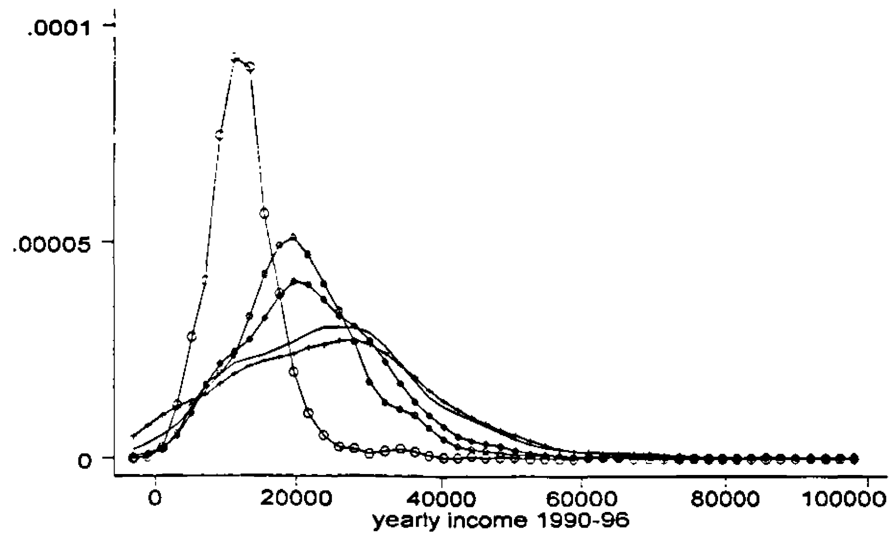
stable and at a high level of around 90%. Consequently, the average duration in non-employment spells increased strongly for both unemployment and inactivity. Finally, the high and increasing percentage in non-employment in general and in inactivity in particular contrasts sharply with the employed's expectations at the beginning of the transition process (cf. table 1.6). At that time, only 6% of the sample population recognised the possibility of dropping out of the labour force. Apart from economy-wide considerations, the combination of a high awareness of a potential job loss (42%) on the one hand and a relatively low propensity to search a new job (33%) or to quit the current job (23%) as of 1990 might explain partially the relatively high drop-out rates from the labour force.

1.3 Returns to job mobility during transition

Paralleling the above described labour market evolution, the income distribution as shown below (cf. figure 1-1) changed significantly throughout transition, most strongly between 1990 and 1992.[8] This suggests that those benefitting most strongly from the transition process could have been either those individuals who stayed on their job throughout the transition process or those who changed to new opportunities right at the beginning of this process.

[8]Among the clear signs of the - initially unexpected - slowdown of wage convergence towards Western levels in some sectors from 1993/94 onwards were the employers' single-sided withdrawal from collective wage agreements at that time and the increased public debate on wage convergence in the public sector. The determinants of wage growth during transition and the issue of wage convergence are discussed in more detail in Hunt (1997,1999a,b).

Figure 1-1: East German income distribution 1990-96



As described in the previous section, in these years both the risk of layoff as well as the accession rates from unemployment or inactivity to employment were also highest. Individuals were thus facing the trade-off between staying on their job - despite the largely unknown risk of layoff - and switching to a new job - despite the largely unknown risk of both job stability on the new job and the overall economic evolution throughout transition. This section briefly reviews theoretical issues related to individual job mobility decisions when facing this trade-off and further presents empirical evidence on returns to job mobility in transition from previous studies. It finally lays out the different approach to be followed in the later empirical analysis.

1.3.1 Theoretical considerations

Individual decision-making during transition is a standard example of decision-making under uncertainty with uncertainty referring not only to the further evolution of the economy in general and the labour market in particular but also to the development of the political and social environment. In this context, and given the dramatic changes in both the structure of labour demand and the reward structure, potential future costs and benefits of individual behaviour are inherently difficult to evaluate, especially at the beginning of the transition process. Variations in labour market policies throughout transition are prone to exacerbate this difficulty further.[9] Individual decisions to move to another job (job mobility) are made under these conditions of uncertainty and are based on the individuals' expectations of future costs and benefits of such a move. The actual (*ex post*) returns to their behaviour, however, will be observed only much later. They might, moreover, differ substantially from the expected (*ex ante*) returns on the basis of which the job mobility decision was made.[10]

Suggestions as to the potential returns of job mobility might be obtained from theoretical models of job mobility decisions and income determination. Models of individual job mobility decision-making like search or matching models usually depart from the standard assumption of maximisation of the present discounted value of the future income stream. They lead naturally to the formulation of stochastic dynamic optimisation problems and the notion of the option value of waiting. In the transition context, this latter option value will depend crucially on both the expected speed of wage convergence to the Western levels and the expected employment prospects in different sectors. Models of income determination, moreover, identify changing rewards to individual characteristics like education, experience or tenure as major components of the potential costs and benefits related to job mobility. Standard human capital theory e.g. predicts positive returns to both experience (general human capital) and tenure (firm-specific human capital). In the present framework of transition from a centrally planned economy to a market economy, however, returns to job mobility depend crucially on the potential transferability of general human capital and on the relative returns that jobs in a market economy (*new jobs*) award to experience and tenure acquired on jobs in a centrally planned economy (*old jobs*). Clearly, new jobs might reward individual characteristics differently from old jobs. Existing studies of the changing reward structure during transition suggest e.g. considerably higher returns to education, especially university education, on new jobs in transition economies.

Theoretical predictions of the returns to early job mobility during transition are thus am-

[9] Even after the establishment of both monetary and political union in July and October 1990, respectively, experts like Heilemann (1991) stressed the inherent unpredictability of the future transition process: "As for the long-term prospects, it is clear that there is neither a theoretical nor an empirical basis so far for a serious forecast (...)".

[10] As pointed out by Angrist (1998) in the context of analysing the returns to voluntary military service, even voluntary movers could "fail to benefit (...) if they do not accurately evaluate future costs and benefits at the time of decision."

biguous. On the one hand, one might expect early job movers to benefit from their move due to high rewards to market economy specific human capital investments and experience (i.e. tenure on a new job) as well as due to learning-by-doing effects. Moreover, they can probably better exploit the potential for job shopping at the beginning of transition.[11] Those moving already at the beginning of the transition period might further signal both higher motivation as well as higher flexibility to adapt to the new labour market situation. On the other hand, early job movers might fall behind the stayers if there is a significantly positive option value of waiting. In this case it could instead pay off to await at least a partial temporal resolution of the transition-specific uncertainty by observing the changes in reward and labour demand structures. The information obtained in the meantime could e.g. help to avoid costs of unprofitable human capital investments in further education or retraining. Initially staying on the job could further ensure benefits from the above average wage growth during the first years of transition. Finally, staying on the job could be considered a signal of high productivity.

Following the theoretical distinction between forced and voluntary mobility, returns to early job mobility, moreover, are likely to depend on the nature of the move and are thus likely to be different across individuals. Immobility is probably most beneficial to individuals with few outside opportunities but maybe detrimental to individuals with good outside opportunities and a high risk of displacement. Job mobility, on the other hand, is probably beneficial to individuals with good outside opportunities and a low risk of displacement due to firm closure, but even more beneficial to individuals with good outside opportunities and a high risk of displacement due to firm closure. Finally, individuals might realise at a later stage that it would have been beneficial to stay on their old job instead of switching to a new one, possibly because employment prospects on the new job unexpectedly turn out to be worse than on the old job or because the wage evolution on the new job is not as favourable as expected. In other words, the realised (*ex post*) returns to early job mobility can be negative.[12] There exist

[11] similar to job shopping behaviour at the beginning of the career

[12] For an analysis of the employment evolution in the period 1992-95 in old firms (already founded in the GDR before 1990) as opposed to new firms (founded after 1989), cf. Steiner et al. (1998), chapter 2, and the references cited therein. The authors present mixed evidence, showing that the relative employment performance of new firms as opposed to old ones depends, among other factors, on the respective sector. While e.g. new firms seem to do better in the service sector, the opposite seems to hold for the retail sector. It is at least questionable whether and to what extent these differences in employment prospects could have been predicted before 1992. In the sample used in the later analysis, moreover, displacement rates in the period 1993-96 are similar for early

thus strong theoretical reasons to believe that returns to (early) job mobility in transition are heterogeneous across different subpopulations. The evaluation of these returns in the case of the East German transition process clearly is an empirical problem that will be addressed in detail in chapter 4.

1.3.2 Previous studies

Despite a huge body of literature on the economics of transition in general and the evolution of earnings and inequality during transition in particular, however, only a few studies have undertaken an evaluation of the returns to job mobility in the transition from a centrally planned to a market economy.[13] In the special case of the East German transitional labour market, previous empirical studies on the returns to job mobility like Buechel and Pannenberg (1994) and Hunt (1999a), have tried to identify the returns to job mobility as coefficients on the job mobility indicators in some appropriate reduced form model. Both studies suggest that individuals moving voluntarily to a new job early in the transition process experience significant gains with respect to job stayers, in terms of wage increase and job or income satisfaction levels respectively, while forced movers do not seem to experience any gains.[14] Moreover, the initially high returns to job mobility are found to decline over the subsequent years.

While all these studies attempt to evaluate the returns to job mobility during transition, they differ considerably, however, in the definition of job mobility, the choice of the outcome variable, the model formulation, and the sample used in the empirical analysis. Not surprisingly, they also lead to different conclusions about the magnitude of returns to job mobility.

Buechel and Pannenberg (1994) evaluate the returns from switching to self-employment and from changing the employer with respect to job stayers as control group. They base their analysis on the balanced sample of all full-time employed individuals in the Eastern German

job-to-job movers (6.4% on average) and for stayers (6.2% on average).

[13] Cf. Boeri and Flinn (1999) in the context of job mobility between the public and the private sector in the Polish transition economy: "(...) returns to mobility under the economic transformation may be too low to motivate people to abandon the most protected and unionised jobs in the public sector for more risky jobs in the private sector. What appears to be lacking in the literature is an evaluation of the private costs and benefits of job mobility."

[14] In these studies, the distinction between forced and voluntary movers is based on survey information about the nature of previous job endings. While job changers who had previously been laid off or whose firm had closed down (and in Buechel and Pannenberg (1994) also those who had come to the end of a fixed-term contract) are considered as forced movers, quitters from their previous jobs are considered as voluntary movers.

subsample of the GSOEP for the years 1990 and 1991. Due to the lack of valid income information for the self-employed the authors restrict attention to subjective income and job satisfaction levels as outcome variables. The returns of mobility are considered to be identified as coefficients on the respective job mobility indicator in a random effects probit model. The authors' main finding is that voluntary job mobility in the first year of transition significantly increases both income and job satisfaction while there are no significant returns to forced job mobility.[15]

Boeri and Flinn (1997), for the case of Poland, jointly model wages and transition rates between the three labour market states private sector employment, state sector employment, and non-employment. In their search-theoretic model, individuals act according to static decision rules comparing costs and benefits of job-to-job shifts between two respective sectors. Identification of these returns to job mobility is achieved by assuming a known functional form of the underlying individual decision mechanism. Based on subsamples of two consecutive quarters each of the Polish Labour Force Survey, their estimation results suggest the lack of incentives to move from the state sector to the private sector due to low or zero returns to experience and age in the private sector as well as due to the greater dispersion of private-sector offers. According to the authors this lack of incentives explains the observed low mobility of workers across both industries and occupations in transition economies.

Finally, Hunt (1999a) analyses the effect of job mobility on year-to-year median wage growth in Eastern Germany for the years 1990-1996. She considers wage growth of job movers relative to stayers, further dividing movers into involuntary movers, namely those having experienced a job loss due to either firm closure or layoff, and voluntary ones. On the basis of estimates from quantile regressions she compares the determinants of changes in the logged gross monthly income for the two subperiods 1990-91 and 1991-96 and concludes that those individuals moving voluntarily between 1990 and 1991 experience large wage gains with respect to job stayers. Like job mobility rates, these high returns to job mobility are found to decline over the subsequent years. When interpreting the estimated coefficients on the various job mobility indicators as

[15] In their own words: "Those people who were already able to realize a turnover in the first phase of the change of the economic system do not only receive pecuniary benefits (...), but also an increase in job satisfaction. (...) Regardless of the objective quality of their jobs, the perspectives for job-movers are much better than those for people remaining in their former jobs."

returns to mobility the gains from voluntary job mobility with respect to stayers amount up to 15% for job mobility within the east and up to 65% for moves to a job in the west. Involuntary job movers are found to have a wage growth similar to job stayers.[16] Year-to-year median wage growth in the period 1991-96, on the contrary, is not found to be significantly different across the three subgroups of stayers and voluntary and involuntary movers.

1.3.3 This study

Recent theoretical contributions to optimal job mobility behaviour under uncertainty, both in general and in the special case of transition from a centrally planned economy to a market economy do suggest that individuals self-select into specific forms of job mobility behaviour on the basis of their unobservable characteristics.[17] In the presence of such self-selection, however, standard results from both simple treatment-control comparisons of average income or from simple regressions of reduced form earnings equations do in general not identify any *causal effect* of job mobility on the outcome, thus questioning the interpretability of the results obtained by Buechel and Pannenberg (1994) or Hunt (1999a).[18]

In order to estimate the causal effects of job mobility, this study will deviate from those mentioned above in several aspects. First, and most importantly, it will take explicitly into consideration the potential endogeneity problem mentioned above. Second, the outcome variables in the above studies are defined to be year-to-year wage increases or satisfaction levels before and after moving, reflecting short-term effects of job mobility. Optimal decision rules obtained from stochastic dynamic optimisation models can imply, however, temporary income decreases as part of the optimal strategy, e.g. due to retraining or further education which, in turn, might increase future employment probabilities. In order to capture both short-run and long-run effects of job mobility, this study will suggest the discounted income stream over

[16] A closer look at the determinants of wage growth further reveals that initial wage growth is highest for those at the lower end of the income distribution, in particular the less skilled, the young, and women.

[17] Cf Alexeev and Kaganovich (1998), Simmonet (1998), and Widerstedt (1998). The model developed in chapter 2 below adds to this literature by showing that such self-selection might e.g. occur if the probabilities of being laid off from the current job or of being hired on a new job vary with unobservable characteristics like ability or the access to networks.

[18] Hunt (1999a) herself, however, recognises that the "coefficients on the moving dummies should not be interpreted as the return to an exogenous move by a random worker, since voluntary and involuntary movers are likely to be unobservably different from each other and from stayers."

the whole transition period instead as outcome variable. Third, although quite standard, the distinction between voluntary and forced movers on the basis of the nature of job losses previous to a job mobility episode is not unproblematic. It is questionable to what extent displacement due to firm closure or layoff is involuntary if it was to be expected with a high probability. By the same line of reasoning, job moves without previous displacement which are actually due to a high expected probability of either firm closure or layoff should probably be considered as involuntary (*forced quits*). This study will therefore base the distinction of forced and voluntary movers on information other than self-declared reasons for the job change.

In this context, the *causal effect* of job mobility on the discounted income stream over the transition period will be defined as the difference between the *observed outcome* given the individual's actual behaviour and the *counterfactual* - or hypothetical - *outcome* in case he/she had behaved differently. In case an individual had actually changed job early in transition, the counterfactual outcome is the individual's hypothetical income stream over the transition period if he/she had stayed on his/her job instead. Accordingly, in case the individual had stayed on his/her job early in transition, the counterfactual outcome is the income stream he/she would have received had he/she decided to change job instead. While not observable at the individual level, these causal effects can be inferred in a *quasi-experimental* situation in which individuals - conditional on their observed characteristics - are selected randomly and independently of their potential outcomes into different states, leading to significant differences in job mobility behaviour. For a given quasi-experiment, three types of individuals can be distinguished with regard to their job mobility behaviour: first, those who change job early in transition independently of the underlying selection mechanism (the so-called *always-takers*); second, those who - also independently of the selection mechanism - do not change job early in transition (the *never-takers*); and finally, those who only change job early in transition if they are selected into some group with high probability of job change but who would not change job if they were selected into some other group (the so-called *compliers*). The chosen approach allows estimation of the causal effects of job mobility for this latter group of compliers only, and it is this implicit dependence of the estimated causal effects on the respective quasi-experimental setting that will finally allow a distinction between *forced* and *voluntary* job movers.[19]

[19]For a detailed description of the quasi-experimental approach to causal inference and a formal definition of

The empirical evidence on returns to job mobility obtained on the basis of this approac and presented in this thesis might help not only to judge the adequacy of mobility incentive provided during transition and their potential impact on the speed of labour reallocation and th human capital stock of the economy in the medium to long run but also to identify 'winners and 'losers' during transition. It could hence prove useful in both understanding individua opinions about the transition process and designing future labour market policies and mobilit incentives.

1.4 References

Alexeev, M. and M. Kaganovich (1998), Returns to human capital under uncertain reform: Good guys finish last, Journal of Economic Behavior and Organization, 37, 53-70

Angrist, J.D. (1998), Estimating the labor market impact of voluntary military service using social security data on military applicants, Econometrica, 66(2), 249-288

Bartel, A. and G. Borjas (1981), Wage growth and job turnover, in: S. Rosen (ed.), Studies in Labor Markets, University of Chicago Press, Chicago

Beblo, M., I.L. Collier and T. Knaus (2001a), The unification bonus (malus) in postwall Eastern Germany, ZEW Discussion Paper No. 01-29, Mannheim

Beblo, M., I.L. Collier and T. Knaus (2001b), Hohe Gewinne aus der deutschen Wiedervereiningung, ZEW aktuell, 9 September 2001, III-IV

Bird, E., J. Schwarze, and G. Wagner (1994), Wage effects of the move towards free markets in East Germany, Industrial and Labor Relations Review, 47(3), 390-400

Boeri, T. and C.J. Flinn (1999), Returns to mobility in the transition to a market economy, CEPR Discussion Paper No. 2098, London

Brainerd, Elizabeth (1998), Winners and losers in Russia's economic transition, American Economic Review, 88(5), 1094-1116

Buechel, F. and M. Pannenberg (1994), Welfare effects of labour mobility in Eastern Germany, A comparison between benefits from switching into self-employment and gains from job turnover, in: J. Schwarze, F. Buttler and G. Wagner (eds), Labour Market Dynamics in Present

the underlying concepts, cf. chapter 3 and the references cited therein.

Day Germany, Campus, Frankfurt/Main, 91-111

Chase, R.S. (1998), Markets for communist human capital: Returns to education and experience in post-communist Czech Republic and Slovakia, Industrial and Labor Reations Review, 51(3), 401-423

Flanagan, R.J. (1998), Were communists good human capitalists?, The case of the Czech Republic, Labour Economics, 5, 295-312

Franz, W. and V. Steiner (1999), Wages in the East German transition process, Facts and explanations, German Economic Review, 1(3), 241-269

Heilemann, U. (1991), The economics of German unification: A first appraisal, Konjunkturpolitik, 37(3), 127-155

Hunt, J. (1997), The transition in East Germany: When is a ten point fall in the gender wage gap bad news?, NBER Working Paper No. 6167

Hunt, J. (1999a), Post-unification wage growth in East Germany, NBER Working Paper No. 6878

Hunt, J. (1999b), Determinants of non-employment and unemployment durations in East Germany, NBER Working Paper No. 7128

Krueger, A.B. and J.S. Pischke (1995), A comparative analysis of East and West German labor markets before and after unification, in: R.B. Freeman and F. Katz (eds), Differences and Changes in Wage Structures, Chicago University Press, Chicago

Mincer, Jacob (1986), Wage changes in job changes, Research in Labor Economics, JAI Press, Volume 8, Part A, 171-197

Munich, D., J. Svejnar, and K. Terrell (2000), Returns to human capital under the communist wage grid and during the transition to a market economy, IZA Discussion Paper No. 122, Institute for the Study of Labor, Bonn

Orazem, P.F. and M. Vodopivec (1995), Winners and losers in the transition: Returns to education, experience, and gender in Slovenia, World Bank Economic Review, 9(2), 201-230

Orazem, P.F. and M. Vodopivec (1998), Unemployment in Eastern Europe, and the value of human capital in the transition to market: Evidence from Slovenia, European Economic Review, 41, 893-903

Roy, A.D. (1951), Some thoughts on the distribution of earnings, Oxford Economic Papers, 3, 135-146

Rutkowski, J. (1994a), Wage determination in late socialism: The case of Poland, Economics of Planning, 27, 135-164

Rutkowski, J. (1994b), Changes in wage structure and in returns to education during economic transition: The case of Poland, Center for International Studies, Princeton

Rutkowski, J. (1996), High skills pay off: The changing wage structure during economic transition in Poland, Economics of Transition, 4(1), 89-112

Rutkowski, J. (1997), Low wage employment in transitional economies of Central and Eastern Europe, MOST, 7, 105-130

Simmonet, V. (1998), The role of unobservables in the measurement of return to job mobility: Evidence from Germany and the United States, Lamia, Université Paris I Panthéon-Sorbonne, mimeo

Steiner, V. et al. (eds) (1998), Strukturanalyse der Arbeitsmarktentwicklung in den neuen Bundeslaendern, ZEW-Wirtschaftsanalysen, Band 30, Nomos-Verlag, Baden-Baden

Steiner, V. and P. Puhani (1997), Economic restructuring, the value of human capital, and the distribution of hourly wages in Eastern Germany 1990-94, Vierteljahreshefte zur Wirtschaftsforschung, 197-210

Svejnar, J. (1998), Labor markets in the transitional Central and East European economies, in: O. Ashenfelter and D. Card (eds), Handbook of labor economics, Vol. 3, North-Holland, Amsterdam

Vecernik, J. (1995), Changing earnings distribution in the Czech Republic: Survey evidence from 1988-1994, Economics of Transition, 3(3), 335-371

Widerstedt, B. (1998), Moving or staying?, Job mobility as a sorting process, Ph.D. thesis, Umea University

1.5 Appendix

Table 1.7: Labour market status 1990-97: East Germany (full sample)

LM status	1990	1991	1992	1993	1994	1995	1996	1997
employed								
total	81.3	71.5	65.0	62.2	60.7	61.1	58.8	57.6
full-time	68.1	62.2	57.5	54.9	53.1	53.0	50.3	49.8
part-time	10.6	6.3	4.8	4.8	5.6	6.6	6.9	6.3
(re)training	2.4	2.6	2.2	2.0	1.3	0.8	0.7	0.8
marginally	0.2	0.5	0.5	0.6	0.8	0.8	0.8	0.8
not employed								
total	18.7	28.5	35.1	37.8	39.3	38.9	41.3	42.4
unemployed		7.9	12.2	13.2	13.6	11.7	12.5	12.4
out of LF	18.7	20.6	22.8	24.6	25.7	27.2	28.8	29.9

NOTES: in % of full sample (N=2696); labour market status as of time of survey
SOURCE: GSOEP, waves 7-14, subsample C

Table 1.8: Labour market status 1990-97: West Germany (full sample)

LM status	1990	1991	1992	1993	1994	1995	1996	1997
employed								
total	63.2	64.5	63.0	61.7	59.6	58.6	57.9	56.2
full-time	49.3	49.7	49.2	48.7	47.3	46.2	44.8	43.7
part-time	8.5	9.0	9.3	9.3	8.8	9.0	9.8	9.8
(re)training	3.4	3.2	2.5	1.8	1.0	0.7	0.5	0.3
marginally	2.0	2.6	1.9	2.0	2.5	2.6	2.8	2.5
not employed								
total	36.8	35.5	37.1	38.3	40.4	41.4	42.1	43.8
unemployed	3.2	2.6	3.2	3.9	5.1	5.1	5.1	5.5
out of LF	33.6	32.9	33.8	34.4	35.2	36.3	37.1	38.2

NOTES: in % of full sample (N=6187); labour market status as of time of survey
SOURCE: GSOEP, waves 7-14, subsamples A and B

Table 1.9: Labour market transitions: East Germany (full sample)

LM transitions from ...	East 90/91	91/92	92/93	93/94	94/95	95/96	96/97
employment to:							
employment	84.0	84.4	86.1	89.8	90.9	88.2	89.8
unemployment	8.1	10.5	9.6	7.9	6.8	9.2	7.5
out of the labour force	7.9	5.1	4.3	2.4	2.3	2.6	2.8
unemployment to:							
employment		34.4	37.3	31.2	35.4	31.0	30.4
unemployment		47.6	46.7	53.4	47.4	50.3	53.6
out of the labour force		17.9	16.1	15.5	17.2	18.7	16.1
out of the LF to:							
employment	17.3	9.0	7.8	3.0	4.3	4.4	3.9
unemployment	6.8	4.9	5.5	6.8	4.5	3.4	4.8
out of the labour force	76.0	86.2	86.7	90.2	91.2	92.2	91.4

LM transitions from ...	West 90/91	91/92	92/93	93/94	94/95	95/96	96/97
employment to:							
employment	93.3	92.2	91.6	90.9	91.6	91.6	91.5
unemployment	1.4	2.1	2.7	4.3	3.3	3.0	3.6
out of the labour force	5.3	5.7	5.7	4.8	5.1	5.4	4.9
unemployment to:							
employment	34.2	22.1	22.5	24.9	30.8	26.7	22.4
unemployment	36.2	46.6	50.0	53.9	51.9	54.1	55.9
out of the labour force	29.6	31.3	27.5	21.2	17.3	19.2	21.7
out of the LF to:							
employment	13.1	8.9	9.9	7.3	6.7	7.8	5.8
unemployment	1.8	1.9	1.8	1.1	1.5	1.5	1.7
out of the labour force	85.1	89.2	88.5	91.5	91.8	90.7	92.6

NOTES: one-year transition probabilities between labour market states; labour market states as of time of survey; SOURCE: GSOEP public-use file, waves 7-14, subsamples A and B (West Germany; N=6187 and)subsample C (East Germany; N=2696)

Chapter 2

A theoretical model of individual job mobility decisions during transition

2.1 Introduction

This chapter[1] derives a 3-period theoretical model of individual job mobility decisions in the transition process from a centrally planned economy to a market economy. The model builds on Alexeev and Kaganovich (1998), adapting their model to the East German case.[2] It demonstrates, first, the possibility of self-selection into early job mobility on the basis of unobservable characteristics, and second, the possibility of divergence between expected and realised returns to early job mobility in transition. Given an appropriate choice of the underlying assignment mechanism, the model is, moreover, easily amenable to an interpretation in a quasi-experimental framework which will be set out in the following chapter. Set in such a framework, it allows the characterisation of those subgroups - already mentioned in the introduction - which are of

[1] A shortened version of this chapter was published as: Siebern, F. (2000), A Theoretical Model of Job Mobility Decisions During Transition, in: W. Welfe and P. Wdowinski (eds), Modelling Economies in Transition, Proceedings of the Fourth International Conference, December 1-4, 1999, Rydzyna, Poland

[2] Alexeev and Kaganovich (1998) is one of the few theoretical models which analyse the potential effect of ability on job mobility behaviour during transition. Following Jovanovic and Nyarko (1996) and building on the concept of leapfrogging in the industrial organisation literature they demonstrate that uncertainty about the structure of future labour demand can give rise to adverse selection of the less productive individuals into the upcoming new occupations. According to their model, leapfrogging among individuals with respect to job mobility "systematically rewards the less able agents" since "the most able individuals may be systematically late to switch to the new 'market' professions (...)". (Alexeev and Kaganovich (1998))

central interest in the treatment effect literature: *always-takers*, *never-takers*, and *compliers*. The model thus provides a framework for the identification, estimation and interpretation of the *causal effects* of job mobility on income or employment probabilities throughout transition.

After presenting the model and its implications for individual job mobility decisions and returns to job mobility in section 2.2, section 2.3 clarifies the role of unobservable individual characteristics such as ability or network access in job mobility decision making and self-selection. Section 2.4 restates the model in a quasi-experimental framework and characterises the subgroups of interest - that of *compliers* in particular - in terms of their unobservable characteristics. It finally suggests an instrumental variables approach for identification and estimation of the returns to early job mobility during transition. The model and its implications are illustrated throughout the chapter by way of numerical examples. Section 2.5 concludes.

2.2 The model

The main feature of the transitional labour market retained in the model is the structural change in labour demand through, first, the unanticipated emergence of new market economy jobs, and, second, the different evolution of both remuneration patterns and employment prospects on the previously existent jobs and the upcoming ones.

2.2.1 The general setup

Individuals are assumed to be unobservably heterogeneous in both their ability and their access to networks.[3] Let $\alpha \in [0,1]$ denote an individual's ability level and $\nu \in [0,1]$ a measure of his access to networks. Assume further a constant discount factor (rate of time preference) $\delta \in (0,1]$.

In $t = 0$, individuals in the centrally planned economy decide on their investment in education (human capital) and on their career (occupation, employment sector and employer),

[3]The consideration of two types of unobservable characteristics is based on the idea that "ability" alone might not fully capture the particularities of a labour market in transition. In addition, "networks" that had been formed in the centrally planned economy might have been of importance for employment prospects at least in an early stage of the transition process. The term "networks" here is meant to denote any private or institutional relationships stemming from the previous economic and social system which might have influenced - positively or negatively - an individual's labour market outcomes throughout transition.

summarised by the level of human capital acquired, $H = h(e;\alpha,\nu)$ where H is assumed to be a function of a choice variable effort, $e \in [0,1]$, and the unobservable characteristics (α,ν).[4]

At this stage, individuals do not anticipate any later structural change in the economic system and hence do not perceive any alternative to jobs in the centrally planned economy (denoted *P-jobs* in the sequel). Income from such jobs is given by $y = y_P(H;\alpha,\nu)$ as a function of both education and unobservable characteristics, and learning-on-the job (tenure) effects lead to a proportional period-to-period increase of income by a factor $\theta_P \geq 0$.

Given their ability level and their access to networks, individuals choose an effort level e^* (implying an amount of leisure of $(1-e^*)$) in order to maximise the present discounted value of expected utility from future income streams, i.e.

$$e^* = \arg\max_{e\in[0,1]} \left\{ v\left[1-e\right] + \delta u\left[y_P(H;\alpha,\nu)\right] + \delta^2 u\left[(1+\theta_P)\, y_P(H;\alpha,\nu)\right] \right\}, \tag{2.1}$$

where $v[.]$ and $u[.]$ denote indirect utility functions and where $H = h(e;\alpha,\nu)$ as defined above.[5] The utility derived from unemployment is normalised to zero. The optimal human capital level is given by $H^* = h(e^*;\alpha,\nu)$ and resulting income by $y = y_P^* = y_P(H^*;\alpha,\nu)$.[6]

At $t = 1$, new market economy jobs (denoted *M-jobs* in the sequel) emerge due to an unexpected exogenous shock to the centrally planned economy. These new jobs pay income $y = y_M(H^*;\alpha,\nu)$ and offer a learning-by-doing (tenure) premium of $\theta_M \geq 0$. Assume for simplicity $y_M(H^*;\alpha,\nu) = \kappa_y y_P(H^*;\alpha,\nu)$, $\kappa_y \geq 0$, and $\theta_M = \kappa_\theta \theta_P$, where the parameters κ_y and κ_θ capture the differences in pay and tenure premia between the two types of jobs.[7]

[4] There exists at least anecdotal evidence stressing the importance of - individual or family - network access for educational attainment and/or career choice.

[5] For pure convenience, v and u will be assumed linear in the sequel. Although far from being innocuous, this assumption will not change results qualitatively. For further details regarding the existence and characterisations of solutions to the above maximisation problem under standard monotonicity and functional form assumptions, see Alexeev and Kaganovich (1998).

[6] For notational simplicity H^* is assumed to subsume optimal levels of both experience in the labour market and tenure on the job as of $t = 0$. Hence, given the dependence of y on H^*, these characteristics will be reflected in the income level y.

[7] This assumption, again, is not innocuous at all. It ensures in fact the independence of job mobility decisions of unobservable characteristics by assuming that the reward structures on P-jobs and M-jobs are not systematically different. They are assumed instead to differ only by a proportionality factor k_y which itself is assumed independent of (α,ν)..This rather restrictive assumption will be relaxed in the following section. This parameterisation of $y_M(H^*;\alpha,\nu)$, however, allows flexibility in modelling the transitional change in the structure of

In addition, at $t = 1$, subsequent reforms of the labour market (restructuring and reallocation of labour across occupations and industries) are announced, without there being any information about the details of these reforms. These reforms might generally act on the probabilities of being hired on and fired from a job as well as on the payoffs and/or the learning-by-doing premia. Assume for simplicity that these reforms act exclusively on the probabilities of being hired on and fired from a job (and not on the payoffs and/or learning-by-doing premia). Assume in particular that, at $t = 1, 2$, individuals are hired on a job with probability $h_P(t) \in [0,1]$ and $h_M(t) \in [0,1]$, respectively, and they are fired from their job with probability $f_P(t) \in [0,1]$ and $f_M(t) \in [0,1]$, respectively. The hiring and firing probabilities for some period t are assumed known to the individuals in that period. At $t = 1$ individuals form expectations about future hiring and firing probabilities at $t = 2$. These expectations are denoted by $E_1[h_P(2)]$, $E_1[h_M(2)]$ and $E_1[f_P(2)]$, $E_1[f_M(2)]$, respectively. Assume finally, for simplicity, $h_P(1) = h_P(2) = 0$, i.e. no individual is hired on another *P-job* after labour restructuring has been announced, and $f_M(1) = 0$, i.e. no individual is laid off immediately from a *M-job* to which he has changed at $t = 1$.[8]

2.2.2 Job mobility decisions during transition

Conditional on H^* and their career choice at $t = 0$, individuals decide on whether to switch to the newly emerging *M-jobs* either at $t = 1$ (*early*) or at $t = 2$ (*late*), or they can decide to stay on their current job throughout both periods. Let $\tau \in \{1, 2, \infty\}$ denote the time at which an individual switches from a *P-job* to an *M-job*, with $\tau \equiv \infty$ for job stayers, and let $d = (d_1, d_2) \in \{0,1\}^2$ denote the sequence of job mobility decisions for periods 1 and 2, with $d_i = 1$ in case of job mobility at $t = i$ and $d_i = 0$ otherwise.

---

labour demand. A demotion e.g. can be represented in the model by $k_y < 1$, and the absence of any systematic difference in the remuneration patterns of *P-jobs* and *M-jobs* can be expressed as $k_y = k_\theta = 1$.

[8] The assumption that the announced labour market reforms influence the probabilities of being fired from the current job and/or hired on a new job is made in order to adapt Alexeev and Kaganovich's (1998) model to the East German case. In their model, a reform is announced, but it is uncertain whether this reform will take place or not. If the reform takes place, moreover, existing (*P-*) jobs are assumed to disappear completely. Individuals form expectations about the realisation of the reform and act upon these expectations. In the East German case, however, there was surely no doubt as to whether there would be any reforms. However, uncertainty existed as to both their implementation and their timing, the Economic, Monetary, and Social Union (EMSU) probably being the most famous example. And even after the decisions regarding its timing and implementation had been taken, considerable uncertainty prevailed as to its consequences, especially with respect to the labour market, as well as with respect to the potential design of future labour market policies mitigating these consequences.

Depending on the individual's job mobility decisions, income in period 1 is given by

$$y_1 = \begin{cases} h_M(1) y_M(H^*; \alpha, \nu) & \text{if } \tau = 1 \\ (1 - f_P(1)) y_P(H^*; \alpha, \nu) & \phantom{\text{if }} \tau > 1 \end{cases}, \tag{2.2}$$

and in period 2 by

$$y_2 = \begin{cases} (1 - f_M(2))\,(1 + \theta_M)\, y_M(H^*; \alpha, \nu) & \tau = 1 \\ h_M(2) y_M(H^*; \alpha, \nu) & \text{if } \tau = 2 \\ (1 - f_P(2))\,(1 + \theta_P)\, y_P(H^*; \alpha, \nu) & \tau > 2 \end{cases}. \tag{2.3}$$

The optimal job mobility decisions are now obtained easily by backward induction. Individuals base their decisions on the calculation of the expected gains from switching to a new job. In period 2, these are given by

$$\Delta_2 = \begin{cases} h_M(2) y_M(H^*; \alpha, \nu) - (1 - f_M(2))\,(1 + \theta_M)\, y_M(H^*; \alpha, \nu) & \text{if } \tau = 1 \\ h_M(2) y_M(H^*; \alpha, \nu) - (1 - f_P(2))\,(1 + \theta_P)\, y_P(H^*; \alpha, \nu) & \phantom{\text{if }} \tau > 1 \end{cases}, \tag{2.4}$$

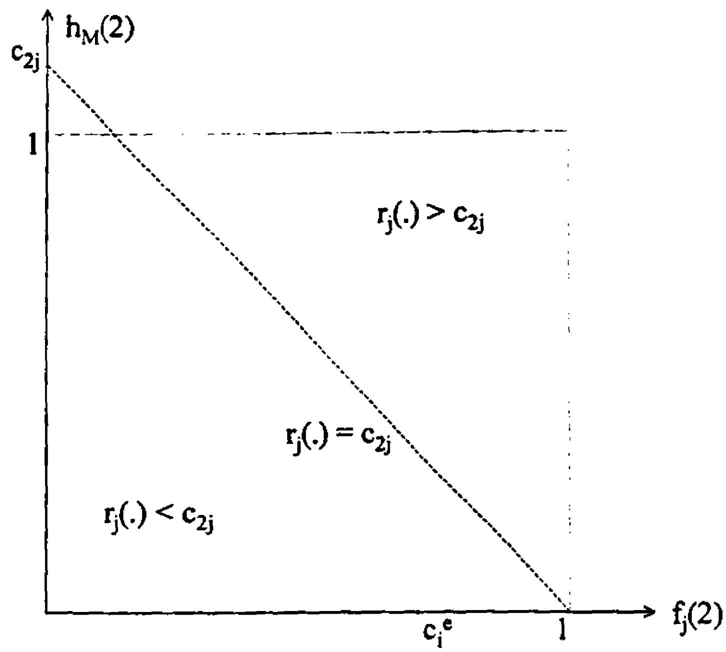
where Δ_2 is a function of the parameters $\kappa_y, \kappa_\theta, \theta_P, h_M(2), f_P(2), f_M(2)$ which are assumed known to the individual at this stage. Conditional on his decision at t=1, an individual will change to a new job in t=2 iff $\Delta_2 > 0$, giving rise to

Lemma 1 *Individuals who changed job in $t = 1$ (in which case $\tau = 1$) switch job again at $t = 2$ iff $\frac{h_M(2)}{1-f_M(2)} > 1 + \kappa_\theta \theta_P$ in case $f_M(2) \in [0,1)$ or iff $h_M(2) > 0$ in case $f_M(2) = 1$.*

Individuals who have not changed job before (in which case $\tau > 1$) switch to a new job at $t = 2$ iff $\frac{h_M(2)}{1-f_P(2)} > \frac{1+\theta_P}{\kappa_y}$ in case $f_P(2) \in [0,1)$ or iff $h_M(2) > 0$ in case $f_P(2) = 1$.

Note that decisions at $t = 2$ are purely deterministic due to the assumption of known period 2 hiring and firing probabilities as of $t = 2$. Moreover, as mentioned before, individual job mobility decisions are independent of unobservable characteristics due to the assumption of proportionality between y_P and y_M and the independence of k_y, k_θ, and θ_P as well as

Figure 2-1: Job mobility decisions at t=2 as a function of period 2 hiring and firing probabilities



$h_M(2), f_P(2), f_M(2)$ of (α, ν). These seemingly rather restrictive assumptions have been made on the basis of the following observations: first, a relatively low degree of income inequality at the beginning of transition together with a relatively high degree of centralised wage bargaining (k_y independent of (α, ν)); second, the unaltered importance of centralised wage bargaining during the transition process (k_θ and θ_P independent of (α, ν)); third, the decreasing influence of networks stemming from the previous system throughout the transition process (f_j and h_j, j=P,M, independent of ν); and fourth, the increasingly similar influence of ability on firing as well as retention probabilities on both types of jobs (f_j and h_j, j=P,M, independent of α).

Letting $r_j(h_M(2), f_j(2)) \equiv \frac{h_M(2)}{1-f_j(2)}$ for $j = P, M$ be a function of the hiring and firing probabilities in period 2, the above decision rule can be illustrated graphically by recognising that individuals switch to a new job late iff $r_j(h_M(2), f_j(2)) > c_{2j}$ where $c_{2M} = 1 + \kappa_\theta \theta_P$ and $c_{2P} = \frac{1+\theta_P}{\kappa_y}$: Given their previous job mobility decision and given the differences in pay and tenure premia, individuals will decide to switch job at $t = 2$ iff the probability of being hired on a new job exceeds the retention probability on the current job by at least the factor c_{2j} as illustrated in Figure 2-1.

According to the above Lemma, four cases (denoted i=1,...,4) have to be distinguished:

$$i = \begin{cases} 1 \text{ if } \frac{h_M(2)}{1-f_P(2)} < \frac{1+\theta_P}{\kappa_y} \text{ and } \frac{h_M(2)}{1-f_M(2)} < 1+\kappa_\theta\theta_P \\ 2 \text{ if } \frac{h_M(2)}{1-f_P(2)} > \frac{1+\theta_P}{\kappa_y} \text{ and } \frac{h_M(2)}{1-f_M(2)} < 1+\kappa_\theta\theta_P \\ 3 \text{ if } \frac{h_M(2)}{1-f_P(2)} < \frac{1+\theta_P}{\kappa_y} \text{ and } \frac{h_M(2)}{1-f_M(2)} > 1+\kappa_\theta\theta_P \\ 4 \text{ if } \frac{h_M(2)}{1-f_P(2)} > \frac{1+\theta_P}{\kappa_y} \text{ and } \frac{h_M(2)}{1-f_M(2)} > 1+\kappa_\theta\theta_P \end{cases} \tag{2.5}$$

with choice sets at $t = 2$ given by:

$$d = (d_1, d_2) \in \begin{cases} C_1 = \{(0,0),(1,0)\} \text{ for } i = 1 \\ C_2 = \{(0,1),(1,0)\} \text{ for } i = 2 \\ C_3 = \{(0,0),(1,1)\} \text{ for } i = 3 \\ C_4 = \{(0,1),(1,1)\} \text{ for } i = 4 \end{cases} \tag{2.6}$$

In the sequel, I will focus on case 1 in which both ratios, $\frac{h_M(2)}{1-f_P(2)}$ and $\frac{h_M(2)}{1-f_M(2)}$, are bounded above by $\frac{1+\theta_P}{\kappa_y}$ and $1+\kappa_\theta\theta_P$, respectively, and in which individuals either switch job in the first period or they stay with their initial job throughout both periods. The results and derivations for the other cases are deferred to the appendix.

The expected present discounted value of future income streams, $Y^e_{(i,j)} = E_1[y_1 + \delta y_2 \,|d]$, as a function of d at t=1 is given by

$$Y^e_{(i,j)} = \begin{cases} \{(1-f_P(1)) + \delta E_1[(1+\theta_P)(1-f_P(2))]\}\, y_P & \text{if } d = (0,0) \\ \{h_M(1) + \delta E_1[(1+\kappa_\theta\theta_P)(1-f_M(2))]\}\, \kappa_y y_P & \phantom{\text{if }} d = (1,0) \end{cases} \tag{2.7}$$

where $E_1[f_P(2)]$ and $E_1[f_M(2)]$ denote the expected layoff probabilities on the respective type of job as of period 1. Individual decisions in period 1 then are described by

Lemma 2 *Define*

$$s(\kappa_y, h_M(1), f_P(1)) \equiv \kappa_y h_M(1) + f_P(1) \tag{2.8}$$

and

$$c_1^e \equiv 1 + \delta\psi_1^e\left(\kappa_y, \kappa_\theta, \theta_P, f_P\left(2\right), f_M\left(2\right), h_M(2)\right) \tag{2.9}$$

where

$$\psi_1^e\left(.\right) = E_1\left[(1+\theta_p)(1-f_P(2)) - \kappa_y(1+\kappa_\theta\theta_P)(1-f_M(2))\right] \tag{2.10}$$

is a function of the expected differences in both pay and learning-by-doing (tenure) effects between P- and M-jobs, and in the layoff probabilities and hiring probabilities, respectively, on the two job types in the second period.

Then, individuals switch to a new job early iff

$$s(\kappa_y, h_M\left(1\right), f_P\left(1\right)) > c_1^e. \tag{2.11}$$

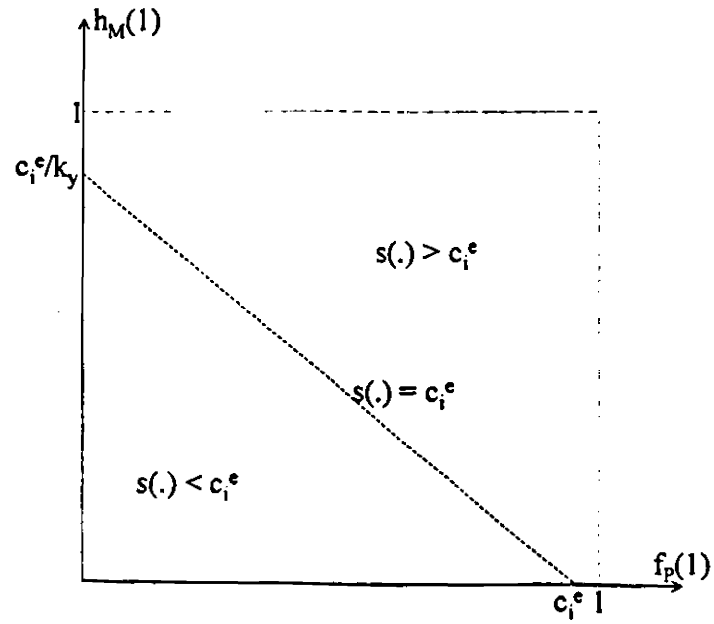
Hence, according to Lemma 2, for given parameters k_y, k_θ, and θ_P and given expectations as of $t = 1$ of the period 2 hiring and firing probabilities, $h_M(2), f_P(2)$, and $f_M(2)$, it is individuals with either a high layoff probability or a high probability of being hired on a new job (or both) who change job early in transition. This is illustrated in figure 2-2.

2.2.3 Comparative statics

This section sheds some light on the individual decisions for different situations of the transitional labour market as described by the values for the parameters $\kappa_y, \kappa_\theta, \theta_p, E_1\left[f_P(2)\right], E_1\left[f_M(2)\right]$. Recalling the above definition of the 'critical value'

$$c_1^e \equiv 1 + \delta E_1\left[(1+\theta_p)(1-f_P(2)) - \kappa_y(1+\kappa_\theta\theta_P)(1-f_M(2))\right],$$

Figure 2-2: Job mobility decisions at t=1 as a function of period 1 hiring and firing probabilities



it follows that

$$
\begin{aligned}
\frac{\partial c_1^e}{\partial \kappa_y} &= -\delta E_1\left[(1+\kappa_\theta \theta_p)(1-f_M(2))\right] \\
\frac{\partial c_1^e}{\partial \kappa_\theta} &= -\delta E_1\left[\kappa_y \theta_p (1-f_M(2))\right] \\
\frac{\partial c_1^e}{\partial \theta_P} &= \delta E_1\left[(1-f_P(2)) - \kappa_y \kappa_\theta (1-f_M(2))\right] \\
\frac{\partial c_1^e}{\partial \delta} &= E_1\left[(1+\theta_p)(1-f_P(2)) - \kappa_y (1+\kappa_\theta \theta_P)(1-f_M(2))\right] \\
\frac{\partial c_1^e}{\partial E_1\left[f_P(2)\right]} &= -\delta(1+E_1\left[\theta_p\right]) \\
\frac{\partial c_1^e}{\partial E_1\left[f_M(2)\right]} &= \delta E_1\left[\kappa_y (1+\kappa_\theta \theta_P)\right]
\end{aligned}
$$

Consequently, ceteris paribus, c_1^e is shifted down by increases in κ_y, κ_θ or $E_1\left[f_P(2)\right]$ and shifted up by increases in $E_1\left[f_M(2)\right]$, while increases in θ_p and δ further shift c_1^e down (up) if

$$E_1\left[(1-f_P(2)) - \kappa_y \kappa_\theta (1-f_M(2))\right] > 0 \ (< 0)$$

and if

$$E_1\left[(1+\theta_p)(1-f_P(2))-\kappa_y(1+\kappa_\theta\theta_P)(1-f_M(2))\right]<0\ (>0),$$

respectively. Furthermore, $\frac{\partial s}{\partial \kappa_y}=h_M(1)$, while $\frac{\partial s}{\partial \kappa_\theta}=\frac{\partial s}{\partial \theta_P}=\frac{\partial s}{\partial \delta}=\frac{\partial s}{\partial E_1[f_P(2)]}=\frac{\partial s}{\partial E_1[f_M(2)]}=0$.

The effects of changes in c_1^e on the individual's job mobility decisions as a function of period 1 hiring and firing functions then follow from Lemma 2: Individuals are more likely to switch job early the bigger the wage differential or the difference in learning-by-doing effects between both types of jobs, and the bigger the expected probability of being laid off from their actual job in the future, while they are less likely to switch to a new job early the bigger the expected probability of being laid off from such a job in the future. The effects of the actual tenure premia on *P-jobs* and the discount factor are ambiguous and do depend on some weighted difference in the expected period 2 firing and hiring probabilities. Finally, the larger the wage differential between new and old jobs, the more weight in decision-making is attributed to hiring probabilities relative to firing probabilities.

2.2.4 The returns to early job mobility

The above discussion of the individual job mobility decisions has shown that the returns to early job mobility do depend on the individual-specific hiring and firing probabilities throughout transition. Two subgroups expect to benefit most from switching to a new career path early: those with a particularly high probability of being fired from their current job and those with a rather high probability of being hired on a new job. Under the above assumptions, other individuals prefer to stay with their job at the beginning of transition.

At t=2, the uncertainty as to future structure of labour demand (i.e. period 2 hiring and firing probabilities) will have been resolved, and the counterfactual returns to the individual job mobility decision as of t=1 can be calculated by comparing the actual income obtained, given the job mobility decision, with the potential income in case the alternative decision had been taken. Then, from Lemma 2 follows immediately

Corollary 3 *The expected (ex ante) returns to early job mobility are given by*

$$\Delta_1^e = \{s(\kappa_y, h_M(1), f_P(1)) - c_1^e\} y_P \tag{2.12}$$
$$= \{\kappa_y h_M(1) + f_P(1) - c_1^e\} y,$$

while the realised (ex post) returns to early job mobility are given by

$$\Delta_1 = \{s(\kappa_y, h_M(1), f_P(1)) - c_1\} y_P \tag{2.13}$$
$$= \{\kappa_y h_M(1) + f_P(1) - c_1\} y_P$$
$$= \{\kappa_y h_M(1) + f_P(1) - 1 - \delta\psi_1(\kappa_y, \kappa_\theta, \theta_P, f_P(2), f_M(2), h_M(2))\} y_P,$$

where

$$c_1 \equiv 1 + \delta\psi_1(\kappa_y, \kappa_\theta, \theta_P, f_P(2), f_M(2), h_M(2)) \tag{2.14}$$

and

$$\psi_1(.) = (1+\theta_p)(1 - f_P(2)) - \kappa_y(1 + \kappa_\theta\theta_P)(1 - f_M(2)) \tag{2.15}$$

denote the realisations of c_1^e and $\psi_1^e(.)$ in period 2, respectively.

Consequently, the difference between realised and expected returns to early job mobility is given by

$$\Delta_1 - \Delta_1^e = \{c_1^e - c_1\} y_P \tag{2.16}$$
$$= \{\psi_1^e - \psi_1\} y_P$$
$$= \{(1+\theta_p)e(f_P(2)) - \kappa_y(1+\kappa_\theta\theta_P)e(f_M(2))\} y_P$$

where

$$\begin{aligned} e(f_P(2)) &\equiv E_1[1 - f_P(2)] - (1 - f_P(2)) = f_P(2) - E_1[f_P(2)] \\ e(f_M(2)) &\equiv E_1[1 - f_M(2)] - (1 - f_M(2)) = f_M(2) - E_1[f_M(2)] \end{aligned} \tag{2.17}$$

denote the individual's 'expectation errors' with respect to the hiring and firing probabilities prevailing in the second period. In case $e(f_{\cdot}(2)) > 0$*, the individual actually underestimates the respective firing probability in period 2 and vice versa.*

The difference of realised and expected returns to early job mobility is hence seen to depend on the difference of the individual's weighted 'expectation errors' with respect to the hiring and firing probabilities prevailing in the second period with the weights given by some function of the parameters that characterise the job-specific remuneration pattern. Among the potentially long list of reasons for such 'expectation errors' one might mention, first, the in many cases unexpectedly bad performance of new firms that were established only after EMSU, and second, the sometimes unexpectedly positive performance of old firms following e.g. a management buyout or a takeover despite their being hit initially by a negative unification shock.[9] Another reason might be the unexpectedly low growth rates and the relatively poor overall economic situation in Eastern Germany from 1993 onwards.

Given this potential for individual misperceptions of the future structure of labour demand, let alone that of remuneration and tenure premia,[10] the conditions under which individuals would incur (counterfactual) losses (or gains) from their actual behaviour - i.e. realise returns from their actual behaviour that are lower (resp. higher) than the hypothetical returns that they would have realised had they taken a different decision - are summarised in the following

Corollary 4 *Given* κ_y *and the period 1 hiring and firing probabilities* $h_M(1)$ *and* $f_P(1)$*, and letting* $s \equiv s(\kappa_y, h_M(1), f_P(1)) = \kappa_y h_M(1) + f_P(1)$*, counterfactual losses from mobility would be incurred if* $c_1 > s > c_1^e$*, whereas* $c_1 < s < c_1^e$ *would imply counterfactual losses from staying*

[9] For some empirical evidence on this issue, cf. Steiner et al. (1998).who, in an analysis of employment evolution in 'old GDR-firms' as opposed to 'newly (i.e. after 1989) founded firms', actually do find mixed evidence, depending on the sector.

[10] As mentioned above, more general results could be obtained from allowing for uncertainty as to the parameters reflecting these differences in the remuneration structure k_y, k_θ,and θ_P, in which case the difference between realised and expected returns to early job mobility would depend additionally on the potential 'expectation errors' as regards the future remuneration structure on the two types of jobs.

on the job at $t = 1$. More generally:

$$
\begin{array}{lcll}
\Delta_1^e > \Delta_1 > 0 & \Leftrightarrow & s > c_1^e > c_1 & \textit{(larger than expected) gains from moving} \\
\Delta_1 > \Delta_1^e > 0 & \Leftrightarrow & s > c_1 > c_1^e & \textit{(lower than expected) gains from moving} \\
\Delta_1^e > 0 > \Delta_1 & \Leftrightarrow & c_1 > s > c_1^e & \textit{counterfactual losses from moving} \\
\Delta_1 > 0 > \Delta_1^e & \Leftrightarrow & c_1^e > s > c_1 & \textit{counterfactual losses from staying} \\
0 > \Delta_1 > \Delta_1^e & \Leftrightarrow & c_1^e > c_1 > s & \textit{(lower than expected) gains from staying} \\
0 > \Delta_1^e > \Delta_1 & \Leftrightarrow & c_1 > c_1^e > s & \textit{(larger than expected) gains from staying}
\end{array}
\tag{2.18}
$$

Clearly, the driving force behind the above results are individual expectation errors. This does by no means, however, amount to assuming irrational behaviour on the side of the individuals, given that individual decisions at $t = 1$ basically have to be taken without any previous experience nor any objective information on the future structure of labour demand. Clearly, more restrictive assumptions on the expectation formation by the individuals would not seem justified in such a transitional setting.

2.2.5 An illustrative example

Assume for illustrative purposes the following parameter values: $\delta = 1$ (no discounting), $\theta_P = 0.25$ (period-to-period learning-by-doing (tenure) effects on existing jobs of 25%), $\kappa_\theta = 1.2$ and hence $\theta_M = 0.3$ (period-to-period learning-by-doing (tenure) effects on new jobs of 30%), and $\kappa_y = 1$ (no wage difference between P- and M-jobs, i.e. no differences in the remuneration of individual characteristics). Assume, moreover, period 2 firing and hiring probabilities given by $h_P(2) = 0$ (no hiring into P-jobs in period 2), $h_M(2) = 0.75$, $f_P(2) = 0.25$, and $f_M(2) = 0.25$, and expectations of period 2 firing probabilities given by $E_1[f_P(2)] = 0.25$ and $E_1[f_M(2)] = 0.15$, so that individuals do actually underestimate the probability of being laid off from an M-job in period 2.

Given the above parameter values, $\frac{h_M(2)}{1-f_P(2)} = 1 < 1.25 = \frac{1+\theta_P}{\kappa_y}$ and $\frac{h_M(2)}{1-f_M(2)} = 1 < 1.3 = 1+\kappa_\theta\theta_P$, and hence case 1 (as defined in (2.5) on the basis of Lemma 1) applies. Further, since $\psi_1^e = -0.17$ and $c_1^e = 0.83$, expected returns to early job mobility amount to $\Delta_1^e = \{s - c_1^e\}\, y_P = \{h_M(1) + f_P(1) - 0.83\}\, y_P$, and according to Lemma 2 an individual decides to switch to a new job at t=1 iff $s = h_M(1) + f_P(1) > 0.83$.

$\frac{\partial f_{P,1}(\alpha,\nu)}{\partial\alpha} \leq 0$, $\frac{\partial f_{P,1}(\alpha,\nu)}{\partial\nu} \leq 0$, $\frac{\partial h_{M,1}(\alpha,\nu)}{\partial\alpha} \geq 0$, and $\frac{\partial h_{M,1}(\alpha,\nu)}{\partial\nu} \geq 0$. The relative influence of the unobservable characteristics on both probabilities, as expressed by the absolute values of the above first derivatives, is less clear, though. Given that current employers should have more information on a worker's unobservable ability than potential new employers, one might expect $\left|\frac{\partial f_{P,1}(\alpha,\nu)}{\partial\alpha}\right| > \left|\frac{\partial h_{M,1}(\alpha,\nu)}{\partial\alpha}\right|$. One might further expect existing networks to lose importance in the transition process and hence $\left|\frac{\partial f_{P,1}(\alpha,\nu)}{\partial\nu}\right| > \left|\frac{\partial h_{M,1}(\alpha,\nu)}{\partial\nu}\right|$. For certain sectors or occupations that see a steep increase in labour demand immediately after the labour market reforms have been announced, however, the opposite relationships might well hold true.

Given this specification of period 1 hiring and firing probabilities as functions of the unobservables, one derives the following

Proposition 5 *Let $g : [0,1]^2 \longrightarrow [0,1]$ be a positive and monotonically increasing function of the unobservable characteristics (α,ν), and let $s(\alpha,\nu) \equiv s(\kappa_y, h_{M,1}(\alpha,\nu), f_{P,1}(\alpha,\nu))$ where $h_{M,1} : [0,1]^2 \longrightarrow [0,1]$ and $f_{P,1} : [0,1]^2 \longrightarrow [0,1]$ are monotonic and twice differentiable functions of the unobservable characteristics (α,ν) such that $\frac{\partial f_{P,1}(\alpha,\nu)}{\partial\cdot} \leq 0$, $\frac{\partial^2 f_{P,1}(\alpha,\nu)}{\partial\cdot^2} \geq 0$, $\frac{\partial h_{M,1}(\alpha,\nu)}{\partial\cdot} \geq 0$ and $\frac{\partial^2 h_{M,1}(\alpha,\nu)}{\partial\cdot^2} \geq 0$. Then there exist two values $\underline{g}$ and $\overline{g}$, $0 \leq \underline{g} \leq \overline{g} \leq 1$, such that $s(\alpha,\nu) > c_1^e$ on $\{(\alpha,\nu)\,|g(\alpha,\nu) < \underline{g}\} \cup \{(\alpha,\nu)\,|g(\alpha,\nu) > \overline{g}\}$.*

Proof. From the above definition of s it follows that s is twice differentiable with $\frac{\partial s(\alpha,\nu)}{\partial\cdot} = \kappa_y \frac{\partial h_{M,1}(\alpha,\nu)}{\partial\cdot} + \frac{\partial f_{P,1}(\alpha,\nu)}{\partial\cdot}$ and $\frac{\partial^2 s(\alpha,\nu)}{\partial\cdot^2} = \kappa_y \frac{\partial^2 h_{M,1}(\alpha,\nu)}{\partial\cdot^2} + \frac{\partial^2 f_{P,1}(\alpha,\nu)}{\partial\cdot^2} \geq 0$, with convexity of s following immediately from the above assumptions. As to the monotonicity of s, it is easily seen that $\frac{\partial s(\alpha,\nu)}{\partial\cdot} \geq (\leq)\, 0$ iff $\kappa_y \left|\frac{\partial h_{M,1}(\alpha,\nu)}{\partial\cdot}\right| \geq (\leq) \left|\frac{\partial f_{P,1}(\alpha,\nu)}{\partial\cdot}\right|$. Following the convexity of s, it is further seen that s is a U-shaped function in each of the unobservables if $\kappa_y \left|\frac{\partial h_{M,1}(\alpha,\nu)}{\partial\cdot}\right|_{\cdot=0} \leq \left|\frac{\partial f_{P,1}(\alpha,\nu)}{\partial\cdot}\right|_{\cdot=0}$, while s is an increasing function in the unobservable characteristics if $\kappa_y \left|\frac{\partial h_{M,1}(\alpha,\nu)}{\partial\cdot}\right|_{\cdot=0} \geq \left|\frac{\partial f_{P,1}(\alpha,\nu)}{\partial\cdot}\right|_{\cdot=0}$. Observing that c_1^e is independent of (α,ν), the existence of the values $\underline{g}$ and $\overline{g}$ then basically follows from the mean value theorem with

$$\underline{g} = \imath\left\{s(0,0) > c_1^e\right\} \min\left\{g(\alpha,\nu)\,|s(\alpha,\nu) \leq c_1^e\right\}$$

and

$$\overline{g} = \imath\left\{s(1,1) \leq c_1^e\right\} + \imath\left\{s(1,1) > c_1^e\right\} \max\left\{g(\alpha,\nu)\,|s(\alpha,\nu) \leq c_1^e\right\}$$

where $\imath$ denotes the indicator function. Note that the above sets are empty if $\underline{g} = 0$ and $\overline{g} = 1$, respectively. ■

Hence, in case the marginal firing and hiring probabilities are non-decreasing in the unobservable characteristics (α, ν), for given parameter values κ_y, κ_θ, and θ_P, and for given expectations of period 2 hiring and firing probabilities $f_P(2), f_M(2)$, and $h_M(2)$, two groups of individuals do self-select into early job mobility: first, individuals with low ability levels or with weak networks (or both), and second, individuals with high ability levels or strong networks (or both). In the sequel, I will refer to the former as LW-individuals and to the latter as HS-individuals. In both cases, depending on the relationship between $\left|\frac{\partial f_{P,1}(\alpha,\nu)}{\partial \alpha}\right|$ and $\left|\frac{\partial f_{P,1}(\alpha,\nu)}{\partial \nu}\right|$ and between $\left|\frac{\partial h_{M,1}(\alpha,\nu)}{\partial \alpha}\right|$ and $\left|\frac{\partial h_{M,1}(\alpha,\nu)}{\partial \nu}\right|$, respectively, self-selection might be dominated by one of the two unobservable characteristics.[13]

Linear firing and hiring probabilities

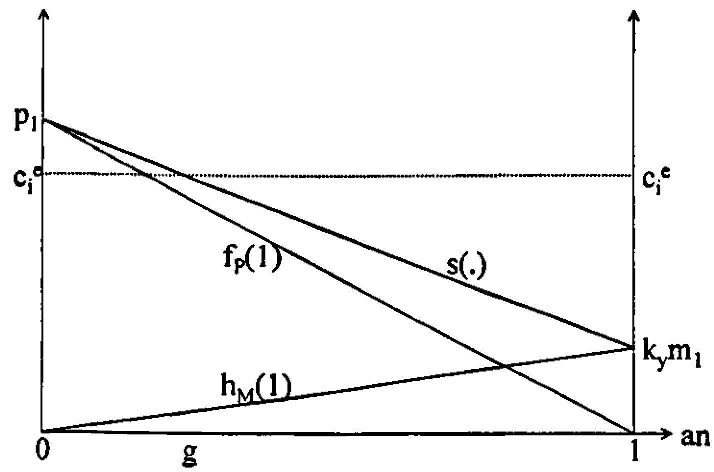
To illustrate the above result, let $g(\alpha, \nu) = \alpha\nu$, and assume the period 1 hiring and firing probabilities to be of the particular functional forms $f_{P,1}(\alpha, \nu) = p_1(1 - \alpha\nu)$, $p_1 \in [0, 1]$, and $h_{M,1}(\alpha, \nu) = m_1\alpha\nu$, $m_1 \in [0, 1]$. Then, $s(\alpha, \nu) = p_1 + (\kappa_y m_1 - p_1)\alpha\nu$, and according to Lemma 2, individuals will self-select into early job mobility iff

$$g(\alpha,\nu) = \alpha\nu \begin{cases} < \frac{c_1^e - p_1}{\kappa_y m_1 - p_1} \equiv \underline{g} & \text{for} \quad \frac{p_1}{m_1} > \kappa_y \\ > \frac{c_1^e - p_1}{\kappa_y m_1 - p_1} \equiv \overline{g} & \phantom{\text{for}} \quad \frac{p_1}{m_1} < \kappa_y \end{cases} \tag{2.19}$$

If $\frac{p_1}{m_1} = \kappa_y$, all individuals will switch to a new job at t=1 if $p_1 > c_1^e$ and stay with their job otherwise. The subgroup of individuals which self-select into early job mobility is seen to depend on both the maximal firing and hiring probabilities, p_1 and m_1, and their ratio $\frac{p_1}{m_1}$. Self-selection will occur only if $\max(p_1, \kappa_y m_1) > c_i^e$. In the case that $\min(p_1, \kappa_y m_1) > c_i^e$, all individuals switch to a new job at t=1, independently of their unobservable characteristics, while all individuals decide to stay with their job if $\max(p_1, \kappa_y m_1) < c_i^e$. If further $p_1 > c_i^e > \kappa_y m_1$,

[13]If, contrary to the above assumption of non-decreasing marginal firing and hiring probabilities. both firing and hiring functions are assumed to be concave, the opposite result holds, with *LW-individuals* and *HS-individuals* self-selecting into *immobility* early in transition. Since later results on the compliance status as a function of unobservable characteristics will remain valid even under this assumption of concavity, attention will be restricted to the case of convex firing and hiring functions in the sequel.

Figure 2-3: Job mobility and self-selectivity: Linear case



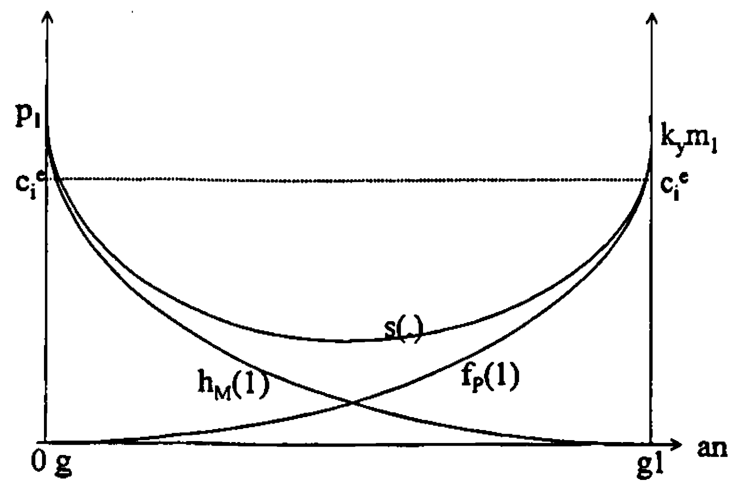
only individuals with low ability or weak networks (*LW-individuals*) self-select into early job mobility. If $\kappa_y m_1 > c_i^e > p_1$, instead, only individuals with high ability or strong networks (*HS-individuals*) do so. The relative importance of each of these two subgroups will be determined by the slope ratio $\frac{p_1}{m_1}$ in comparison to κ_y. Figure 2-3 illustrates this result for the case $p_1 > c_i^e > \kappa_y m_1$ in which only individuals with low ability or weak networks ($\alpha\nu < \underline{g}$) will self-select into early job mobility.

Quadratic firing and hiring probabilities

Assuming quadratic period 1 hiring and firing probabilities $f_{P,1}(\alpha,\nu) = p_1(1-\alpha\nu)^2$, $p_1 \in [0,1]$, and $h_{M,1}(\alpha,\nu) = m_1(\alpha\nu)^2$, $m_1 \in [0,1]$, one has $s(\alpha,\nu) = p_1 - 2p_1(\alpha\nu) + (\kappa_y m_1 + p_1)(\alpha\nu)^2$, and hence, according to Lemma 2, individuals will self-select into early job mobility iff

$$g(\alpha,\nu) = \alpha\nu \in [0,\underline{g}) \cup (\overline{g},1] \tag{2.20}$$

Figure 2-4: Job mobility and self-selectivity: Quadratic case



with

$$\underline{g} = \max(0, \frac{p_1 - \sqrt{p_1^2 - (p_1 - c_1^e)(\kappa_y m_1 + p_1)}}{(\kappa_y m_1 + p_1)})$$

and

$$\overline{g} \equiv \min(1, \frac{p_1 + \sqrt{p_1^2 - (p_1 - c_1^e)(\kappa_y m_1 + p_1)}}{(\kappa_y m_1 + p_1)}).$$

In this case, either LW-individuals or HS-individuals or both will self-select into early job mobility, depending on p_1 and $\kappa_y m_1$. As in the case of linear firing and hiring probabilities, self-selection will occur only if $\max(p_1, \kappa_y m_1) > c_i^e$. However, as opposed to the case illustrated above, in the case where $\min(p_1, \kappa_y m_1) > c_i^e$, no longer will all individuals switch to a new job at t=1 independently of their unobservable characteristics. Figure 2-4 summarises this finding for the case in which both types of individuals - those with low ability or weak networks ($\alpha\nu < \underline{g}$) and those with high ability or strong networks ($\alpha\nu > \overline{g}$) - self-select into early job mobility.

More general results can be obtained from more general functional forms of the firing and hiring probabilities, allowing for non-multiplicative and potentially asymmetric influence of the two unobservables.

2.3.2 The illustrative example continued

Assume the following specifications of the firing and hiring probabilities: $p_1 = 0.85$ and $m_1 = 0.6$, and hence $f_P(1)(\alpha,\nu) = 0.85(1-\alpha\nu)^\gamma$ and $h_M(1)(\alpha,\nu) = 0.6(\alpha\nu)^\gamma$ with $\gamma = 1$ in case of linear firing and hiring probabilities and $\gamma = 2$ in the quadratic case.

Then, for the linear case, $c_1^e - p_1 = 0.83 - 0.85 = -0.02 < 0$ and $\frac{p_1}{m_1} = \frac{0.85}{0.6} > 1 = \kappa_y$. According to Lemma 2, the individual thus moves to a new job in period 1 iff $\alpha\nu < \frac{c_1^e - p_1}{\kappa_y m_1 - p_1} = \frac{-0.02}{-0.25} = 0.08$. Hence, individuals with very low levels of either unobservable characteristic will move to a new job in this case at t=1. This can represented in a graph which plots the contour lines for the difference $s(\alpha,\nu) - c_1^e = 0.02 - 0.25\alpha\nu$ on the z-axis as a function of the unobservable characteristics α (x-axis) and ν (y-axis). Following Lemma 2, only those individuals for whom this difference is positive will self-select into early job mobility.

Similarly, for the quadratic case, $\frac{p_1 - \sqrt{p_1^2 - (p_1 - c_1^e)(\kappa_y m_1 + p_1)}}{(\kappa_y m_1 + p_1)} = \frac{0.85 - \sqrt{0.85^2 - 0.02 * 1.45}}{1.45} = 0.011885$ and $\frac{p_1 + \sqrt{p_1^2 - (p_1 - c_1^e)(\kappa_y m_1 + p_1)}}{(\kappa_y m_1 + p_1)} = \frac{0.85 + \sqrt{0.85^2 - 0.02 * 1.45}}{1.45} = 1.1605$, so that $\underline{g} = 0.011885$ and $\overline{g} = 1$, and hence only individuals with $\alpha\nu < 0.011885$ will self-select into early job mobility. As in the linear case, a graph showing the contour lines of the difference $s(\alpha,\nu) - c_1^e = 0.02 - 1.7\alpha\nu + 1.45(\alpha\nu)^2$ can illustrate the combinations of α and ν for which self-selection into early job mobility occurs.

2.4 A quasi-experimental framework for the analysis of the returns to job mobility during transition

This section restates the above model in a quasi-experimental framework. Moreover, the subgroups of interest in this framework are characterised and an instrumental variables approach is suggested for identification and estimation of the returns to early job mobility.

2.4.1 Shocks to labour demand

It has been assumed above that some unanticipated exogenous shock to the centrally planned economy acts on the period 1 hiring and firing probabilities for both types of jobs. This shock is likely to be specific to either a firm or sector, an occupation or a region while uncorrelated to the individuals' unobservable characteristics. The reason for the latter being that the individuals self-selected into the various sectors, firms, occupations or regions at $t = 0$ by their choices of the optimal human capital level H^* and their implied career choices without anticipating any such exogenous shock. It is hence inconceivable that the direct consequences of this exogenous shock - in terms of changes in hiring or firing probabilities - should be correlated to unobservable individual characteristics.

In the sequel, this shock will be modelled as a uniform upward shift of either $f_{P,1}$ or $h_{M,1}$ by some constant u_f or u_h, respectively: Assume that the firing probability is increased uniformly (i.e. at all levels of the unobservable characteristics) by u_f if the respective firm or sector is hit disproportionately by the shock or if the announced reforms lead to more pressure on labour reallocation and restructuring in this firm or sector to be expected. Such a firm or sector will be denoted as "bad" in the sequel, while firms or sectors not hit disproportionately will be called "good". Similarly, assume a uniform increase in the hiring probability by u_h if the respective occupation or region turns out to be "good" after the exogenous change in the structure of labour demand. By "good" is meant that the occupation or region is not facing any negative shock in labour demand or is even facing an increase in demand, and thus benefits disproportionately from either the exogenous shock itself or from the expectations related to the announced reforms. In the case of East Germany, one might interpret these shocks as *negative* firm/sector-specific *unification shock* and *positive* occupation/region-specific *unification shock*, respectively. These shocks are likely to be one of two types of demand shock. On the one hand, a demand shock associated with the dramatic *de facto* revaluation of the East German currency, the steep increase in wages and labour costs, and the breaking up of trade agreements together with the loss of export channels in the former COMECON area. On the other hand, a demand shock associated with the immediate modernisation of infrastructure and housing, and the privatisation and restructuring of the financial and service sectors.

The modified period 1 hiring and firing probabilities are then given by

$$\begin{aligned}\tilde{f}_{P,1}(\alpha,\nu) &= f_{P,1}(\alpha,\nu)+u_f z_f \\ &= p_1(1-\alpha\nu)^{\gamma}+u_f z_f\end{aligned} \tag{2.21}$$

and

$$\begin{aligned}\tilde{h}_{M,1}(\alpha,\nu) &= h_{M,1}(\alpha,\nu)+u_h z_h \\ &= m_1(\alpha\nu)^{\gamma}+u_h z_h\end{aligned} \tag{2.22}$$

where $u_f > 0$ and $u_h > 0$ denote the actual increase in either probability as a consequence of the exogenous shock and where $\gamma = 1$ in the linear case and $\gamma = 2$ in the quadratic case. z_f and z_h are indicators of the quality of the firm/sector or occupation as perceived at the beginning of the transition process[14] such that

$$z_f = \begin{cases} 0 & \text{if firm or sector is "good"} \\ 1 & \text{if firm or sector is "bad"} \end{cases}$$

and

$$z_h = \begin{cases} 0 & \text{if occupation or region is "bad"} \\ 1 & \text{if occupation or region is "good"} \end{cases}.$$

Individual job mobility decisions in period 1 then are made according to

Corollary 6 *Let $u \equiv \kappa_y u_h z_h + u_f z_f$. Given the exogenous shocks to labour demand u, individuals decide to switch job early iff*

$$s(\kappa_y, h_M(1), f_P(1)) > \tilde{c}_1^e \equiv c_1^e - u. \tag{2.23}$$

Given further the assumptions in Proposition 5, there exist two values $\underline{g}$ and $\tilde{g}$, $0 \leq \underline{g} \leq \underline{g} \leq$

[14]Most importantly, this assessment is assumed to be made *before* any partial resolution of uncertainty about the labour market consequences of EMSU or the design of future labour market policies.

$\tilde{g} \leq \overline{g} \leq 1$, *such that* $s(\alpha,\nu) > \tilde{c}_1^e$ *on* $\{(\alpha,\nu)|g(\alpha,\nu) < \underset{\sim}{g}\} \cup \{(\alpha,\nu)\,|g(\alpha,\nu) > \tilde{g}\}$.

Proof. Follows immediately from Lemma 2 and Proposition 5 by replacing $s(\alpha,\nu)$ by $\tilde{s}(\alpha,\nu) \equiv \kappa_y \tilde{h}_{M,1}(\alpha,\nu) + \tilde{f}_{P,1}(\alpha,\nu) = s(\alpha,\nu) + u$ and noting that c_i^e remains unchanged. Given the assumptions stated in Proposition 5, it is seen that

$$\underset{\sim}{g} = \imath\{s(0,0) > \tilde{c}_1^e\} \min\{g(\alpha,\nu)\,|s(\alpha,\nu) \leq \tilde{c}_1^e\}$$

and

$$\tilde{g} = \imath\{s(1,1) \leq \tilde{c}_1^e\} + \imath\{s(1,1) > \tilde{c}_1^e\} \max\{g(\alpha,\nu)\,|s(\alpha,\nu) \leq \tilde{c}_1^e\}.$$

■

2.4.2 Subpopulations of interest

The above model can be interpreted as a quasi-experimental framework (*Rubin's causal model*)[15] by defining z_f or z_h or some combination of them (like $\min(z_f, z_h)$ or $\max(z_f, z_h)$) as binary *assignment to treatment*, and early job mobility status (d_1) as binary *treatment indicator.*[16]

As mentioned in the introduction and as discussed in detail in chapter 3, different subpopulations of interest can be distinguished in this framework: *always-takers*, *never-takers*, and *compliers*. Always-takers do select themselves into the treatment status independently of the assignment to treatment they receive, as opposed to compliers who receive only treatment if they are assigned to treatment. The never-takers, finally, select themselves into the control group of stayers, independently of their assignment to treatment status. With respect to job mobility during decision, always-takers do change job early in transition independently of any firm/sector- or occupation-specific shocks to labour demand at the beginning of the transition process, as opposed to never-takers who decide to stay on their job under any circumstances.

[15] For an authoritative introduction into this model cf. e.g. Angrist et al. (1996) and Imbens and Rubin (1994). Cf. also chapter 3.

[16] The main intuition behind this quasi-experimental interpretation is that a higher layoff probability on the current job due to a *negative firm-specific unification shock* is likely to result in higher rates of job changes at an early stage of transition. Similarly, a higher probability of being hired on a new job due to a *positive occupation-specific unification shock* will also increase the propensity to switch to a new job early in transition.

Compliers, finally, change job only in case of such firm/sector- or occupation-specific shocks to labour demand at the beginning of the transition process but would not do so otherwise. Given this definition, the above-mentioned subpopulations of interest are given by the following

Lemma 7 *Let* $u \equiv \kappa_y u_h z_h + u_f z_f$. *Then, one can distinguish always-takers (AT), compliers (C), and never-takers (NT) with respect to the given assignment to treatment definition on the basis of their unobservable characteristics as:*

$$\begin{aligned} AT &\equiv \{(\alpha,\nu)\,|s(\alpha,\nu) \geq c_1^e\} \\ &= \{(\alpha,\nu)|0 \leq g(\alpha,\nu) < \underline{g}\} \cup \{(\alpha,\nu)|\overline{g} < g(\alpha,\nu) \leq 1\} \end{aligned} \tag{2.24}$$

$$\begin{aligned} C &\equiv \{(\alpha,\nu)|c_1^e - u \leq s(\alpha,\nu) \leq c_1^e\} \\ &= \{(\alpha,\nu)|\underline{g} \leq g(\alpha,\nu) < \underset{\sim}{g}\} \cup \{(\alpha,\nu)|\widetilde{g} < g(\alpha,\nu) \leq \overline{g}\} \end{aligned} \tag{2.25}$$

$$\begin{aligned} NT &\equiv \{(\alpha,\nu)|\, s(\alpha,\nu) \leq c_1^e - u\} \\ &= \{(\alpha,\nu)|\underset{\sim}{g} \leq g(\alpha,\nu) \leq \widetilde{g}\} \end{aligned} \tag{2.26}$$

The interpretation of these subpopulations depends on the choice of the assignment to treatment indicator. While generally four different assignment to treatment status $z = (z_h, z_f) \in \{(0,0),(0,1),(1,0),(1,1)\}$ are possible in the above setting, for practical reasons and for better interpretability one would probably restrict attention to either z_h or z_f as the assignment variable. When increasing only the firing probability at each level of ability and network accessibility, the shock mainly induces individuals with relatively low ability levels and weak networks to change job early, their main motivation being the increased probability of layoffs. When increasing the hiring probability at each ability-network combination instead, mainly individuals with relatively high ability and strong networks are induced to change job early due to good outside opportunities.

Linear firing and hiring probabilities

In the linear case, always-takers (AT), compliers (C), and never-takers (NT) are given by:

$$AT = \begin{cases} \left\{(\alpha,\nu) \middle| 0 \le \alpha\nu < \max\left(\frac{c_1^e - p_1}{\kappa_y m_1 - p_1}, 0\right)\right\} & \frac{p_1}{m_1} > \kappa_y \\ \{(\alpha,\nu) \mid p_1 > c_1^e\} & \text{if } \frac{p_1}{m_1} = \kappa_y \\ \left\{(\alpha,\nu) \middle| \min\left(\frac{c_1^e - p_1}{\kappa_y m_1 - p_1}, 1\right) < \alpha\nu \le 1\right\} & \frac{p_1}{m_1} < \kappa_y \end{cases}$$

$$C = \begin{cases} \left\{(\alpha,\nu) \middle| \max\left(\frac{c_1^e - p_1}{\kappa_y m_1 - p_1}, 0\right) \le \alpha\nu < \max\left(\frac{(c_1^e - u) - p_1}{\kappa_y m_1 - p_1}, 0\right)\right\} & \frac{p_1}{m_1} > \kappa_y \\ \{(\alpha,\nu) \mid c_1^e - u < p_1 \le c_1^e\} & \text{if } \frac{p_1}{m_1} = \kappa_y \\ \left\{(\alpha,\nu) \middle| \min\left(\frac{(c_1^e - u) - p_1}{\kappa_y m_1 - p_1}, 1\right) < \alpha\nu \le \min\left(\frac{c_1^e - p_1}{\kappa_y m_1 - p_1}, 1\right)\right\} & \frac{p_1}{m_1} < \kappa_y \end{cases}$$

$$NT = \begin{cases} \left\{(\alpha,\nu) \middle| \max\left(\frac{(c_1^e - u) - p_1}{\kappa_y m_1 - p_1}, 0\right) \le \alpha\nu \le 1\right\} & \frac{p_1}{m_1} > \kappa_y \\ \{(\alpha,\nu) \mid p_1 \le c_1^e - u\} & \text{if } \frac{p_1}{m_1} = \kappa_y \\ \left\{(\alpha,\nu) \middle| 0 \le \alpha\nu \le \min\left(\frac{(c_1^e - u) - p_1}{\kappa_y m_1 - p_1}, 1\right)\right\} & \frac{p_1}{m_1} < \kappa_y \end{cases}$$

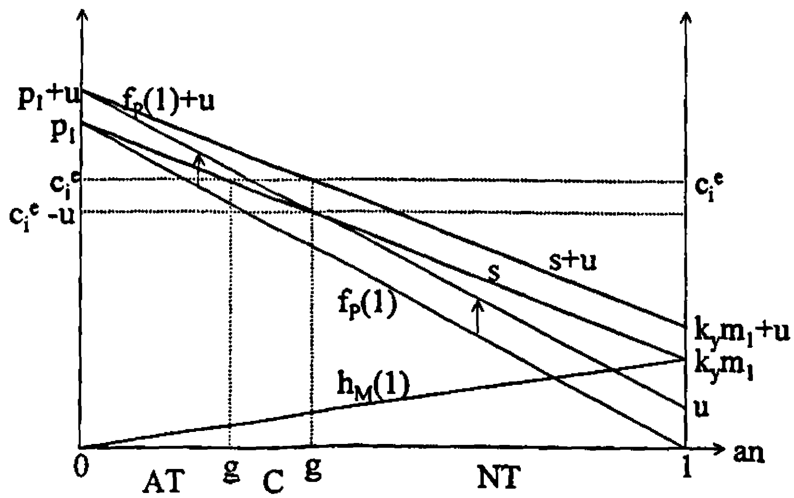
In the case $p_1 > c_i^e > \kappa_y m_1$, the above subpopulations can be represented graphically as in figure 2-5:

In this case, the subgroups of always-takers and compliers are made up of individuals with relatively low levels of ability or weak networks ($\alpha\nu < \underline{g}$). Compliers will self-select into early job mobility because of the increased firing probability, while individuals with high levels of ability or networks decide to stay on their job nevertheless.

Quadratic firing and hiring probabilities

In the quadratic case, letting $g_{l1} = \frac{p_1 - \sqrt{p_1^2 - (p_1 - c_1^e)(\kappa_y m_1 + p_1)}}{(\kappa_y m_1 + p_1)}$, $g_{u1} = \frac{p_1 + \sqrt{p_1^2 - (p_1 - c_1^e)(\kappa_y m_1 + p_1)}}{(\kappa_y m_1 + p_1)}$, $g_{l2} = \frac{p_1 - \sqrt{p_1^2 - (p_1 - c_1^e + u)(\kappa_y m_1 + p_1)}}{(\kappa_y m_1 + p_1)}$, and $g_{u2} = \frac{p_1 + \sqrt{p_1^2 - (p_1 - c_1^e + u)(\kappa_y m_1 + p_1)}}{(\kappa_y m_1 + p_1)}$ always-takers (AT),

Figure 2-5: Job mobility and compliance status: Linear case



compliers (C), and never-takers (NT) are given by

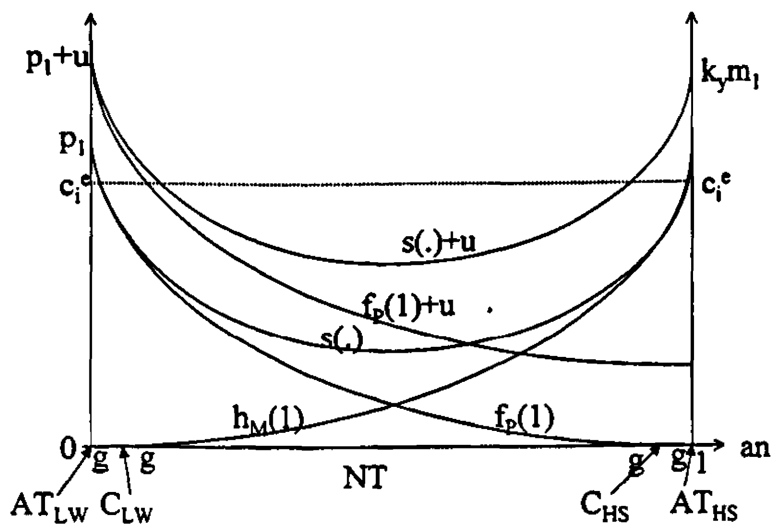
$$\begin{aligned} AT &= AT_{LW} \cup AT_{HS} \text{ with} \\ AT_{LW} &\equiv \{(\alpha,\nu)\,|0 \le \alpha\nu < \max(0, g_{l1})\} \\ AT_{HS} &\equiv \{(\alpha,\nu)\,|\min(g_{u1},1) < \alpha\nu \le 1\} \end{aligned}$$

$$\begin{aligned} C &= = C_{LW} \cup C_{HS} \text{ with} \\ C_{LW} &\equiv \{(\alpha,\nu)\,|\max(0, g_{l1}) \le \alpha\nu < \max(0, g_{l2})\} \\ C_{HS} &\equiv \{(\alpha,\nu)\,|\min(g_{u2},1) < \alpha\nu \le \min(g_{u1},1)\} \end{aligned}$$

$$NT = \{(\alpha,\nu)\,|\max(0, g_{l2}) \le \alpha\nu \le \min(g_{u2},1)\}$$

The above subpopulations can be represented graphically as in figure 2-6:

Figure 2-6: Job mobility and compliance status: Quadratic case



Contrary to the linear case discussed above, in this case individuals with intermediate levels of ability or network access decide to stay on their job despite the increase in the firing probability, while the groups of always-takers and compliers are both made up of two distinct subgroups, namely individuals with either very low or very high levels of ability and network access.

2.4.3 The returns to early job mobility for the different subpopulations

Following Corollary 3, the expected (ex ante) returns to early job mobility as a function of unobservable characteristics are given by

$$\begin{aligned} \Delta_1^e &= \{s(\alpha,\nu) - \tilde{c}_1^e\}\, y_p(H^*;\alpha,\nu) \\ &= \{s(\alpha,\nu) - c_1^e + u\}\, y_p(H^*;\alpha,\nu) \end{aligned} \tag{2.27}$$

while the realised (ex post) returns to early job mobility are given by

$$\Delta_1 = \{s(\alpha,\nu) - c_1 + u\}\, y_p(H^*;\alpha,\nu). \tag{2.28}$$

Consequently, the average returns to early job mobility for the subgroups of always takers and compliers are given by

$$\int_{(\alpha,\nu)\in AT} \{s(\alpha,\nu) - c_i + u\}\, y_p(H^*;\alpha,\nu) d\lambda(\alpha,\nu) \tag{2.29}$$

and

$$\int_{(\alpha,\nu)\in C} \{s(\alpha,\nu) - c_i + u\}\, y_p(H^*;\alpha,\nu) d\lambda(\alpha,\nu) \tag{2.30}$$

respectively, where λ denotes some measure of the weight of ability-network combinations in the subpopulation of interest. The subgroup of never-takers, finally, is not considered here since they will not move under any circumstance by definition. It is thus questionable to what extent information on their 'realised' (ex post) returns to job mobility can be of interest.

2.4.4 The illustrative example continued

In the case of linear firing and hiring probabilities, as discussed above, for the given parameter values, always-takers - those who self-select into early job mobility even without increases in either the hiring or the firing probabilities due to some exogenous shock - are characterised in terms of their unobservable characteristics as $AT = \{(\alpha,\nu)\,|\alpha\nu < 0.08\}$.

Now assume a negative shock to some firm or sector which is modelled by $u = u_f = 0.15$. Then, individuals will decide to change job early iff $\alpha\nu < \underline{g} = \frac{\tilde{c}_1^e - p_1}{\kappa_y m_1 - p_1} = \frac{c_1^e - u - p_1}{\kappa_y m_1 - p_1} = \frac{0.83-1}{-0.25} = 0.68$, and the group of compliers will be given by $C = \{(\alpha,\nu)\,|0.08 \leq \alpha\nu < 0.68\}$. This group is made up of individuals with either relatively low ability or weak networks or both. In the absence of an exogenous increase in the firm-specific layoff probability, these individuals would not want to change to another job or employer. However, given this increase, they are forced to switch to a new job early.

The never-takers are given by $NT = \{(\alpha,\nu)\,|\alpha\nu \geq 0.68\}$. Given their either relatively high ability level or their good access to networks, or both, they prefer staying on their job even in case their firm turns out to be a "bad" one.

As for the subpopulation self-selecting into early-job mobility, the compliance status can be represented graphically by the modified contour lines for the difference $s(\alpha,\nu) - (c_1^e - u) = 0.17 - 0.25\alpha\nu$. For never-takers, this difference remains negative, while for compliers it is shifted to positive values.

Similarly, in the case of quadratic firing and hiring probabilities, the subpopulations of interest are given by $AT = \{(\alpha,\nu)\,|\alpha\nu < 0.011885\}$, $C = \{(\alpha,\nu)\,|0.011885 \leq \alpha\nu < 0.11039\}$, and $NT = \{(\alpha,\nu)\,|\alpha\nu \geq 0.11039\}$. The modified difference necessary for the plot of contour lines is $s(\alpha,\nu) - (c_1^e - u) = 0.17 - 1.7\alpha\nu + 1.45(\alpha\nu)^2$ in this case.

The realised returns to early job mobility amount to $\Delta_1 = \{0.04 - 0.25\alpha\nu\}\, y_p(H^*;\alpha,\nu)$ in the linear case and to $\Delta_1 = \left\{0.04 - 1.7\alpha\nu + 1.45(\alpha\nu)^2\right\} y_p(H^*;\alpha,\nu)$ in the quadratic case. In both cases, they are in the range $\Delta_1 \in (0.02y_p, 0.04y_p]$ for always takers and $\Delta_1 \in (-0.13y_p, 0.02y_p]$ for compliers and hence on average positive for always takers while likely to be negative for compliers. Given the nature of the assignment mechanism, the latter can probably best be interpreted as returns to *forced* early job mobility during transition.

Assume now a positive shock specific to some occupation or region which is modelled by $u = u_h = 0.4$ instead. Then, given the substantial increase in the probability of being hired on a new job, in the linear case all individuals will decide to change job early with the groups of compliers and never takers consequently given by $C = \{(\alpha,\nu)\,|0.08 \leq \alpha\nu \leq 1\}$ and, $NT = \varnothing$, respectively. In the quadratic case, the subpopulations of interest are given by $AT = \{(\alpha,\nu)\,|\alpha\nu < 0.011885\}$, $C = \{(\alpha,\nu)\,|0.011885 \leq \alpha\nu < 0.20721\} \cup \{(\alpha,\nu)\,|0.9652 < \alpha\nu \leq 1\}$, and $NT = \{(\alpha,\nu)\,|0.20721 \leq \alpha\nu \leq 0.9652\}$. In this case, the subgroup of compliers is hence made up of both types of individuals, *LW-individuals* and *HS-individuals*, who can to be interpreted as *voluntary* early job movers in this case.

The realised returns to early job mobility amount to $\Delta_1 = \{0.29 - 0.25\alpha\nu\}\, y_p(H^*;\alpha,\nu)$ in the linear case and to $\Delta_1 = \left\{0.29 - 1.7\alpha\nu + 1.45(\alpha\nu)^2\right\} y_p(H^*;\alpha,\nu)$ in the quadratic case. In both cases, they are in the range $\Delta_1 \in (0.27y_p, 0.29y_p]$ for always takers. For compliers, $\Delta_1 \in (0.04y_p, 0.27y_p]$ in the linear case, and in the quadratic case $\Delta_1 \in (1.8 \times 10^{-7}y_p, 0.27y_p]$

for *LW-individuals* who comply with the assignment, and $\Delta_1 \in (-4 \times 10^{-6} y_p, 0.04 y_p]$ for *HS-individuals* who comply with the assignment. The returns to early job mobility are again seen to be positive on average for always takers and this time also positive on average for compliers.[17]

2.5 Summary and conclusions

This chapter has presented a model of individual decision-making in a transitional labour market which helps to clarify the role of unobservable characteristics in job mobility decision making and rationalises the endogeneity status of the job mobility indicator. More precisely, under the model assumptions it can be shown that mainly individuals with either relatively low levels of ability and network access or with relatively high levels of ability and good access to networks do self-select into early job mobility.[18]

Rephrasing the model in a quasi-experimental framework allows, first, a characterisation of the subgroups of compliers in terms of their unobservable characteristics, and second, the calculation of counterfactual gains or losses from job mobility for this subgroup. On the basis of appropriate assignment to treatment mechanisms, the compliers can be interpreted as the subgroup of either forced early job movers or voluntary early job movers. Moreover, the returns to each of these types of job mobility are seen to differ significantly across the subgroups of always takers and compliers.

The model thus underlines that the so-called *local average treatment effect* (LATE) can be an economically interesting parameter in this context and, finally, suggests IV estimation of the (causal) returns to job mobility on the basis of appropriate instruments like e.g. firm-level announced layoffs (as suggested by the above described shift in $f_P(1)$ due to a sector- or firm-specific shock induced by the announced reforms) or like occupation-specific outside options (as suggested by the shift in $h_M(1)$ due to an occupation-specific shock induced by the announced

[17]As mentioned above, there is no interest in calculating the 'realised' returns to early job mobility for the subgroup of never-takers in this framework. Doing so nevertheless yields a range of $\Delta_1 \in (-0.13 y_p, -0.21 y_p]$ in the case $u = u_f = 0.15$ (in both linear and quadratic case), and $\Delta_1 \in (-4 \times 10^{-6} y_p, 1.8 \times 10^{-7} y_p]$ in the case $u = u_h = 0.4$ (in the quadratic case only, since $NT = \emptyset$ in the linear case).

[18]Steiner et al. (1998) present some evidence that, in the group of early job movers, two 'occupational' groups seem to be dominating, namely the unqualified (probably individuals with lower ability levels being affected by negative firm/sector-specific demand shocks) and service sector workers (probably individuals with higher ability levels and/or good networks who are mainly affected by positive occupation-specific demand shocks).

reforms). After reviewing the necessary results on identification and estimation of causal effects by instrumental variables methods in the next chapter, this empirical issue will be addressed in chapter 4.[19]

2.6 References

Alexeev, M. and M. Kaganovich (1998), Returns to human capital under uncertain reform: Good guys finish last, Journal of Economic Behavior and Organization, 37, 53-70

Angrist, J., G.W. Imbens and D.W. Rubin (1996), Identification of causal effects using instrumental variables, Journal of the American Statistical Association, 91(434), 444-472

Imbens, G.W. and J.D. Angrist (1994), Identification and estimation of local average treatment effects, Econometrica, 62(2), 467-475

Jovanovic, B. and Y. Nyarko (1996), Learning-by-doing and the choice of technology, Econometrica, 64, 1299-1310

Siebern, F. (2000), Better LATE?, Instrumental variables estimation of the returns to job mobility during transition, German Economic Review, 1(3), 335-362

Steiner, V. et al. (eds) (1998), Strukturanalyse der Arbeitsmarktentwicklung in den neuen Bundeslaendern, ZEW-Wirtschaftsanalysen, Band 30, Nomos-Verlag, Baden-Baden

2.7 Appendix

This section presents the above lemmata and propositions for the cases i=2,3,4:

The expected present discounted value of future income streams as a function of d at t=1 is given by

$$\begin{aligned} Y^e_{(i,j)} &\equiv E_1\left[Y \mid d_1 = i, d_2 = j\right] \qquad \text{(ad (2.7))} \\ &= \begin{cases} \{(1 - f_P(1)) + \delta(1+\theta_P)E_1\left[1 - f_P(2)\right]\} y_P & d = (0,0) \\ \{h_M(1) + \delta(1+\kappa_\theta\theta_P)E_1\left[1 - f_M(2)\right]\} \kappa_y y_P & d = (1,0) \\ \{(1 - f_P(1)) + \delta\kappa_y E_1\left[h_M(2)\right]\} y_P & d = (0,1) \\ \{h_M(1) + \delta E_1\left[h_M(2)\right]\} \kappa_y y_P & d = (1,1) \end{cases} \quad \text{if} \end{aligned}$$

[19] Cf. also Siebern (2000).

Define

$$s(\kappa_y, h_M(1), f_P(1)) \equiv \kappa_y h_M(1) + f_P(1) \tag{ad (2.8)}$$

and

$$c_i^e \equiv 1 + \delta\psi_i^e(\kappa_y, \kappa_\theta, \theta_P, f_P(2), f_M(2), h_M(2)), \tag{ad (2.9)}$$

where

$$\psi_1^e(.) = \begin{cases} E_1[(1+\theta_p)(1-f_P(2)) - \kappa_y(1+\kappa_\theta\theta_P)(1-f_M(2))] & i=1 \\ E_1[\kappa_y h_M(2) - \kappa_y(1+\kappa_\theta\theta_P)(1-f_M(2))] & \text{if } i=2 \\ E_1[(1+\theta_P)(1-f_P(2)) - \kappa_y h_M(2)] & i=3 \\ 0 & i=4 \end{cases} \tag{ad (2.10)}$$

are functions of the expected differences in both pay and learning-by-doing (tenure) effects between P- and M-jobs, and in the layoff probabilities and hiring probabilities, respectively, on the two jobs in the second period.

Then, individuals switch to a new job early iff

$$s(\kappa_y, h_M(1), f_P(1)) > c_i^e. \tag{ad (2.11)}$$

The realised returns to early job mobility are given by

$$\begin{aligned} \Delta_i &= \kappa_y h_M(1) + f_P(1) - c_i \qquad \text{(ad (2.13))} \\ &= \kappa_y h_M(1) + f_P(1) - 1 - \delta\psi_i(\kappa_y, \kappa_\theta, \theta_P, f_P(2), f_M(2), h_M(2)), \end{aligned}$$

where

$$\psi_i(.) = \begin{cases} (1+\theta_p)(1-f_P(2)) - \kappa_y(1+\kappa_\theta\theta_P)(1-f_M(2)) & i=1 \\ \kappa_y h_M(2) - \kappa_y(1+\kappa_\theta\theta_P)(1-f_M(2)) & i=2 \\ (1+\theta_P)(1-f_P(2)) - \kappa_y h_M(2) & i=3 \\ 0 & i=4 \end{cases} \text{ for } \qquad \text{(ad (2.15))}$$

denotes the realisation of $\psi_i^e(.)$ in period 2.

Consequently, the difference of realised and expected returns to early job mobility is given by

$$\Delta_i - \Delta_i^e = \{\psi_i^e - \psi_i\}\, y_P \qquad \text{(ad (2.16))}$$

where

$$\psi_i^e - \psi_i = \begin{cases} (1+\theta_p)e(f_P(2)) - \kappa_y(1+\kappa_\theta\theta_P)e(f_M(2)) & i=1 \\ \kappa_y e(h_M(2)) - \kappa_y(1+\kappa_\theta\theta_P)e(f_M(2)) & i=2 \\ (1+\theta_P)e(f_P(2)) - \kappa_y e(h_M(2)) & i=3 \\ 0 & i=4 \end{cases} \text{ for } . \qquad \text{(ad (2.16))}$$

with expectation errors defined as

$$\begin{aligned} e(f_P(2)) &\equiv E_1[1-f_P(2)] - (1-f_P(2)) = f_P(2) - E_1[f_P(2)] \\ e(f_M(2)) &\equiv E_1[1-f_M(2)] - (1-f_M(2)) = f_M(2) - E_1[f_M(2)] \, . \\ e(h_M(2)) &\equiv E_1[h_M(2)] - h_M(2) \end{aligned} \qquad \text{(ad (2.17))}$$

Chapter 3

Identification and estimation of causal effects in observational studies

3.1 Introduction

The previous chapter has developed a model according to which individuals self-select into specific job mobility patterns during transition on the basis of (to the econometrician) unobservable characteristics like ability or access to networks that are correlated with earnings. This *selection* (or *endogeneity*) *problem* generally invalidates inference on the returns to job mobility in standard regression approaches.

In the applied statistics literature, this problem has been treated extensively as problem of inference from nonrandom samples (e.g. Smith (1983) and references cited therein). Applied economists, too, have been aware of this problem, and work by Roy (1951) and Gronau (1974) figures among the earliest and most prominent applications in labour economics. Starting with the influential work by Heckman (1976,1978,1979) on the so-called *dummy endogenous variable model*, they have developed a range of identification strategies and estimation methods that take account explicitly of self-selectivity.[1]

[1] For overviews, cf. e.g. Angrist and Krueger (1998), Heckman (1990,1999b), Manski (1993, 1994, 1995) and

These methods generally require rather strong parametric identifying assumptions regarding both the functional form and the error distributions. Two problems are therefore related to this parametric identification approach. First, the assumptions made to achieve identification are generally restrictive and untestable. Second, further (even more restrictive) implicit assumptions have to be made for some parameter of interest - like the coefficient on a job mobility indicator in an earnings equation - to be interpretable as a *causal effect*.[2] More recent contributions to the literature allow for semi- and nonparametric identification of selectivity models, thus avoiding unnecessarily strong identifying assumptions.[3] The second criticism, however, also applies to these semi- and nonparametric versions of the dummy endogenous variable model.

The main reason for this criticism lies in the understanding that causal effects can in general only be identified in a *potential outcome* framework. This framework stems originally from the experimental biostatistical literature and has found its way into the economics literature in particular in the context of evaluating the labour market impact of training programmes and active labour market measures,[4] the effects of military service on subsequent earnings, the returns to education and various other applications based on non-experimental data.[5]

In the form of a commented literature review, this chapter presents the *potential outcome model* as a general statistical framework for causal inference in observational studies and clarifies the assumptions needed for the identification of causal effects by instrumental variables methods.[6] The presentation will be related to the subsequent empirical applications: first, the estimation of returns to job mobility during transition - where job mobility is interpreted as treatment the effect of which on income and employment probabilities as outcome variable is to be evaluated; and second, the estimation of returns to education - with duration of schooling

Rosenbaum (1995).

[2] Cf. e.g. Heckman (1997,1999) and section 3 below.

[3] Cf. e.g. Heckman and Honore (1990), Ahn and Powell (1993), Chen (1999) and Froehlich (1998).

[4] Cf. e.g. Friedlander et al. (1997), Ham and LaLonde (1990), LaLonde (1986,1995), Heckman (1992), Heckman and Hotz (1989), Heckman et al. (1999), Heckman and Robb (1995a,b) and Hotz (1992). For applications to German data or other transition economies, cf. also Fitzenberger and Prey (1996,1997), Hübler (1997), Kraus et al. (1997), Lechner (1998a,b,1999), Puhani (1998,1999), Schmidt (1999) Steiner and Kraus (1995).

[5] Cf. e.g. Angrist (1990,1998), Angrist and Krueger (1991,1998), Card (1995,1999), Ichino and Winter-Ebmer (1999,2000) and Meyer (1995a,b).

[6] The chapter builds extensively on Angrist and Imbens (1995), Angrist et al. (1996), Heckman (1997), Heckman and Vytlacil (1999,2000a,b,c), Holland (1986), Ichino (1999), Imbens and Angrist (1994) and Imbens and Rubin (1997).

as treatment and earnings as outcome variable. The chapter is organised as follows:

Section 3.2 provides the basic notation, defines the treatment effects of interest and derives the relationship between them. Section 3.3 reviews potential identification strategies and states the assumptions sufficient for instrumental variables estimates to have a causal interpretation, in particular as *local average treatment effect* (*LATE*). It also discusses some extensions to more general outcome variables and further reviews results on the identification and estimation of local average treatment effects in the case of variable treatment intensity or, equivalently, multivalued treatment variables. Section 3.4 concludes.

3.2 A general statistical framework for causal inference: The potential outcome model

The *potential outcome model* (or *Rubin's causal model*) considers the *causal effect* of some *treatment* on potential outcomes *(counterfactuals)*.[7] The underlying hypothetical question in this model is of the type "*What would the outcome have been had state 2 been realised instead of state 1?*" or, in the concrete applications of the later chapters, "*What would a job mover have earned had he stayed in his previous job?*", "*What would a person have earned if he had stayed one more year in school?*" and the like. *Causal effects* are then defined as the difference between the realised outcome and the relevant counterfactual outcome.

This section reviews the model as a framework for causal inference in non-experimental studies, provides the basic notation, defines the (unit-level) causal effects and (average) treatment effects of interest and clarifies the relationship between them.

3.2.1 Basic notation and definitions

Let Y_{1i} and Y_{0i} denote some potential outcome of interest in case of treatment and non-treatment, respectively, and let Y_i denote the (observable) outcome for individual (or, equivalently, *unit*) i with observable individual characteristics X_i in a population I. Assume that

[7]Cf. the seminal contributions Rubin (1974, 1977, 1978). Acknowledging the early contributions on experimental approaches in the statistical literature by Neyman (1923), Fisher (1935) and Cox (1958), Heckman and Vytlacil (2000a,b,c) also speak of the *Neyman - Fisher - Cox - Rubin model of potential outcomes*. Moreover, as noted by Heckman and Vytlacil, this model is isomorphic to both the Roy (1951) model of income distribution and the switching regression model of Quandt (1972).

the potential outcome functions are of the form $Y_{1i} = \mu_1(X_i, u_{1i})$ in case of treatment and $Y_{0i} = \mu_0(X_i, u_{0i})$ otherwise, with generally unrestricted functions μ_0 and μ_1 and $E[u_{0i}|X_i] = E[u_{1i}|X_i] = 0$. Conditional on observable characteristics X_i, $(\mu_0(.), \mu_1(.))$ and (u_{0i}, u_{1i}) denote the common components and the idiosyncratic components of the potential outcomes across the two treatment status, respectively. It is generally further assumed that the potential outcome functions are additively separable in the common and idiosyncratic components, i.e. $Y_{1i} = \mu_1(X_i, u_{1i}) = \mu_1(X_i) + u_{1i}$ in case of treatment and $Y_{0i} = \mu_0(X_i, u_{0i}) = \mu_0(X_i) + u_{0i}$ otherwise.

Let further D_i denote some binary *treatment* indicator taking the value 1 in case of (receipt of) treatment of individual i and 0 otherwise, and let Z_i be some binary *intention-to-treat* variable that describes the *assignment* mechanism into treatment and control groups. Z_i takes the value 1 if individual i is assigned to the treatment group and 0 otherwise. Assignment to treatment is described by the conditional probability of assignment, conditional on observable individual characteristics and potential outcomes. It is defined as *strongly ignorable* iff, conditional on observable characteristics, the probability of assignment is independent of potential outcomes. In case $D_i = Z_i$ for all $i \in I$, there is full compliance with the assignment. Let D_{1i} and D_{0i} denote the potential treatment status in case of assignment to treatment and in case of non-assignment to treatment, respectively.[8]

Assume an individual's treatment status to be potentially a function of all individuals' assignment status, $D_i = D(\tilde{Z}; X_i)$, where $\tilde{Z} = (Z_1, ..., Z_N)$ denotes the full vector of assignment status in the population. Similarly, let the outcome variable Y_i for the individual potentially depend on all individuals' assignment and treatment status, $Y_i = Y(\tilde{D}, \tilde{Z}; X_i)$, where further $\tilde{D} = (D_1, ..., D_N)$ denotes the full vector of treatment status in the population. The observed cross-sectional data are $((z_i, d_i, y_i), X_i)_{i=1,...,N}$.

[8] One might, further, also think of the variables in X_i as dependent on the treatment status and hence define X_{i1} and X_{i0} as the potential individual characteristics in the case of treatment and in its absence, respectively. In general, however, the assumption $X_i = X_{i0} = X_{i1}$ is made, since it would otherwise be difficult to separate direct causal effects of D_i on Y_i from the consequences of changing individual characteristics X_i for Y_i. (Cf. Holland (1986) and Lechner (1998a,b)).

3.2.2 Subpopulations of interest

Within the population under consideration, several subpopulations might be distinguished on the basis of their observed or potential treatment status. The most common subpopulation of interest in evaluation studies is the *treatment group* (or simply *the treated*), i.e. all those individuals who actually receive treatment, $I_T \equiv \{i \in I \,|D_i = 1\}$ as opposed to the *control group*, $I_{\overline{T}} \equiv \{i \in I \,|D_i \neq 1\}$.

On the basis of the individuals' potential treatment status, one might further distinguish between the following groups:

$$always-takers : I_{AT} \equiv \{i \in I \,|D_{0i} = D_{1i} = 1\}$$

$$never-takers : I_{NT} \equiv \{i \in I \,|D_{0i} = D_{1i} = 0\}$$

$$compliers : I_C \equiv \{i \in I \,|D_{1i} - D_{0i} = 1\}$$

$$defiers : I_D \equiv \{i \in I \,|D_{1i} - D_{0i} = -1\}$$

While the first two groups are not affected by the assignment to treatment, the latter two groups are both affected, but they differ in their compliance behaviour. As suggested by the above names, the compliers fully comply with the assignment mechanism. In contrast, the defiers do right the opposite and self-select into treatment only if they are not assigned to it.

The relationship between these different subgroups depends on the experimental design, i.e. on the assignment mechanism and the compliance behaviour. An individual is generally said to comply with the assignment Z_i if $D_i = Z_i$. However, some individual with $D_i = Z_i = 1$ might have self-selected into treatment even in the (counterfactual) case of non-assignment, or an individual with $D_i = Z_i = 0$ might not have complied with an assignment to treatment. The potential outcome model draws a distinction between these different types of *observed compliers*, denoting any *potential non-complier* among these as either always-taker or never-taker. Observed non-compliers, i.e. $D_i \neq Z_i$, on the other hand, can be part of one of these two categories or they can be defiers. The compliers are all those individuals who only change treatment status because of assignment.

3.2.3 Parameters of interest

The notion of *causal effect* in the potential outcome model refers to the effect of D_i on Y_i, $\Delta_i \equiv (Y_{1i} - Y_{0i})$. Δ_i is the so-called *(unit-level or individual-specific) treatment effect.* This effect has a counterfactual interpretation by definition. Since only $Y_i = Y_{0i}(1 - D_i) + Y_{1i}D_i = Y_{0i} + \Delta_i D_i$ can be observed, it cannot be inferred directly.[9] Interest therefore generally focuses on some feature of the distribution of Δ_i in some (sub-)population of interest, generally some *average* (or *mean*) *treatment effect.* According to the population of interest one might distinguish between three different treatment effects, namely the *average treatment effect* or *treatment effect on a person selected randomly from the population,* $\Delta_{ATE} \equiv E[\Delta_i | X_i]$, the *treatment effect on the treated,* $\Delta_{TT} \equiv E[\Delta_i | X_i, D_i = 1]$, and the *local average treatment effect* (*LATE*), $\Delta_{LATE} \equiv E[\Delta_i | X_i, (D_{1i} - D_{0i}) \neq 0]$.[10] [11]

Under the above assumption of additively separable potential outcome functions, the unit-level treatment effect can then be written as $\Delta_i = (\mu_1(X_i) - \mu_0(X_i)) + (u_{1i} - u_{0i})$. The

[9] Holland (1986) refers to this as the *fundamental identification problem of causal inference.*

[10] If interest is restricted to some subpopulation characterised by a vector of characteristics $X_i = \bar{x}$, like e.g. all unemployed or all public service employees, these effects can be defined accordingly as $E[\Delta_i | X_i = \bar{x}]$, $E[\Delta_i | D_i = 1, X_i = \bar{x}]$, and $E[\Delta_i | (D_{i1} - D_{i0}) \neq 0, X_i = \bar{x}]$.

[11] While based on different conditioning sets all of these effects are *mean treatment effects.* In evaluating the effect of some treatment, one might, however, also be interested in distributional effects on the outcome variable. More general parameters of interest for causal inference which might capture such effects can be defined. Among these are: median treatment effects, quantile treatment effects (Abadie et al. (1998)); the distribution of programme gains $F(Y_1 - Y_0 | X)$; and the difference of full outcome distributions by treatment status, $F(Y_1 | X) - F(Y_0 | X)$. The main difficulty generally consists in identifying the joint distributions of the potential outcomes, viz. $F(Y_0, Y_1 | X)$. For a more in depth discussion, cf. Manski (1993, 1994, 1995). In this work, we will follow advice by Lechner (1998a): "However, the latter quantities are impossible to estimate without knowledge of the *joint* distribution of the potential outcomes, whereas the expectation solely depends on the two marginal distributions of the potential outcomes. Therefore, causal analysis using the other features would need far more stringent assumptions to achieve identification than causal analysis based on first moments only. Since identification of the expectations will prove to be difficult enough, identifying these other quantities will not be attempted in this work."

first term, $(\mu_1(X_i) - \mu_0(X_i))$, denotes the common gain from treatment which is equal for all individuals with the same observable characteristics X_i. The second term, $(u_{1i} - u_{0i})$, the idiosyncratic gains from treatment, generally differs even across observably identical individuals. While they might be observed by the individuals themselves, they are generally unobservable to the econometrician.

Conditional on individual characteristics X_i, treatment effects are heterogeneous in this model unless $u_{0i} = u_{1i}\ \forall i = 1, ..., N$. In the special case of $\mu_1(X_i) = \mu_0(X_i) + \Delta$ and $u_{0i} = u_{1i}$ for all $i \in I$, the treatment effect is in fact $\Delta_i = \Delta$ for all $i \in I$ and hence homogenous across individuals conditional on individual characteristics X_i.

In the case of job mobility, the return to job mobility for a randomly selected individual who is subsequently forced to change job and hence the average treatment effect Δ_{ATE} are certainly not of interest. It is less clear, however, which of the remaining two effect, Δ_{TT} or Δ_{LATE}, constitutes the main parameter of interest and which can actually be identified on the basis of the observed data.

3.2.4 Relationship between the parameters of interest

Recalling the definition of (unit-level) treatment effects, $\Delta_i = (\mu_1(X_i) - \mu_0(X_i)) + (u_{1i} - u_{0i})$, and assuming the absence of any systematic difference in the idiosyncratic components of potential income, $E[u_{0i}|X_i] = E[u_{1i}|X_i]$, the above defined average treatment effects can be expressed in terms of the common and idiosyncratic components of the potential outcomes as follows:

$$\begin{aligned}
\Delta_{ATE} &\equiv E[\Delta_i|X_i] \\
&= E[(\mu_1(X_i) - \mu_0(X_i)) + (u_{1i} - u_{0i})|X_i] \\
&= \mu_1(X_i) - \mu_0(X_i) + E[u_{1i} - u_{0i}|X_i] \\
&= \mu_1(X_i) - \mu_0(X_i)
\end{aligned} \tag{3.1}$$

$$
\begin{aligned}
\Delta_{TT} &\equiv E[\Delta_i | X_i, D_i = 1] \\
&= E[(\mu_1(X_i) - \mu_0(X_i)) + (u_{1i} - u_{0i}) | X_i, D_i = 1] \\
&= \mu_1(X_i) - \mu_0(X_i) + E[u_{1i} - u_{0i} | X_i, D_i = 1] \\
&= \mu_1(X_i) - \mu_0(X_i) + E[u_{1i} - u_{0i} | D_i = 1]
\end{aligned} \tag{3.2}
$$

$$
\begin{aligned}
\Delta_{LATE} &\equiv E[\Delta_i | X_i, (D_{i1} - D_{i0}) \neq 0] \\
&= E[(\mu_1(X_i) - \mu_0(X_i)) + (u_{1i} - u_{0i}) | X_i, (D_{1i} - D_{0i}) \neq 0] \\
&= \mu_1(X_i) - \mu_0(X_i) + E[u_{1i} - u_{0i} | X_i, (D_{1i} - D_{0i}) \neq 0] \\
&= \mu_1(X_i) - \mu_0(X_i) + E[u_{1i} - u_{0i} | (D_{1i} - D_{0i}) \neq 0]
\end{aligned} \tag{3.3}
$$

In the absence of idiosyncratic responses to treatment, $u_{1i} = u_{0i}\ \forall i \in I$, all individuals with the same characteristics respond in the same way to the treatment and hence all three average treatment effects are in fact equal. The same result holds in the case of strongly ignorable assignment and full compliance with the assignment, in which case potential outcomes are independent of the assignment mechanism, and necessarily $D_{1i} = Z_i = 1$ and $D_{0i} = Z_i = 0$.

In most empirical applications, these conditions are unlikely to be met and the above average treatment effects will not coincide. In particular the average treatment effect for the treated and the average treatment effect for a randomly selected person will differ from each other unless the average idiosyncratic gains for the treated are zero or, in other words, if "agents do not select into treatment on the basis of expected (idiosyncratic) gains from it." (cf. Heckman (1997)) Similarly, the local average treatment effect will differ from these two effects unless the average idiosyncratic gains for the subpopulations of compliers and defiers are zero and the treatment effects for these two groups are further identical with those for always-takers. More generally:

$$\Delta_{ATE} = \Delta_{TT} \Leftrightarrow E[u_{1i} - u_{0i} \,|D_i = 1] = 0, \tag{3.4}$$

$$\Delta_{ATE} = \Delta_{LATE} \Leftrightarrow E[u_{1i} - u_{0i} \,|(D_{1i} - D_{0i}) \neq 0] = 0 \tag{3.5}$$

Hence, *a priori* information on the assignment mechanism and the compliance behaviour is of crucial importance for the identification of treatment effects. In the absence of such information, identifying assumptions have to be invoked. This will be discussed in more detail below.

3.2.5 Relationship to economic models: The marginal treatment effect

An additional treatment effect which is explicitly derived from economic theory and discrete choice formulations of individual decisions on the treatment status can be defined in a latent variable framework: Let potential outcome functions and observed outcome be defined as above, and assume that the treatment status D_i is generated according to a latent variable (or single index) model

$$\begin{aligned} D_i &= \imath\{D_i^* > 0\} \\ &= \imath\{u_{Di} < \mu_D(Z_i, X_i)\} \end{aligned} \tag{3.6}$$

where

$$\begin{aligned} D_i^* &= \mu_D(Z_i, X_i, u_{Di}) \\ &= \mu_D(Z_i, X_i) - u_{Di} \end{aligned} \tag{3.7}$$

with unobservable random variable u_{Di} which is generally interpreted as unobserved net utility or net gain from choosing one of several alternatives.

Definition 8 *(marginal treatment effect, Heckman and Vytlacil (1999,2000a,b,c))*

Additional to the above "index condition" 3.6 assume the following regularity conditions:

(1) Conditional on $X = x$, $\mu_D(Z_i, X_i = x)$ is a nondegenerate random variable.

(2) (u_{0i}, u_{Di}) and (u_{1i}, u_{Di}) are absolutely continuous.

(3) (u_{0i}, u_{Di}) and (u_{1i}, u_{Di}) are independent of (Z_i, X_i).

(4) $\Pr(D_i = 1, X_i) > 0$, *i.e. the probability of treatment is strictly positive at each level of $X = x$ for each individual i.*

Define, moreover, the propensity score as

$P(z) \equiv \Pr(D_i = 1 \,| X_i, Z_i = z) = F_{u_D}(\mu_D(Z_i = z, X_i))$

where F_{u_D} is the distribution function of u_D. Then, conditional on $X_i = x$ and $P(Z_i) = P(z)$, the marginal treatment effect $\Delta_{MTE}(x, P(z))$ is defined as

$$\Delta_{MTE}(x, P(z)) \equiv \frac{\partial E[Y_i \,| X_i = x, P(Z_i) = P(z)]}{\partial P(z)} \tag{3.8}$$

Like the local average treatment effect Δ_{LATE}, the marginal treatment effect Δ_{MTE} is defined on the basis of an instrument Z and thus varies with it. This implies that, using different instruments, one should expect different estimates of causal effects for different subpopulations. Contrary to the local average treatment effect, however, the marginal treatment effect is - by definition - a continuous function of $P(z)$.

The relationships between the marginal treatment effect and the above parameters are derived in Heckman and Vytlacil (1999,2000a,b,c). They show in particular that the above defined average treatment effects can be generated by integrating the marginal treatment effect over some well-defined range of the propensity score used as instrument. The marginal treatment effect Δ_{MTE} thus is a limit form of the local average treatment effect. Heckman and Vytlacil (1999) interpret it as "the average effect for people who are just indifferent between participation or not at the given value of the instrument". For values of $P(z)$ close to 1, it is "the average effect for individuals with unobservable characteristics that make them the most inclined to participate" and, similarly, for values of $P(z)$ close to 0, it is "the average effect for individuals with unobservable characteristics that make them the least inclined to participate". The relationships between the above treatment effects are summarised in the following

Corollary 9 *(relationship between the different average treatment effects, Heckman and Vyt-*

lacil (1999,2000a,b,c))

Let $P(z) \neq P(z')$ at each pair (z, z') where $P(z) \equiv P(Z_i)$, and let $\mu_D(z) \equiv \mu_D(Z_i = z, X_i = x, u_{Di})$. In the context of the above latent variable model, the conditional versions of the above average treatment effects conditional on $X_i = x$ then can be written as

$$\Delta_{ATE}(x) \equiv E[\Delta_i | X_i = x],$$

$$\Delta_{TT}(x, P(z)) \equiv E[\Delta_i | X_i = x, P(Z_i) = P(z), D_i = 1]$$

$$\Delta_{LATE}(x, P(z), P(z')) = E\left[\Delta_i | X_i = x, (\imath\{\mu_D(z) > 0\} - \imath\{\mu_D(z') > 0\}) \neq 0\right]$$

Given the assumptions from the above definition, it follows that these (conditional) average treatment effects are an appropriately weighted average of the marginal treatment effect for some range of the propensity score with

$$\Delta_{ATE}(x) = \int_0^1 E[\Delta_i | X_i = x, \widetilde{u}_{Di} = u] du$$

$$\Delta_{TT}(x, P(z)) = P(z)^{-1} \int_0^{P(z)} E[\Delta_i | X_i = x, \widetilde{u}_{Di} = u] du$$

$$\Delta_{LATE}(x, P(z), P(z')) = (P(z) - P(z'))^{-1} \int_{P(z')}^{P(z)} E[\Delta_i | X_i = x, \widetilde{u}_{Di} = u] du$$

where $\widetilde{u}_{Di} \equiv F_{u_D}(u_{Di})$ denotes the on the interval $[0, 1]$ uniformly distributed probability transform of u_{Di}.

Each of the average treatment effects thus is an average of the marginal treatment effect for a well-specified range of the propensity score defined by the instruments considered. The average treatment effect is obtained by integrating the marginal treatment effect over the full support of $\tilde{u}_{Di}$, while the average treatment effect for the treated is the integral over the range $[0, P(z)]$ as defined by the given value of the instrument. According to Heckman and Vytlacil (1999), "it is primarily determined by the average effect for individuals whose unobserved characteristics make them the most inclined to participate in the program." The LATE, finally, integrates the marginal treatment effect over the range $[P(z'), P(z)]$ and is "the average treatment effect for someone who would not participate if $P(Z) \leq P(z')$ and would participate if $P(Z) > P(z)$." (ibidem)[12]

3.3 Causal inference

Causal inference attempts at identifying and estimating the above average treatment effects on the basis of a comparison between average outcomes across treatment and control groups. In order to identify the above effects, various identifying assumptions are necessary that allow the construction of counterfactuals from observations on the treatment group and the control group. Both the choice of the parameter of interest and the strength of the necessary identifying assumptions depend on the nature of the available data (experimental vs observational), their dimension (cross-section, longitudinal, or panel data), the (sub-)population of interest, and the underlying (implicit) experimental design (assignment mechanism and compliance behaviour), or, in economic parlance, the way in which individuals select into treatment.

This section discusses the identification of causal effects by instrumental variables techniques. It reviews in particular the assumptions under which instrumental variables estimates have a causal interpretation in the simple case of a binary assignment-binary treatment model with a continuous outcome variable.

3.3.1 The econometric model

As defined above, the general potential outcome model is given by $(Z_i, (D_{0i}, D_{1i}), (Y_{0i}, Y_{1i}))_{i \in I}$ with potential outcomes $Y_{0i} = \mu_0(X_i) + u_{0i}$ and $Y_{1i} = \mu_1(X_i) + u_{1i}$, parameter of interest $\Delta_i \equiv (Y_{1i} - Y_{0i})$, and observability rules $D_i = D_{0i}(1 - Z_i) + D_{1i}Z_i$ and $Y_i = Y_{0i}(1 - D_i) +$

[12] For empirical applications of this concept cf. e.g. Aakvik (1999) and Aakvik et al. (1999).

$Y_{1i}D_i = Y_{0i} + \Delta_i D_i$. Its econometric implementation is generally based on the so-called *dummy endogenous variable model* (with either fixed or random coefficient) or the *additively separable correlated random coefficient model.* Consider a simple simultaneous equations specification including an outcome equation

$$Y_i = X_i'\beta + D_i\delta_i + u_i \tag{3.9}$$

and a participation (or selection) equation

$$D_i = X_i'\alpha + Z_i'\gamma + \nu_i \tag{3.10}$$

This model is interpreted as a *heterogeneous treatment effect model* with outcome variable Y_i, treatment indicator D_i, assignment-to-treatment indicator Z_i, and (unit-level) treatment effect δ_i. For the estimation of the *average* causal effect of the treatment D_i on the potential outcome Y_i, a fixed coefficient specification of the heterogeneous treatment effect is imposed, interpreting δ as some average treatment effect of D_i on Y_i, conditional on individual characteristics X_i.[13]. Because of the selection (or endogeneity) bias due to $E[D_i u_i] = E[u_i \,|D_i = 1] \neq 0$, two difficulties arise in causal inference: first, to estimate consistently δ, and second, to ensure the interpretability of the estimate obtained as an average causal effect in the above potential outcome model. Several methods can be applied which aim at taking into account both of the above problems directly[14]: first, control for confounding variables, using regression or matching techniques (matched treatment-control-comparisons), second, pre-post comparisons on the same units of observation to reduce bias from unobserved differences (and in a similar vein: fixed-effects models and difference-in-difference estimates), and third, instrumental variables estimation, interpreted in a quasi-experimental framework. In the sequel, attention will be restricted to the identification of causal effects by instrumental variables methods, adopting the view expressed *inter alia* by Lechner (1998a,b) that nonparametric identification strategies

[13] Note that, assuming $\delta_i \equiv \delta$ for all $i \in I$ would implicitly assume $\Delta_i = \Delta$ for all $i \in I$, and hence impose equality between all average treatment effects ($\Delta_{ATE} = \Delta_{TT} = \Delta_{LATE}$) defined above.

[14] Cf. e.g. Angrist and Krueger (1998).

are a "more credible way of identifying [causal] effects."

3.3.2 Identification of causal effects by instrumental variables techniques

A well-known solution to the problem of endogenous treatment status D_i in the above outcome equation 3.9 consists in the use of instrumental variables as a source of exogenous variation in D_i. These instruments can be interpreted as generating a quasi-*natural experiment* by assigning individuals to treatment independently of unobserved characteristics.[15] Estimates are obtained from comparing the outcomes across groups which differ with respect to the value of an instrumental variable that is related to the outcome of interest only by virtue of correlation with the probability of treatment. Instrumental variables techniques consistently estimate δ in the above model in the presence of an endogenous regressor D_i if - loosely speaking - conditional on X_i, Z_i is correlated with the outcome Y_i only through treatment status D_i, i.e. if conditions (IV1) $p\lim N^{-1}D'Z = \Sigma > 0$ and (IV2) $p\lim N^{-1}Z'u = 0$ hold. In the special case without covariates, the IV estimator has the simple Wald estimator as probability limit, where the Wald estimator is defined as the ratio of the causal effect of the assignment on the outcome $(E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0])$ and the causal effect of the assignment on the treatment status $(E[D_i|Z_i = 1] - E[D_i|Z_i = 0])$.

Identification of causal effects hence requires the existence of a variable Z_i that affects the selection into treatment but does not affect the outcome directly. However, even if IV estimates are consistent for δ, it is a highly controversial matter in the literature under what (further implicit) conditions this estimate does actually identify one of the above-mentioned average treatment effects of interest.

The following theorems each state the respective conditions for the instrumental variables estimate to have a causal interpretation as an average treatment effect on a randomly selected person (ATE), an average treatment effect on the treated (TT), and a local average treatment effect (LATE).

[15] Cf. Angrist (1998): "In labor economics, at least, the current popularity of quasi-experiments stems precisely from this concern: because it is typically impossible to control for all relevant variables it is often desirable to seek situations where one has a reasonable presumption that the omitted variables are uncorrelated with the variables of interest. Such situations may arise if the researcher can use random assignment, or if the forces of nature or human institutions provide something close to random assignment."

Theorem 10 *(Heckman (1997), conditions for a causal interpretation of linear IV estimates in the heterogeneous treatment effect model as average treatment effect on a randomly selected person)*

(1) "non-zero treatment effect"

$E[D_i|Z_i, X_i] = \Pr[D_i = 1|Z_i, X_i] \neq 0$, *i.e. the treatment probability is a nontrivial function of* Z_i *given* X_i.

(2) "exclusion restriction"

The instrumental variable Z_i *is mean-independent of the idiosyncratic components* u_{0i} *and* u_{1i} *(and hence of the potential outcomes* Y_{0i} *and* Y_{1i}*), i.e.* $E[u_{0i} + D_i(u_{1i} - u_{0i})|X_i, Z_i] = 0$.

This latter condition implies:

(2a) $E[u_{0i}|Z_i, X_i] = E[u_{0i}|X_i] = 0$, *i.e. the instrument is uncorrelated with the potential idiosyncratic component of the potential outcome in the case of non-treatment.*

(2b) $E[D_i(u_{1i} - u_{0i})|X_i, Z_i] = E[u_{1i} - u_{0i}] = 0$, *i.e. conditional on the instrument, the idiosyncratic gain is uncorrelated with the treatment.*

Theorem 11 *(Heckman (1997), conditions for a causal interpretation of linear IV estimates in the heterogeneous treatment effect model as average treatment effect on the treated)*

The instrumental variable estimate of δ *in the heterogeneous treatment effect model has a causal interpretation as average treatment effect for the treated under the following assumptions:*

(1) "non-zero treatment effect"

$E[D_i|Z_i, X_i] = \Pr[D_i = 1|Z_i, X_i] \neq 0$, *i.e. the treatment probability is a nontrivial function of* Z_i *given* X_i.

(2) "exclusion restriction"

The instrumental variable Z_i *is mean-independent of the idiosyncratic gains* u_{0i} *and* u_{1i} *(and hence of the potential outcomes* Y_{0i} *and* Y_{1i}*), i.e.*

$E\{u_{0i} + D_i(u_{1i} - u_{0i}) - E[u_{1i} - u_{0i}|D_i = 1, X_i]|X_i, Z_i\} = 0.$

This latter condition implies:

(2a) $E[u_{0i}|Z_i, X_i] = E[u_{0i}|X_i] = 0$, *i.e. the instrument is uncorrelated with the potential idiosyncratic component of the potential outcome in the case of non-treatment.*

(2b) $E[u_{1i} - u_{0i}|D_i = 1, Z_i, X_i] = E[u_{1i} - u_{0i}|D_i = 1, X_i]$, *i.e. conditional on the treatment, the idiosyncratic gain is uncorrelated with the instrument.*

In many applications based on observational studies, the implicit assumptions underlying theorems 10 or 11 do not seem reasonable. This applies in particular to assumption (2b) which states that the average idiosyncratic gains of those who are assigned to treatment and do comply with this assignment equal the average idiosyncratic gains in the whole population or in the group of all the treated, respectively. Since these groups are relatively heterogeneous and since, in most economic applications, individuals are likely to self-select into treatment according to their expected idiosyncratic gains from doing so, these implied assumptions do not seem warranted. The fact that these assumptions are not needed for a causal interpretation of linear IV estimates as local average treatment effects is one of the main differences regarding the identifiability of the treatment effect of the treated, on the one hand, and the local average treatment effect, on the other. It is also, however, one of the main reasons for conceptual controversies in the field of causal inference as will be discussed briefly below.

Theorem 12 *(AIR (1996), conditions for a causal interpretation of linear IV estimates in the heterogeneous treatment effect model as local average treatment effect)*[16]

The instrumental variable estimate of δ in the heterogeneous treatment effect model has a causal interpretation as local average treatment effect under the following assumptions:

(1) "stable unit treatment value assumption" (SUTVA)

Potential outcomes for each individual i are unrelated to the treatment status of other individuals.

(2) "strongly ignorable assignment to treatment"

Conditional on observables, the assignment to treatment is random, i.e. $\Pr(Z_i = 1\,|X_i) = \Pr(Z_i = 0\,|X_i)$, *or equivalently, treatment status is independent of potential outcomes.*

(3) "strong monotonicity"

The treatment probability is a nontrivial and monotonic function of the instrument, i.e.

$E\left[D_{1i} - D_{0i}\,|X_i\right] > 0.$

[16] *Cf. Angrist et al. (1996), here referred to as AIR(1996). Cf. also Imbens and Angrist (1994), Angrist and Imbens (1995), and Angrist and Krueger (1998). For a different point of view, cf. e.g. Heckman (1997) who shows that in the heterogenous treatment effect model the instrumental variable estimate of δ_{ATE} only identifies a causal effect if individuals do not self-select into the treatment status on the basis of idiosyncratic gains to treatment. If this strong condition is fulfilled, the identified effect is necessarily the average treatment effect for the treated. The author generally questions the validity and usefulness of the LATE.*

(4) "exclusion restrictions"

The (unit-level) potential outcome variables depend on the assignment status Z_i only through the treatment status D_i, i.e. $(Y_{0i}, Y_{1i}) \perp Z_i \,|D_i$.

Under certain assumptions, causal effects in the above heterogeneous treatment effect model can thus be estimated by means of standard two-stage least squares methods for a given appropriate instrument. The more difficult question is which of the above average treatment effects can be identified by IV estimation. The answer hinges on the assumptions - implicit or explicit - one is willing to make on the assignment and selection mechanisms underlying the given quasi-natural experiment or - as Heckman (1997) puts it - on the way in which individuals "process information". While e.g. AIR (1996) argue that the only treatment effect that IV can consistently estimate is the local average treatment effect, Heckman (1997) argues that IV can at best identify the average treatment effect for the treated, or no treatment effect of interest at all.[17] In more recent work, Heckman and Vytlacil (2000a,b,c) suggest to estimate treatment parameters in a latent variable framework when they are identified or bound them otherwise. While, contrary to the identification of local average treatment effects, no assumptions on monotonicity or strong ignorability are required for the identification of such marginal treatment effects, statistical support conditions for the propensity score are required which are difficult to interpret economically. Furthermore, even for bounding treatment parameters, problematic assumptions of known lower and upper bounds on potential outcomes have to be invoked.

These controversies are problematic since, in the presence of heterogeneity, the above treatment effects will differ significantly. Since the assumptions underlying the above theorems generally refer to unobservable individual-level potential outcomes and treatment status and are thus untestable,[18] the interpretation of IV estimates as one or the other causal effect has

[17] With respect to the case where the average treament effect for the treated and the local average treatment effect differ, Heckman (1997) states: "In the likely case in which individuals possess and act on private information about gains from a program that cannot be fully predicted by variables in the outcome equation, instrumental variables methods do not estimate economically interesting evaluation parameters." He further adds that "[a]ny valid application of the method of instrumental variables for estimating (...) treatment effects in the case where the response to treatment varies among persons requires a behavioral assumtion about how persons make their decisions about program participation." In Heckman's (1997) view, this issue "cannot be settled by statistical analysis."

[18] Cf. Lechner (1998a) on test procedures suggested by e.g. Heckman and Hotz (1989) for the evaluation of

to rely entirely on the presumed plausibility of these identifying assumptions. Even here, however, there are strong differences in the literature: while AIR (1996) consider the local average treatment effect as the only treatment effect of interest which can effectively be estimated by instrumental variables methods, Heckman (1997) claims that this would imply both "ignorance or irrationality about unobserved components of gain on the part of the people being studied" and absence of "private information that is useful in forecasting the gains that they use in making their decisions but that is not available to the analyst."

The predominant role of inherently untestable identifying assumptions notwithstanding, there exists an agreement in the literature to provide diagnostics on the instrument relevance. Following the suggestions by Bound et al. (1995) and Staiger and Stock (1997), first-stage F-statistics and partial R^2-measures will be reported in the later empirical applications. Other methods have been suggested to investigate in particular the plausibility of the exclusion restriction[19] but these will not discussed here any further.

3.3.3 Extensions

The previous section focussed on the simple case of a binary assignment-binary treatment model with a continuous outcome variable. The methods discussed, however, can in general be adapted easily to more general outcome variables such as discrete or non-negative outcomes such as frequencies, counts or durations.[20] Other extensions of interest are, first, the estimation of

different sample selection models: "Test procedures do only work when there are overidentifying restrictions to test. It should be clear by now that the fundamental lack of identification inherent in causal analysis requires identifying restrictions in the first place. Such restrictions are never testable."

[19] Cf. e.g. Blomquist and Dahlberg (1999), Angrist and Krueger (1995) and Angrist et al. (1995).

[20] Angrist (2000) e.g. argues that "the dificulty with endogenous variables in non-linear LDV models is, in fact, more apparent than real." According to him, at least for binary endogenous regressors, technical challenges related to LDV models "come primarily from what I see as a counterproductive focus on structural parameters such as latent index coefficients or censored regression coefficients, instead of directly interpretable causal effects."

The main question of interest in this context is whether standard estimation procedures such as 2SLS remain applicable in the case of dependent variables which do not have a continuous support. While Angrist (2000) presents arguments in favour of both, standard linear IV methods as well as semi-parametric nonlinear IV estimators, he argues in particular that any of these approaches have two main advantages over structural models based on a latent index or censored regression framework: computational simplicity and weak identification requirements.

For evaluation studies with discrete or non-negative outcomes, cf. e.g. Angrist (1998, 2000), Card and Sullivan (1998) (binary outcome variables); Angrist (2000), Cameron and Trivedi (1998) and Mullahy (1998) (count data); Abbring and van den Berg (1999), Angrist (2000), Ham and LaLonde (1996), Hujer et al. (1996, 1997a,b, 1999) and Ridder (1986,2000) (duration data); and Cockx and Bardoulat (1999), Ham and LaLonde (1990) and Lubyova and van Ours (1998) (transition rates).

full outcome distributions for compliers, and second, the identification and estimation of causal effects in the case of multivalued treatments.

Estimation of full outcome distributions for compliers

As shown in Imbens and Rubin (1997), under the assumptions stated in theorem 12 it is possible to estimate the entire marginal outcome distributions by treatment status, Y_{0i} and Y_{1i}, for the subgroup of compliers. These distributions are implicitly estimated by the IV estimator of local average treatment effects - generally without imposing any non-negativity restrictions. The estimated outcome distributions for compliers can be used as a diagnostic tool - in particular with a view to checking the plausibility of the exclusion restrictions. The authors further show that IV estimates based on the restriction of non-negativity of these outcome distributions might differ substantially from the unrestricted estimates. A comparison of the unrestricted and restricted IV estimates of local average treatment effects therefore represents an additional intuitive diagnostic check in causal inference.

Observe first that one can estimate both the population proportions and the outcome distributions for the two groups of *non-compliers* (always-takers and never-takers). Let ϕ_{NT} and ϕ_{AT} denote the population proportions of these two groups, and let g_{NT} and g_{AT} denote their outcome distributions. Let further f_{zd} denote the directly estimable outcome distribution in the subsample defined by $z_i = z$ and $d_i = d$. Then,

$$\phi_{NT} = \Pr(d_i = 0 \,|z_i = 1) \tag{3.11}$$

$$\phi_{AT} = \Pr(d_i = 1 \,|z_i = 0) \tag{3.12}$$

and

$$g_{NT}(y_i) = f_{01}(y_i) = \Pr(y_i \,|d_i = 0, z_i = 1) \tag{3.13}$$

$$g_{AT}(y_i) = f_{10}(y_i) = \Pr(y_i | d_i = 1, z_i = 0) \tag{3.14}$$

are directly estimable from the observed data $((z_i, d_i, y_i), X_i)_{i=1,...,N}$.

From independence of assignment and compliance behaviour it follows that the directly estimable sampling distributions f_{00} and f_{11} are mixtures of the above (directly estimable) distributions for never-takers or always-takers, on the one hand, and of the (not directly estimable) outcome distributions for compliers, on the other. Letting ϕ_C denote the population proportion of compliers and g_{1C} and g_{0C} the potential outcome distributions for compliers with and without treatment, respectively,[21] one can thus write

$$f_{00}(y_i) = \frac{\phi_C}{\phi_C + \phi_{NT}} g_{0C}(y_i) + \frac{\phi_{NT}}{\phi_C + \phi_{NT}} g_{NT}(y_i) \tag{3.15}$$

$$f_{11}(y_i) = \frac{\phi_C}{\phi_C + \phi_{AT}} g_{1C}(y_i) + \frac{\phi_{AT}}{\phi_C + \phi_{AT}} g_{AT}(y_i) \tag{3.16}$$

It follows that

$$\phi_C = 1 - \phi_{NT} - \phi_{AT} \tag{3.17}$$

and

$$g_{0C}(y_i) = \frac{\phi_C + \phi_{NT}}{\phi_C} f_{00}(y_i) - \frac{\phi_{NT}}{\phi_C} f_{10}(y_i) \tag{3.18}$$

[21] For the above subpopulations of always-takers and never-takers only one distribution had to be defined in each case since it is conceptually impossible to envisage either treatment of a never-taker or non-treatment of an always-taker.

$$g_{1C}(y_i) = \frac{\phi_C + \phi_{AT}}{\phi_C} f_{11}(y_i) - \frac{\phi_{AT}}{\phi_C} f_{01}(y_i) \tag{3.19}$$

On the basis of these relationships, by replacing the population values by their sample estimates, one can thus estimate the full (marginal) outcome distributions for compliers, Y_{0i} and Y_{1i}, as well as their means $E\left[Y_{0i}\,|D_{1i} - D_{0i} = 1\right]$ and $E\left[Y_{1i}\,|D_{1i} - D_{0i} = 1\right]$ and not just their difference equal to $\Delta_{LATE} = E\left[Y_{1i} - Y_{0i}\,|D_{1i} - D_{0i} = 1\right]$. In chapter 4, the results of the estimation of the full (marginal) potential outcome distributions for compliers will be discussed as part of the discussion on the plausibility of the identifying assumptions and the robustness of the results obtained.

Variable treatment intensity

This section extends the above quasi-experimental Rubin's Causal Model to the case of multi-valued treatments - or so-called variable treatment intensity, building on Angrist and Imbens (1995). Let Y_{ji} denote the potential outcome of interest for individual (or, equivalently, *unit*) $i \in I$ in case of treatment $j = 0, 1, ..., J$. As above, Y_{0i} denotes the potential outcome in the absence of treatment, and Y_{ji}, $j = 1, ..., J$, the potential outcome in case the multivalued treatment variable takes the value j. In the later empirical analysis, j will indicate the number of years of schooling and Y_{ji} the potential earnings of individual i for j years of schooling. As above, Y_i is the observed outcome variable for individual i. Let further $D_{ki} \in \{0, 1, ..., J\}$ denote the individual's potential *(multivalued) treatment status* as a function of the binary assignment status Z_i. As before, Z_i takes the value $k = 1$ if individual i is assigned to the treatment group and $k = 0$ otherwise. The observed cross-sectional data is $((z_i, d_i, y_i), X_i)_{i=1,...,N}$, where $d_i = d_{0i}$ in case of non-assignment and $d_i = d_{1i}$ in case of treatment and where further $y_i = y_{0i}$ in the case of non-assignment and $y_i = y_{ji}$ in the case of treatment $j \in \{1, 2, ..., J\}$.

In this framework, both the subpopulations and the parameters of interest are defined with respect to a specific value j of the multivariate treatment variable. The unit-level treatment effect of treatment j on outcome Y_i is given by $\Delta_{ji} = Y_{ji} - Y_{j-1,i}$ - the causal effect of an additional jth year of schooling in the later empirical analysis.

Estimates of δ in the outcome equation 3.9 have a causal interpretation only if they have probability limit equal to a weighted average of $E[Y_{ji} - Y_{j-1,i}]$ for all j in the respective subpopulation of interest. Furthermore, as in the case of bivariate treatments, the linear IV estimate identifies a causal effect for a well-defined subpopulation.

Theorem 13 *(Angrist and Imbens (1995), conditions for a causal interpretation of linear IV estimates in the heterogeneous treatment effect model with multivalued treatment as weighted average of per-unit average causal effects)*

Assume

(1) "independence"

The (unit-level) potential outcome and treatment variables $D_{0i}, D_{1i}, Y_{0i}, Y_{1i}, ..., Y_{Ji}$ are jointly independent of Z_i.

(2) "strict monotonicity"

With probability 1, either $D_{1i} - D_{0i} \geq 0$ or $D_{1i} - D_{0i} \leq 0$ for each individual $i \in I$ and $\Pr(D_{1i} \geq j > D_{0i}) > 0$ for at least one $j = 0, 1, ..., J$.

Then, the probability limit of the linear 2SLS-IV estimator is given by

$$\underset{N\to\infty}{p\lim}\,\widehat{\delta} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} = \sum_{j=1}^{J} \omega_j \cdot r(j) \equiv \delta \tag{3.20}$$

where

$$\omega_j = \frac{\Pr(D_{1i} \geq j > D_{0i})}{\sum_{i=1}^{J} \Pr(D_{1i} \geq i > D_{0i})} \tag{3.21}$$

and

$$r(j) \equiv E[Y_{ji} - Y_{j-1,i}|D_{1i} \geq j > D_{0i}]. \tag{3.22}$$

The fact that $0 \leq \omega_j \leq 1$ and $\sum_{j=1}^{J} \omega_j = 1$ implies that δ is a weighted average of average causal responses to a unit change in the treatment value for those whose treatment status is affected by the instrument, $E[Y_{ji} - Y_{j-1,i}|D_{1i} \geq j > D_{0i}]$. The weight ω_j is proportional to the number of compliers, $\Pr(D_{1i} \geq j > D_{0i})$. Angrist and Imbens (1995) refer to the parameter δ as the *average causal response* (ACR). The *weights* ω_j reflect the weight of the subpopulation in calculating this average causal response.$r(j)$ further shows the weights of the respective schooling levels in computing the average treatment effect.

To identify a meaningful average treatment effect, the literature typically assumes a constant unit treatment effect, $Y_{ji} - Y_{j-1,i} = \alpha$, for all $j = 0, 1, ..., J$ and all $i \in I$. By means of the above monotonicity assumption, however, Angrist and Imbens (1995) impose a nonparametric restriction on the process determining D as a function of Z instead of restricting treatment effect heterogeneity. The authors further show that, for multivalued treatments, this assumption has the testable implication that the respective conditional cumulative distribution functions of D given $Z = 1$ and $Z = 0$ should not cross.

This approach will be applied in the analysis of returns to schooling in Germany in chapter 5. It allows in particular to take into account the following facts: first, different subgroups are affected by different instruments; second, individuals in these subgroups are affected by the respective instrument in different ways; and third, the instrument may induce changes of behavior at different levels of schooling.

3.4 Summary and conclusions

This chapter - in the form of a commented literature review - has presented Rubin's Causal Model and its extension to multivalued treatment variables as a general framework for the identification and estimation of causal effects in observational studies. This model allows naturally for a focus on causal effects of individual decisions - job mobility during transition or educational choices. Empirical applications of Rubin's Causal Model below will be based on the identification results reviewed in this chapter, in particular on those stated in theorem 12. Both on theoretical grounds (following the results of chapter 2) and because of data availability, the following empirical applications will concentrate on the identification and estimation of local average treatment effects in the case of, first, binary treatment (job mobility), and second, variable treatment intensity (schooling).

3.5 References

Aakvik, A. (1999), Assessing the effects of labour market training in Norway, University of Bergen, Department of Economics, mimeo

Aakvik, A., J.J. Heckman, and E.J. Vytlacil (1999), Training effects on employment when the training effects are heterogeneous: An application to Norwegian vocational rehabilitation programs, University of Chicago, Department of Economics, mimeo

Abadie, A., J.D. Angrist and G.W. Imbens (1998), Instrumental variables estimation of quantile treatment effects, NBER Technical Working Paper No. 229

Abbring, J.H. and G.J. van den Berg (1999), The non-parametric identification of treatment effects in duration models, Vrije Universiteit Amsterdam, Faculteit der Economische Wetenschappen en Econometrie, mimeo

Ahn, H. and J.L. Powell (1993), Semi-parametric estimation of censored selection models with a nonparametric selection mechanism, Journal of Econometrics, 58, 3-29

Angrist, J.D. (1990), Lifetime earnings and the Vietnam era draft lottery: Evidence from social security administrative records, American Economic Review, 80(3), 313-336

Angrist, J.D. (1998), Estimating the labor market impact of voluntary military service using social security data on military applicants, Econometrica, 66(2), 249-288

Angrist, J.D. (2000), Estimation of limited-dependent variable models with dummy endogenous regressors: Simple strategies for empirical practice, NBER Technical Working Paper No. 248

Angrist, J.D. and G.W. Imbens (1995), Two-stage least squares estimation of average causal effects in models with variable treatment intensity, Journal of the American Statistical Association, 90(430), 431-442

Angrist, J., G.W. Imbens and A.B. Krueger (1995), Jackknife instrumental variables estimation, NBER Technical Working Paper No. 172

Angrist, J., G.W. Imbens and D.B. Rubin (1996), Identification of causal effects using instrumental variables, Journal of the American Statistical Association, 91(434), 444-472

Angrist, J.D. and A.B. Krueger (1991), Does Compulsory School Attendance Affect Schooling and Earnings, Quarterly Journal of Economics, 106, 979-1014

Angrist, J.D. and A.B. Krueger (1995), Split-sample instrumental variables estimates of the return to schooling, Journal of Business and Economic Statistics, 13, 225-235

Angrist, J.D. and A.B. Krueger (1998), Empirical strategies in labor economics, Massachusetts Institute of Technology Working Paper No. 98-7, Cambridge, forthcoming in: O. Ashenfelter and D. Card (eds) (1999), Handbook of Labor Economics, Vol. 3A, North-Holland, Amsterdam

Blomquist, S. and M. Dahlberg (1999), Small sample properties of LIML and Jackknife IV estimators: Experiments with weak instruments, Journal of Applied Econometrics, 14, 69-88

Bound, J., D.A. Jaeger and R.M. Baker (1995), Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak, Journal of the American Statistical Association, 90(430), 443-450

Cameron, A.C. and P.K. Trivedi (1998), Nonrandom samples and simultaneity, in: Regression Analysis of Count Data, Econometric Society Monograph No. 30, Cambridge University Press, Cambridge, chapter 11

Card, D. (1995), Using Geographic Variation in College Proximity to Estimate the Return to Schooling, in: L.N. Christofides, E.K. Grant, and R. Swidinsky (eds), Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp, University of Toronto Press, Toronto, 201-222

Card, D. (1999), The Causal Effect of Education on Earnings, in: O. Ashenfelter and D. Card (eds), Handbook of Labor Economics, Vol. 3, 1801-63

Card, D. and D. Sullivan (1988), Measuring the effect of subsidised training programs on movements in and out of employment, Econometrica, 56(3), 497-530

Chen, S. (1999), Distribution-free estimation of the random coefficient dummy endogenous variable model, Journal of Econometrics, 91, 171-199

Cockx, B. and I. Bardoulat (1999), Vocational training: Does it speed up the transition rate out of unemployment?, Université Catholique de Louvain, IRES and Département des Sciences Economiques et Sociales

Cox, D.R. (1958), The planning of experiments, Wiley, New York

Fisher, R. (1935), The design of experiments, Oliver and Boyd, London

Fitzenberger, B. and H. Prey (1996), Training in East Germany: An evaluation of the effects on employment and wages, Center for International Labor Economics Discussion Paper

No. 36-1996, Universität Konstanz

Fitzenberger, B. and H. Prey (1997), Assessing the impact of training on employment: The case of East Germany, ifo-Studien, 43, 71-116

Froehlich, M. (1998), Semiparametric estimation of selectivity models, Diplomarbeit, Universitaet Konstanz

Gronau, R. (1974), Wage comparisons: A selectivity bias, Journal of Political Economy, 82, 1119-1143

Ham, J.C. and R.J. LaLonde (1990), Using social experiments to estimate the effect of training on transition rates, in: J. Hartog, G. Ridder, and J. Theeuwes (eds), Panel data and labor market studies, North-Holland, Amsterdam, 152-172

Ham, J.C. and R.J. LaLonde (1996), The effect of sample selection and initial conditions in duration models: Evidence from experimental data on training, Econometrica, 64, 175-205

Heckman, J.J. (1976), The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator for such models, Annals of Economic and Social Measurement, 5, 479-492

Heckman, J.J. (1978), Dummy endogenous variables in a simultaneous equation system, Econometrica, 46(6), 931-960

Heckman, J.J. (1979), Sample selection bias as a specification error, Econometrica, 47(1), 153-161

Heckman, J.J. (1990), Varieties of selection bias, Econometrica, Papers & Proceedings, 80(2), 313-318

Heckman, J.J. (1992), Randomization and social policy evaluation, in: C.F. Manski and I. Garfinkel (eds), Evaluating welfare and training programs, Harvard University Press, Cambridge MA

Heckman, J.J. (1997), Instrumental variables, A study of implicit behavioral assumptions used in making program evaluations, Journal of Human Resources, 441-462

Heckman, J.J. (1999), Causal parameters and policy analysis in economics: A twentieth century perspective, NBER Working Paper No. 7333

Heckman, J.J. and B. Honoré (1990), The empirical content of the Roy model, Econometrica, 58(5), 1121-1149

Heckman, J.J. and V.J. Hotz (1989), Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training, Journal of the American Statistical Association, 84, 862-880

Heckman, J.J., R.J. LaLonde, and J.A. Smith (1999), The economics and econometrics of active labor market programs, in: O. Ashenfelter and D. Card (eds), Handbook of Labor Economics, Vol. 3, North-Holland, Amsterdam

Heckman, J.J. and R. Robb (1985a), Alternative methods for the evaluating the impact of interventions: An overview, Journal of Econometrics, 30, 239-267

Heckman, J.J. and R. Robb (1985b), Alternative methods for the evaluating the impact of interventions, in: J.J. Heckman and B. Singer (eds), Longitudinal analysis of labor market data, Econometric Society Monograph No. 10, Cambridge University Press, New York, 156-245

Heckman, J.J. and R. Robb (1986a), Alternative methods for solving the problem of selection bias in evaluating the impact of treatment on outcomes, in: H. Wainer (ed.), Drawing inferences from self-selected samples, Springer-Verlag, Berlin

Heckman, J.J. and E.J. Vytlacil (1999), Local instrumental variables and latent variable models for identifying and bounding treatment effects, Proceedings of the National Academy of Sciences, 96, 4730-4734

Heckman, J.J. and E.J. Vytlacil (2000a), The relationship between treatment parameters within a latent variable framework, Economics Letters, 66(1), 33-39

Heckman, J.J. and E.J. Vytlacil (2000b), Local instrumental variables, in: C. Hsiao, K. Morimune, and J. Powell (eds), Nonlinear Statistical Inference: Essays in honor of Takeshi Amemiya, Cambridge University Press

Heckman, J.J. and E.J. Vytlacil (2000c), Econometric evaluations of social programs, in: J.J. Heckman and E. Leamer (eds), Handbook of Econometrics, Vol. 5, North-Holland, Amsterdam

Holland, P.W. (1986), Statistics and causal inference (with discussion), Journal of the American Statistical Association, 81, 945-960

Hotz, V.J. (1992), Designing an evaluation of the Job Training Partnership Act, in: C.F. Manski and I. Garfinkel (eds), Evaluating welfare and training programs, Harvard University Press, Cambridge MA

Hübler, O. (1997), Evaluation beschäftigungspolitischer Maßnahmen in Ostdeutschland, Jahrbücher für Nationalökonomie und Statistik, 216, 21-44

Hujer, R., K.-O. Maurer, and M. Wellner (1996), The impact of training on employment: A survey of microeconometric studies, Discussion Papers in Economics No. 69, Universität Frankfurt/Main

Hujer, R., K.-O. Maurer, and M. Wellner (1997a), Estimating the effects of training on unemployment duration in West Gemany: A discrete hazard-rate model with instrumental variables, Discussion Papers in Economics No. 73, Universität Frankfurt/Main

Hujer, R., K.-O. Maurer, and M. Wellner (1997b), The impact of training on unemployment duration in West Germany: Combining a discrete hazard-rate model with matching techniques, Discussion Papers in Economics No. 74, Universität Frankfurt/Main

Hujer, R., K.-O. Maurer, and M. Wellner (1999), Estimating the effect of vocational training on unemployment duration in West Germany, Jahrbücher für Nationalökonomie und Statistik, 218(5/6), 619-646

Ichino, A. (1999), The problem of causality in the analysis of educational choices and labor market outcomes, Lecture notes, European University Institute, Florence, mimeo

Ichino, A. and R. Winter-Ebmer (1999), Lower and Upper Bounds of Returns to Schooling: An Exercise in IV Estimation with Different Instruments, European Economic Review, 43(4-6), 889-901

Ichino, A. and R. Winter-Ebmer (2000), The Long-Run Educational Cost of World War Two, European University Institut, Florence, mimeo

Imbens, G.W. and J.D. Angrist (1994), Identification and estimation of local average treatment effects, Econometrica, 62(2), 467-475

Imbens, G.W. and D.B. Rubin (1997), Estimating outcome distributions for compliers in instrumental variables models, Review of Economic Studies, 64, 555-574

Kraus, F., P.A. Puhani, and V. Steiner (1997), Employment effects of publicly financed training programmes: The East German experience, ZEW Discussion Paper No. 97-33, Mannheim

LaLonde, R. (1986), Evaluating the econometric evaluations of training programs with experimental data, American Economic Review, 76, 604-620

LaLonde, R. (1995) The promise of public sector-sponsored training programmes, Journal of Economic Perspectives, 9(2) ,149-168

Lechner, M. (1998a), Training the East German labour force: Microeconometric evaluations of continuous vocational training after unification, Studies in Contemporary Economics, Physica-Verlag, Heidelberg

Lechner, M. (1998b), Mikrooekonometrische Evaluationsstudien, Anmerkungen zu Theorie und Praxis, in: F. Pfeiffer and W. Pohlmeier (eds), Qualifikation, Weiterbildung und Arbeitsmarkterfolg, ZEW-Wirtschaftsanalysen, Nomos-Verlag, Baden-Baden

Lechner, M. (1999), Earnings and employment effects of continuous off-the-job training in East Germany after unification, Journal of Business and Economic Statistics, 17, 74-90

Lubyova, M. and J.C. van Ours (1998), Effects of active labour market market programmes on the transition rate from unemployment into regular jobs in the Slovak Republic, CentER Discussion Paper No. 98127, Tilburg University

Manski, C.F. (1990), Nonparametric bounds on treatment effects, Econometrica, Papers & Proceedings, 80(2), 319-323

Manski, C.F. (1993), The selection problem in econometrics and statistics, in: C.R. Rao and G.S. Maddala (eds), Handbook of Statistics, Vol. 11, Econometrics, Elsevier Science Publishers, 73-84

Manski, C.F. (1994), The selection problem, in: C. Sims (ed.), Advances in econometrics, Cambridge University Press, New York, 143-170

Manski, C.F. (1995), Identification problems in the social sciences, Harvard University Press, Cambridge MA

Meyer, B. (1995a), Lessons from the U.S. unemployment insurance experiments, Journal of Economic Literature, 33(2), 91-131

Meyer, B. (1995b), Natural and quasi-natural experiments in economics, Journal of Business and Economic Statistics, 13(2), 151-161

Mullahy, J. (1998), Instrumental variables estimation of count data models: Applications to models of cigarette smoking behavior, Review of Economics and Statistics, 11, 586-593

Neyman, J. (1923), Statistical problems in agricultural experiments, Supplement to the Journal of the Royal Statistical Society, 2(2), 107-180

Puhani, P.A. (1998), Advantage through training?, A microeconometric evaluation of the employment effects of active labour market programmes in Poland, ZEW Discussion Paper No.

98-25, Mannheim

Puhani, P.A. (1999), What works? An evaluation of active labour market policies in Poland during transition, ZEW Economic Studies, Mannheim

Quandt, R. (1972), Methods for estimating switching regressions, Journal of the American Statistical Association, 67(338), 306-310

Ridder, G. (1986), An event history approach to the evaluation of training, recruitment, and employment programmes, Journal of Applied Econometrics, 11, 109-126

Ridder, G. (2000), Identification of the effect of treatment on the treated, lecture notes, Johns Hopkins University, Department of Economics, Baltimore

Roy, A.D. (1951), Some thoughts on the distribution of earnings, Oxford Economic Papers, 3, 135-146

Rosenbaum, P.R. (1995), Observational Studies, Springer Series in Statistics, Springer-Verlag, Heidelberg, New York

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. (1977), Assignment to treatment group on the basis of a covariate, Journal of Educational Statistics, 2, 1-26

Rubin, D.B. (1978), Bayesian inference for causal effects: The role of randomization, Annals of Statistics, 6, 34-58

Rubin, D.B. (1990), Formal models of statistical inference for causal effects, Journal of Statistical Planning and Inference, 25, 279-292

Rubin, D.B. (1991), Practical implications of models of statistical inference for causal effects and the critical role of the assignment mechanism, Biometrics,6 47, 1213-1234

Schmidt, C.M. (1999), Knowing what works: The case for rigorous program evaluation, IZA Discussion Paper No. 77, Institute for the Study of Labor, Bonn

Smith, T.M.F. (1983), On the validity of inferences from non-random samples, Journal of the Royal Statistical Society, A, 146(4), 394-403

Staiger, D. and J.H. Stock (1997), Instrumental variables regression with weak instruments, Econometrica, 65, 557-586

Steiner, V. and F. Kraus (1995), Haben Teilnehmer an Arbeitsbeschaffungsmaßnahmen in Ostdeutschland bessere Wiederbeschäftigungschancen als Arbeitslose?, in: V. Steiner and L. Bellmann (eds), Mikroökonomik des Arbeitsmarktes, Beiträge aus der Arbeitsmarkt- und Berufsforschung No. 192, Nürnberg

Chapter 4

Identification and estimation of the returns to early job mobility during transition

4.1 Introduction

As described in chapter 1, the incidence of both job endings and job mobility in the East German transition process was dramatic. The massive job loss especially at the beginning of the transition period, when displacement rates in the East exceeded those in the West by a factor of 5 to 10, was accompanied by a considerable reallocation of labour through job mobility, with job mobility rates in the East generally doubling those in the West. Rates of both job losses and job changes peaked early in transition and declined considerably afterwards.

The considerably higher incidence of displacement in East Germany throughout the transition process suggests a much higher portion of so-called *forced movers* in East Germany as compared to the West. As discussed in chapter 2, the continuously high displacement rates are, moreover, likely to yield negative realised returns to early job mobility for those individuals who did not expect displacement rates on new jobs to remain at such a high level.

This chapter[1] estimates the returns to early job mobility in the transition process. Fol-

[1] A shortened version of this chapter was published as: Siebern, F. (2000), Better LATE? Instrumental

lowing the above observation that individuals are likely to self-select into specific forms of job mobility behaviour on the basis of their unobservable characteristics, standard results from both simple treatment-control comparisons of average income or from simple regressions of reduced form earnings equations do in general not identify any *causal effect* of job mobility. This chapter attempts to identify the *causal effects* of job mobility during transition in the quasi-experimental framework (*Rubin's causal model*) established in chapter 3 and to estimate the returns to early job mobility in the East German transition process for the period 1990-96 while taking explicitly into account the potential endogeneity problem. Job mobility early in the transition process is viewed as a *treatment* the effect of which on two different *outcome* variables - total income and employment probabilities throughout transition - is to be evaluated. Results from simple treatment-control comparisons and from standard OLS regressions are contrasted with those obtained from instrumental variables estimation, exploiting different sources of exogenous variation in the observed job mobility behaviour. Identification of the returns to job mobility is achieved through the choice of adequate instrumental variables without referring to either assumptions about the exogeneity of job mobility or distributional or functional form assumptions. Two different instruments are proposed which imply different *assignment-to-treatment* mechanisms in the quasi-experimental framework. The use of these instruments allows, moreover, a distinction between the returns to forced job mobility and those to voluntary job mobility without recurring to survey information on the nature of previous job losses.

The results obtained suggest both the presence of strong self-selection and the heterogeneity of returns to job mobility. Although suffering potentially from instrument weakness and finite sample bias, the results thus seem to question the validity of simple regression approaches for the estimation of returns to job mobility. They additionally represent an example in which the interpretation of instrumental variables estimates as *local average treatment effects* (*LATE*) as defined in Angrist et al. (1996) seems the most appropriate one.

The remainder of the chapter is organised as follows: Section 4.2 describes the database used and reviews the sample definitions. Section 4.3 defines the *treatment* and *outcome* variables

Variables Estimation of the Returns to Job Mobility During Transition, German Economic Review, 1(3), 335-362 used in the later analysis. Section 4.4 presents results from simple treatment-control comparisons and from OLS estimation. Section 4.5 discusses the choice of instruments in the light of the conditions for IV estimates to identify *causal effects* and reports the results from instrumental variables estimation. Section 4.6 presents results from sensitivity analyses and discusses the robustness of the results obtained with respect to alternative sample definitions, outcome measures or treatment variables. It also reports results from the estimation of full outcome distributions for compliers. Section 4.7 discusses the employment effects of early job mobility, contrasting again results from OLS estimation with IV estimates based on the above two instruments. Section 4.8 addresses some frequently asked questions regarding the identification and estimation of causal effects. Section 4.9 concludes.

4.2 Database and sample selection

The empirical analysis in this paper is based on data from the German Socio-Economic Panel (GSOEP) The GSOEP is a representative panel survey of the German population that was started in 1984 for the then Federal Republic of Germany and was extended to East Germany at the beginning of 1990. Most notably, it is the only German panel survey which managed to gather information on a representative East German sample before the start of the Economic, Monetary and Social Union (EMSU) in July of that year.[2]

For the purpose of the evaluation of job mobility during the transition from a centrally planned economy to a market economy this paper makes use of the East German subsample (subsample C) only. Information on this subsample is available for the years 1990 to 1997 with the first wave containing some important retrospective information for 1989.

In order to follow the same individuals over the first 8 years of the transition period the sample is restricted to those individuals with valid survey responses in all eight waves, yielding a balanced panel of 2696 individuals (denoted "full sample" in the sequel) of which more than 80% (2192 individuals) were gainfully employed in 1990.

[2] A short introduction into the public use version of the GSOEP is presented in Wagner et al. (1993) while Schupp and Wagner (1990) concentrate on the implementation of the survey in the then German Democratic Republic in 1990. A detailed description of the GSOEP public-use file is DIW (1998). It is available at http://www.diw.de.

Table 4.1: Sample definitions

			observations
Full sample			2696
Sample restrictions	not employed 1990	-504	2192
	in vocational training or in self-employment 1990	-128	2064
	age below 20 or above 55 in 1990	-228	1836
	missing information on the instruments	-48	1788
	missing information on exogenous variables	-121	1667
	missing information on outcome variable	-383	
Restricted sample			1284

SOURCE: GSOEP public-use file, waves 7-14, subsample C

In the empirical analysis those individuals not employed in 1990 and those in either vocational training or in self-employment are excluded from the sample. Further sample restrictions due to missing values on crucial variables (outcome variable, exogenous or instrumental variables) and due to the concentration on the age group 20-55 in 1990 leave a sample of 1284 individuals (denoted "restricted sample" in the sequel) (cf. table 4.1).[3]

If not mentioned otherwise, all sample statistics and estimation results in this paper are based on this restricted sample. For the purpose of comparison, a comparable restricted sample has been constructed for West Germany from subsamples A and B in the GSOEP. The full West German balanced sample consists of 6187 individuals (of which 63% employed at the time of the interview in 1990), its restricted version of 2160 individuals.

4.3 Definitions

This section defines the *treatment* and *outcome* variables used in the subsequent analysis.

[3]This study concentrates on the analysis of returns to job-to-job mobility and does not discuss issues related to individual retirement decisions. For this reason, individuals aged over 55 in 1990 are excluded from the sample.

4.3.1 Definition of the *treatment variable*: early job mobility

The job mobility indicators have been constructed on the basis of a set of yearly repeated survey questions regarding the individual's job situation.

First, all individuals are asked whether they have ended an employment relation since the beginning of the previous calendar year and, if so, for what reason. According to the reason stated, *job endings / job separations* can be classified into four categories: quits, displacement (firm closure or individual layoff), (early) retirement, and other job endings.[4]

Second, individuals who are employed at the time of the survey are asked whether they have had any change in their professional situation since the beginning of the previous calendar year. In the affirmative case, further information on the type of this change is provided which enables the classification of *job changes* into the following three categories: job-to-job shifts, within-firm job changes, and re-entry into employment after some period of non-employment. Job-to-job shifts are moves between two different employers (including moves into or out of self-employment) whereas individuals re-taking employment after some break (e.g. periods of further training or maternity leave) report a re-entry into employment. Note, however, that intervening spells in unemployment might also occur in the case of job-to-job shifts since the classification of the type of job change is left to the respondent.

Third, in the case of both job endings and job changes, additional information on the timing of the event is provided. All this information is used to define the job mobility indicators presented in table 4.2 below.[5]

Given the responses to the above questions, the *treatment* group is defined as the group of early job-to-job movers, i.e. those individuals who, by the end of 1991, had indicated at least one job-to-job shift. This group covers 20.2% of the East German sample as opposed to 9% in the West German sample. All remaining individuals are part of the *control* group.[6] Table 4.3

[4] Other job endings include e.g. regular time-limited contract endings or within-firm job changes.

[5] Given that the survey response refers to both the calendar year previous to the survey year as well as to the time up to the survey in the survey year itself, further controlling for the timing of the respective event is necessary in order both to give a correct picture of the incidence of job mobility during transition and to avoid counting some mobility episodes twice.

[6] In the following analysis, results will also be reported when defining the *treatment group* as job mobility in 1990 ("early (1990)") or job mobility between 1990 and 1992 ("early (1990-92)"). When defining as early job-to-job movers all those individuals who experienced at least one job-to-job shift in 1990, the *treatment group* covers 8.3 % of the East German sample. When defining as early job-to-job movers all those individuals who,

Table 4.2: Definitions of job mobility status: Overview

	definition	obs	%
job stayers	**neither job ending nor job change in 1990-96**	**248**	**19.31**
job movers	**any type of job change 1990-96**	**860**	**66.98**
within-firm	within-firm job change 1990-96	307	23.91
job-to-job	change of employer 1990-96	541	42.13
early (1990)	job-to-job shift 1990	106	8.26
early (1990-91)	job-to-job shift 1990-91	259	20.17
early (1990-92)	job-to-job shift 1990-92	336	26.17
re-entries	re-entry into employment 1990-96	281	21.88
job losers	**job ending without job change in 1990-96**	**176**	**13.71**

NOTE: last column reports percentage of restricted sample (N=1284)
SOURCE: GSOEP, waves 7-14, subsample C

below contains an overview of the two subgroups of treatment group and control group.

4.3.2 Definition of the *outcome variable*: total income 1990-96

Following the suggestions from theoretical models of individual decision-making under uncertainty, the *outcome* variable of interest is defined as the present discounted sum of total yearly incomes for the years 1990 to 1996.[7]

Total yearly income is defined as sum of various potential types of income received throughout the calendar year, namely labour income (employment or self-employment income including premia), income from a second job, unemployment benefit or relief, social assistance payments, payments during maternity leave, and grants.[8] Each of the single income components is calcu-

by the end of 1992, had indicated at least one job-to-job shift, the *treatment group* covers 26.2 % of the East German sample as opposed to 12.5 % in the West German sample.

[7] For arguments in favour of chosing an aggregate measure of income instead of year-to-year increases in income, cf. also the discussion of adequate outcome measures in studies of the returns to job mobility in Simmonet (1998).

[8] Given that e.g. job search off-the-job, unemployment, or continued education might represent the optimal behaviour for some individuals, all the above-mentioned potential income sources necessarily have to be included into the outcome variable.

Table 4.3: Definition of treatment and control groups: Overview

	description	obs	%
treatment group ($D = 1$)	**early (1990-91) job-to-job movers**	259	20.17
control group ($D = 0$)	**all remaining:**	1025	79.83
	job stayers	248	19.31
	other job movers	601	46.81
	late job-to-job movers	282	21.96
	only within-firm	175	13.63
	only re-entry	123	9.58
	within-firm and re-entry	21	1.64
	job losers	176	13.71

NOTE: last column reports percentage of restricted sample (N=1284)
SOURCE: GSOEP, waves 7-14, subsample C

lated as product of the number of months the income was received in a calendar year and the average monthly amount for the income source.[9]

As in Hunt (1999), the calculated nominal values are then deflated to 1991 values, using two different region-specific deflators for West and East Germany, respectively, according to the individual's residence.

The *outcome* variable of interest is obtained by adding up these discounted yearly income measures. It is defined as missing only if information on the amount of yearly labour income is missing. In the case other components of the total yearly income measure exhibit missing values these components are set to zero instead.[10] Table 4.4 contains summary statistics for these income variables.

In the empirical analysis, the log of the present discounted sum of total yearly income will

[9] This income information is extracted from the individual-specific calendar files in the GSOEP.

[10] This treatment of missing information is motivated both by the observation that labour income is by far the most important component of total yearly income and by the wish to reduce the number of missing values of the *outcome* variable to a reasonable minimum. It has to be noted, however, that in the many cases of missing information on the average amount of unemployment benefits it might be more accurate to replace the missing value by some fraction of previous labour income than by zero. Alternatively, one could use total yearly labour income only as *outcome* variable. The robustness of the results obtained with respect to alternative outcome variables will be discussed more in detail in section 4.6.

Table 4.4: Variable definitions and descriptive statistics: Total income throughout transition

variable	description	mean	st. dev.	min	max
TYI90	total income 1990	12.864,80	5.925,79	0	65.100
TYI91	total income 1991	21.192,15	10.441,71	0	99.519
TYI92	total income 1992	22.913,48	11.423,00	0	108.134
TYI93	total income 1993	24.771,13	12.877,90	0	125.859
TYI94	total income 1994	25.623,60	14.593,66	0	140.089
TYI95	total income 1995	25.674,38	16.367,92	0	215.600
TYI96	total income 1996	25.487,78	15.623,07	0	118.863
PDVTYI06	discounted sum 1990-96	158.527,30	75.924,73	16.716	757.495

NOTE: All income values gross in 1991 DM
SOURCE: GSOEP, waves 7-14, subsample C (N=1284)

be used as dependent variable.

4.4 Income effects of early job-to-job mobility: Some first evidence

This section presents some first empirical evidence on the returns to early job-to-job mobility on the basis of simple *treatment-control* comparisons and ordinary least squares estimates.

4.4.1 Treatment-control comparisons

Given the above definition of early job-to-job mobility as *treatment* (D=1), this section compares the income of early job-to-job movers with that from the *control* group (D=0). As shown in table 4.5 below, the mean income for the *treatment* group is considerably higher than that for the control group, suggesting positive returns to early job mobility of at least 16%. In the case of early (1990-91) job-to-job mobility, the mean income for the *treatment* group amounts e.g. to 178.493,80 DM (with a standard deviation of 90.234,73 DM) as compared to 153.482,10 DM for the *control* group (with a standard deviation of 71.026,77 DM), thus yielding a difference in mean income between early job-to-job movers and the *control* group of roughly 25.000 DM or 3.600 DM per year. On the basis of the other definitions of early job mobility, the returns

Table 4.5: Outcome variable by treatment status: Total income

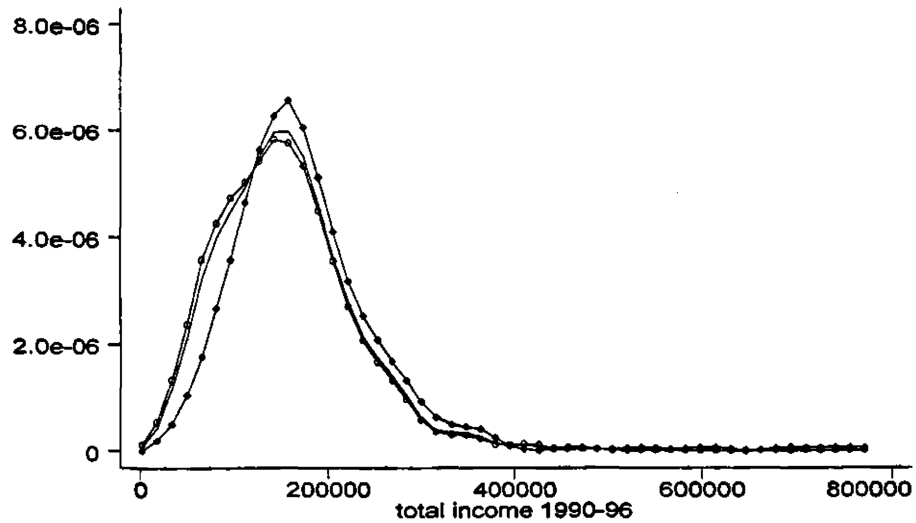
treatment group	E[Y|D=1]	E[Y|D=0]	t-test	KS-test
early job-to-job movers (1990)	209.809,20 (112.313,10)	153.912,80 (70.001,80)	-5.037 (0.000)	0.275 (0.000)
early job-to-job movers (1990-91)	178.493,80 (90.234,73)	153.482,10 (71.026,77)	-4.148 (0.000)	0.144 (0.000)
early job-to-job movers (1990-92)	177.845,10 (86.659,78)	151.680,50 (70.525,89)	-4.981 (0.000)	0.166 (0.000)

NOTES: standard deviations in parentheses; t-test refers to the one-sided two-sample t-test of equal means allowing for unequal variances with alternative $H_1 : E[Y|D=1] > E[Y|D=0]$;KS-test refers to the two-sample Kolmogorov-Smirnov test statistic of equal outcome distributions for treatment and control group with alternative $H_1 : F_{Y|D=1} < F_{Y|D=0}$;for both tests p-values in parentheses; SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284)

to early job mobility seem even higher.

Standard tests on the equality of means and distributions across treatment groups respectively clearly reject the respective null of equal means and equal distributions: A one-sided two-sample t-test of equal means (allowing for unequal variances) yields a test statistic of -4.98, thus clearly rejecting the null of equal means at any standard significance level. A two-sample Kolmogorov-Smirnov-test of equal outcome distributions for treatment and control group gives a test statistic of 0.17 which, again, allows a clear rejection of the null of equal distributions. Similar tests on the yearly mean incomes and income distributions, respectively, also indicate significant differences between treatment and control groups. This result is also supported by a closer look at the distribution of the *outcome* variable by *treatment* status in Figure 4-1. The comparison of the kernel density estimates of the present discounted sum of total yearly income between 1990 and 1996 clearly suggests a considerable income gap between the *treatment* and *control* groups.

Figure 4-1: Distribution of total income 1990-96 by treatment status



◇ treatment status; ○ control group

4.4.2 Selection on covariates

The above simple *treatment-control* comparisons are likely to suffer from ignoring systematic differences in observable characteristics between the *treated* and the *control* group. A natural extension therefore consists in estimating potential returns to job mobility conditional on observable individual and job characteristics.

OLS regression

For this purpose, in a first step the outcome equation (3.9)

$$Y_i = X_i'\beta + D_i\delta + u_i \quad (i = 1, ..., N)$$

has been estimated on the cross-section defined above by simple OLS. Y_i denotes the *outcome* variable, X_i a matrix of exogenous regressors, including individual characteristics, firm characteristics, monthly income 1989, and potentially a set of occupation or industry dummies, D_i a binary indicator of the job mobility status under consideration, and u_i the error term. The parameter of interest is (β, δ), with δ considered as the returns to job mobility.

Descriptive statistics on the observable characteristics in the sample and by treatment status are presented in tables 4.29 and 4.30 in the appendix. As can be seen from the descriptive statistics by treatment status and as one would expect in the case of self-selection / non-randomisation, the means respectively distributions of some "pre-treatment" observable characteristics do indeed differ significantly with treatment status: Men, the young, those with lower tenure on the job, the more educated (10 or 12 years of schooling, university degree), employees in the non-state sector, workers in agriculture and employees in trade and retail are more likely to be treated. Women, employees in the transport or public sector, and those in a low position are less likely to be treated. It is important to note, however, that neither the mean nor the distribution of gross monthly earnings in the year previous to the interview do differ by treatment status. The "pre-treatment" labour market or employment status are further equal by definition of the sample.

Table 4.6 below contains the results from the regression of the log of the present discounted sum of total yearly income over the transition period on different sets of exogenous variables. Already the results from the first specification which includes individual characteristics only indicate the explanatory power of these characteristics for the *outcome* variable under consideration: Additionally to gender and age effects, the outcome variable also differs considerably with educational and vocational attainment as of 1990.[11] The inclusion of further regressors like firm characteristics and lagged income in 1989 as well as either a public service dummy (specification 2), occupation dummies (specification 3) or industry dummies (specification 4) does not change these results qualitatively. The estimation results suggest, finally, considerable returns to early job-to-job mobility of 8-12%.

While the OLS estimates of the returns to early (1990-92) job mobility are similar, the OLS estimates of the returns to early (1990) job mobility seem considerably higher of around 20%

[11]Most noteworthy seems the considerably higher income stream for men with technical or university degree. Given that it is generally understood that these qualifications were undervalued in a centrally planned economy with respect to vocational training, one might want to classify graduates of either technical schools or universities as among the 'winners' during transition. On the other hand, elder individuals, women, and individuals with low schooling and / or without any vocational degrees seem to be among the 'losers'.

Table 4.6: OLS estimates of the outcome equation

	(1)	(2)	(3)	(4)
constant	13.885	12.566	12.651	12.553
	(0.623)	(0.684)	(0.675)	(0.672)
D	0.078	0.101	0.090	0.113
	(0.026)	(0.026)	(0.026)	(0.026)
sex (female=1)	-0.338	-0.265	-0.290	-0.254
	(0.023)	(0.026)	(0.027)	(0.028)
age	-0.200	-0.245	-0.228	-0.241
	(0.053)	(0.054)	(0.054)	(0.053)
$age^2/100$	0.647	0.739	0.691	0.725
	(0.145)	(0.146)	(0.146)	(0.144)
$age^3/10.000$	-0.652	-0.714	-0.675	-0.701
	(0.128)	(0.128)	(0.128)	(0.126)
8 years schooling	-0.135	-0.112	-0.078	-0.106
	(0.033)	(0.031)	(0.031)	(0.031)
12 years schooling	0.032	0.025	-0.004	0.009
	(0.051)	(0.049)	(0.048)	(0.048)
no vocational degree	-0.302	-0.209	-0.084	-0.205
	(0.061)	(0.054)	(0.058)	(0.052)
masters degree	0.064	0.036	0.023	0.038
	(0.043)	(0.040)	(0.044)	(0.039)
technical degree	0.324	0.241	0.152	0.245
	(0.033)	(0.033)	(0.038)	(0.034)
university degree	0.415	0.293	0.201	0.308
	(0.064)	(0.062)	(0.067)	(0.063)
small firm (<20)	-	-0.101	-0.116	-0.117
	(-)	(0.043)	(0.043)	(0.044)
large firm (>200)	-	0.029	0.004	0.023
	(-)	(0.023)	(0.024)	(0.024)
monthly income 1989	-	0.282	0.250	0.272
	(-)	(0.037)	(0.037)	(0.037)
public service dummy	no	yes	no	no
occupation dummies	no	no	yes	no
industry dummies	no	no	no	yes
N	1284	1284	1284	1284
R^2	0.382	0.441	0.452	0.454

NOTES: dependent variable: log of the discounted sum of total income 1990-96; D: indicator of early job-to-job mobility (1990-91) (treatment status); robust standard error estimates in parentheses;
SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284)

(Cf. tables 4.39 and 4.42 in the appendix).

Mincer (1986) approach

As mentioned in Mincer (1986), heterogeneity in workers' wage profiles creates a selectivity bias in using all stayers as the control group.[12] Additionally to the selection on covariates, the author therefore suggests to restrict the control group implicit in the regression to "those stayers whose mobility behavior, in addition to other observable characteristics, is similar to that of current period movers", considering their wage growth on the job a proxy of the wage growth foregone by the job movers. Mincer (1986) proposes to proxy the expected wage growth of movers by the wage growth of those stayers who, besides having similar personal and job characteristics, in particular tenure on the job, are observed to change job one year later. Table 4.7 presents the estimation results based on the restriction of the control group to these "comparable stayers" or "next period movers". These suggest again gains from early job mobility in the order of 15% when compared to late (1992-96) movers. Late job movers are found to incur losses with respect to non-movers in general and job stayers in particular, while early movers experience clear income gains with respect to both of these groups. Gains seem to be highest for those moving still in 1990, reaching income gains of around 30% with respect to those moving later. Movers in the first two years of transition, however, do not seem to experience any significant gains with respect to those moving in 1992 (Cf. also tables 4.37 and 4.40 in the appendix).

4.4.3 Interpretation of results

The results obtained from simple *treatment-control* comparisons and from standard regression analysis suggest considerable positive income effects of early job-to-job mobility. They can in

[12] "Since wages change (usually grow) for stayers as well as for movers, a correct estimate of the wage gain from moving is the difference between the actual wage gain of movers over the interval and the unobserved but expected wage gain of movers had they not moved over the same interval. In the usual procedure, the coefficient of a job change dummy (S_t) in a wage growth equation is used to estimate the wage transition. What it really measures is the difference between the wage gain of movers and wage gain of stayers both defined over this particular time interval (t). (...) What is known as the "selectivity problem" is that wage growth of stayers (...) is not likely to be the same as the expected wage growth of movers (...) had they stayed. Put more strongly, the coefficient on S_t tells us how much better (or worse) movers fared compared to stayers, but this is an irrelevant and even faulty question. It is prima facie faulty, because any answer would suggest that one or the other group is acting irrationally (...). It is irrelevant because economic optimization means that movers are doing their best by moving and stayers by staying. Strictly speaking this is true ex ante, as well as, on average, ex post, so long as most people are not misled by incomplete information."

Table 4.7: OLS estimates of the returns to early (1990-91) job mobility following Mincer (1986): Overview

	(1)	(2)	(3)	(4)	estimated returns
control: movers in 1992					
$G(m) = b_t - b_{t+1}$	-0.014	0.010	0.004	0.029	0%
coefficient on D	0.086	0.109	0.098	0.121	
(s.e.)	(0.027)	(0.026)	(0.026)	(0.027)	
coefficient on $\tilde{D}$	0.100	0.099	0.094	0.093	
(s.e.)	(0.038)	(0.038)	(0.037)	(0.037)	
R^2	0.385	0.443	0.454	0.456	
control: late movers in 1992-96					
$G(m) = b_t - b_{t+1}$	0.130	0.142	0.139	0.156	13.9 % - 16.9 %
coefficient on D	0.056	0.083	0.069	0.093	
(s.e.)	(0.028)	(0.028)	(0.027)	(0.028)	
coefficient on $\tilde{D}$	-0.074	-0.059	-0.070	-0.063	
(s.e.)	(0.026)	(0.026)	(0.024)	(0.025)	
R^2	0.386	0.443	0.455	0.456	

NOTES: dependent variable: log of the discounted sum of labour income 1990-96; D: early job-to-job mobility (1990-91); model specifications (1) to (4) as defined above; robust standard error estimates in parentheses; SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284) general not, however, be considered as estimates of *causal effects* due to the inherent problem of self-selection.

The same negative result holds true for OLS estimates: If, as discussed in chapter 2, job mobility status is correlated with unobservable characteristics like ability or the access to networks, these estimates are biased. As suggested in the above theoretical discussion, the direction of this bias is generally unknown. In other words: The conditional independence (or *unconfoundedness*) assumption implicit in the selection on observables is not valid.

Moreover, even in the case these conditional independence assumptions were to be met, the above results would represent the treatment effect on the treated when assuming treatment effect homogeneity in the treated population. The next section therefore turns to a different approach to identification and estimation of returns to job mobility which addresses explicitly the problem of self-selection while allowing for treatment effect heterogeneity across different subpopulations.

4.5 Instrumental variables estimation of the returns to job mobility during transition

As discussed in chapter 3, a well-known solution to the problem of endogenous *treatment* status D_i in the above outcome equation (3.9) consists in the use of instrumental variables as a source of identifying information. The instrumental variables are interpreted as generating a quasi-*natural experiment* by *assigning* individuals to *treatment* independently of their unobserved characteristics. Under the conditions AIR(1996) laid out in theorem 12 in chapter 3, the IV estimate of the coefficient on the *treatment* indicator, D_i, is not only a consistent estimate of δ, but it has, moreover, a *causal* interpretation as *local average treatment effect* (*LATE*).

4.5.1 Choice of instruments and discussion of identifying assumptions

Two instruments have been chosen for the instrumental variables estimation: first, an indicator of firm-level announced layoffs, and second, an indicator of occupation-specific labour market opportunities. Both instruments stem from the 1990 survey and thus convey pre-EMSU information.[13] They are motivated in turn.

Firm-level announced layoffs as instrument

Individuals were asked in the first wave whether their current employer had already announced layoffs on the firm level. In the restricted sample defined above, 601 individuals answer affirmatively and 683 negatively. I define the first instrument as indicator taking the value 1 if the response is affirmative.

This instrument is correlated with the treatment status of early job-to-job mobility while only slightly (and negatively) correlated with the outcome (Corr(Z,D)=0.14 and Corr(Z,Y)=-0.03). The treatment probabilities, moreover, differ significantly by assignment status: Of those assigned to treatment 26% switch to a new job early, as opposed to only 15% of those not assigned to treatment. A one-sided t-test of equal treatment probabilities for different assignment status yields a statistic of -4.67, thus clearly rejecting the null.

The instrument seems to be reasonably equally distributed across the sample population. Men, employees of large firms, blue collar workers, and workers in the manufacturing sector are over-represented in the group of individuals assigned to treatment due to their employers' announcement of mass layoffs at the beginning of transition. On the other hand, especially employees in the transport sector and the public sector faced a significantly lower probability of early mass layoffs. There is no statistical evidence for differences in firm-specific layoff probabilities with respect to either educational and vocational degrees or hierarchical positions on the current job. Further evidence on the distribution of the instrument in the sample population is presented in table 4.31 in the appendix.

Average outcomes by instrument and treatment status are summarised in table 4.8 below. They show that, while there is a positive effect of firm-level announced layoffs on early job mobility, firm-level announced layoffs have a negative causal effect on total income 1990-96. The simple (unconditional) Wald estimate of the causal effect of early job mobility on total income

[13]It is crucial to understand that it is only the framework of transition which allows to employ this information as an exogenous source of variation in observed job mobility behaviour. Even if it were available, the same information on firm-level announced layoffs in West German firms or on occupation-specific labour market opportunities of West German survey respondents would certainly not constitute valid instruments in an analysis of job mobility in West Germany!

Table 4.8: Average outcome by instrument and treatment status (instr. Z1)

	D_{obs}=0	D_{obs}=1	Total	$\overline{D}$
$Z_{1,obs} = 0$	158.881,20 (72.573,99) (N=580)	171.865,50 (90.285,35) (N=103)	160.839,30 (75.579,65) (N=683)	0.151 (0.358) (N=683)
$Z_{1,obs} = 1$	146.445,10 (68.401,16) (N=445)	182.870,10 (90.223,91) (N=156)	155.899,90 (76.292,84) (N=601)	0.260 (0.439) (N=601)
Total	153.482,10 (71.026,77) (N=1025)	178.493,80 (90.234.,73) (N=259)	158.527,30 (75.924,73) (N=1284)	0.202 (0.401) (N=1284)

NOTES: Z_1 = firm-level announced layoffs as of 1990; standard deviation and number of observations in parentheses; SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284)

1990-96 based on firm-level announced layoffs as instrument is -0.28, suggesting considerable negative returns to early job mobility of around -25%.

As to the plausibility of the conditions AIR(1996) stated in theorem 12 in chapter 3, one might, first, want to accept SUTVA as the usual assumption of the absence of general equilibrium effects in analyses of this kind on the microlevel. Second, it is important to note that the instrument conveys information about potential job loss in a given firm and/or sector. It does, especially, not contain information about the *individual-specific* probability of job loss due to unobservable characteristics unless individuals have self-selected into "good" and "bad" firms in the centrally planned economy.[14] Since such a self-selection seems unlikely, there is good reason to believe that, conditional on the covariates retained, assignment to treatment is strongly ignorable. Third, it seems unlikely that individuals who decided not to change job in the case their firm had announced future layoffs would have done so if their firm had not announced any layoffs. Together with the evidence of non-zero average treatment effects of the chosen instrument, this underlines the plausibility of the assumption of strong monotonicity in

[14] "Good" and "bad" here refer to the firms' reactions (in terms of labour demand and labour restructuring) to the negative shocks following unification as reflected by the immediate announcement of future firm-level layoffs. Cf. also the discussion in chapter 2, section 2.4.

this case. Finally, since, as stated above, the announcement of firm-level layoffs is unlikely to be correlated with unobserved income components, and since it is not clear whether individuals in these firms have a significantly lower employment probability throughout the transition process than individuals in firms that did not announce layoffs immediately, the exclusion restrictions might hold.[15]

As discussed in detail in chapter 2, an increase in the layoff probability due to a firm-specific negative unification shock is likely to induce mainly individuals with either low ability or weak networks (*LW-individuals*) to switch early to a new job.

Occupation-specific labour market opportunities as instrument

Individuals were further asked in the first wave to evaluate the possibility of finding an *equivalent* job if they were to lose their current job. In the restricted sample defined above, 213 individuals report that it would be easy while the remaining 1071 consider it as difficult or impossible. I define the second instrument as indicator taking the value 1 if an individual considers finding an equivalent job easy.

Like the first instrument, also this second instrument is correlated with the treatment status of (early) job-to-job mobility and to a much lesser degree (and positively) with the outcome (Corr(Z,D)=0.12 and Corr(Z,Y)=0.05). The treatment probabilities also differ significantly by assignment status. Of those assigned to treatment by the instrument, 31% do actually switch to a new job early, as opposed to only 18% in the case of no assignment to treatment. The value of the t-statistic of a test for equal treatment probabilities for different assignment status is -3.83 thus clearly rejecting the null hypothesis of equal treatment probabilities.

As can be seen from table 4.32 in the appendix, the instrument further seems to be reasonably equally distributed across the sample population, although to a much lesser degree than the first instrument. Most noteworthy are probably the differences between blue and white collar workers in evaluating their occupation-specific labour market chances: The former seem to face

[15] Indeed, only half (54%) of those individuals who are in firms that had announced mass layoffs in early 1990 also declared that their individual layoff was likely, contrary to the other half (46%) who declared a small likelihood of individual layoff. Independently of the treatment status, layoff rates later in the transition process are further found to be similar for those in firms that had announced layoffs and those whose firms had not done so. Finally, the results reported later in this chapter are found to be robust against excluding those individuals laid off from the sample.

Table 4.9: Average outcome by instrument and treatment status (instr. Z2)

	D_{obs}=0	D_{obs}=1	Total	$\overline{D}$
$Z_{2,obs} = 0$	152.902,60	174.661,80	156.823,70	0.180
	(71.878,27)	(87.018,84)	(75.255,31)	(0.385)
	(N=878)	(N=193)	(N=1071)	(N=1071)
$Z_{2,obs} = 1$	156.943,30	189.699,50	167.093,10	0.310
	(65.835,75)	(98.895,14)	(78.829,99)	(0.464)
	(N=147)	(N=66)	(N=213)	(N=213)
Total	153.482,10	178.493,80	158.527,30	0.202
	(71.026,77)	(90.234.,73)	(75.924,73)	(0.401)
	(N=1025)	(N=259)	(N=1284)	(N=1284)

NOTES: Z_2 = occupation-specific labour market opportunities as of 1990; standard deviation and number of observations in parentheses; SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284)

significantly more favourable employment prospects at the beginning of transition. Moreover, with respect to educational attainment and hierarchical employment position in 1990, both the low-level and the high-level groups report significantly less favourable occupation-specific labour market prospects. There are, further, significant age- and tenure-differences in evaluating the chances of finding an equivalent job. Among the sectors the construction sector offers the most favourable outside options while the manufacturing sector offers the least favourable.

Average outcomes by instrument and treatment status are summarised in table 4.9. They show that there is a positive effect of occupation-specific labour market opportunities on both early job mobility and total income 1990-96. The simple (unconditional) Wald estimate of the causal effect of early job mobility on total income 1990-96 based on occupation-specific labour market opportunities as instrument is 0.50, suggesting considerable positive returns to early job mobility of around 65%.

As in the case of the first instrument, SUTVA, strongly ignorable treatment assignment, and strong monotonicity seem to be plausible assumptions: It seems unlikely that individuals self-selected into "good" and "bad" occupations in the centrally planned economy according to their unobservable characteristics.[16] Hence, conditional on the covariates retained, assignment to treatment on the basis of the second instrument might be considered as strongly ignorable. Strong monotonicity further seems a plausible assumption because, first, there is a clear non-zero average treatment effect of occupation-specific labour market opportunities as assignment and, second, it seems again unlikely that some individual would only change jobs in the case of rare outside options, but not do so if it was easy to find an equivalent job.

The validity of the exclusion restriction, however, is probably more difficult to justify in this second case. As before, it is important to note that the instrument conveys information about the availability of outside options for a given occupation and/or hierarchical position, more precisely about the availability of an option *equivalent* to the current job. It should not, again, contain information about the *individual-specific* chances to simply find *some* job due to unobservable characteristics like e.g. motivation, and should therefore be uncorrelated with unobserved income components. Moreover, it was certainly not obvious at the beginning of the transition process whether the same occupations would remain favoured throughout the whole process. These observations provide some evidence for the validity of the exclusion restrictions also in the case of the second instrument.

Finally, according to the model in chapter 2, one might expect the group of individuals complying with the assignment that is implicit in this second instrument to consist mainly of individuals with either high ability or strong networks (*HS-individuals*).

Extensions: Interaction of instruments and multiple instruments

As discussed above, the instruments proposed are likely to refer to different groups of compliers, either those who only change job early because of firm-specific risk of job loss or those who only change job early because of occupation-specific outside opportunities on the labour market. By using different instruments, different assignment mechanisms and groups of compliers are captured. This section suggests as an extension to the above choice of instruments to look also at the interactions of the above instruments (namely the minimum and the maximum of the two instruments above).

[16] Here the adjectives "good" and "bad" refer to increases and decreases in occupation-specific labour demand at the beginning of transition. Cf. also the discussion in chapter 2, section 2.4.

Table 4.10: Average outcome by instrument and treatment status (instr. Z3)

	D_{obs}=0	D_{obs}=1	Total	$\overline{D}$
$Z_{3,obs} = 0$	153.967,60	172.773,20	157.518,20	0.189
	(71.636,40)	(83.028,66)	(74.250,72)	(0.392)
	(N=971)	(N=226)	(N=1197)	(N=1197)
$Z_{3,obs} = 1$	144.751,70	217.671,40	172.410,90	0.379
	(58.875,01)	(123942,10)	(95.491,22)	(0.488)
	(N=54)	(N=33)	(N=87)	(N=87)
Total	153.482,10	178.493,80	158.527,30	0.202
	(71.026,77)	(90.234.,73)	(75.924,73)	(0.401)
	(N=1025)	(N=259)	(N=1284)	(N=1284)

NOTES: Z_3 =min(Z_1,Z_2) where Z_1 = firm-level announced layoffs as of 1990 and Z_2 = occupation-specific labour market opportunities as of 1990; standard deviation and number of observations in parentheses; SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284)

The above arguments in favour of the choice of instruments remain valid when combining these, and both interaction instruments show a clear non-zero average treatment effect. Average outcomes by instrument and treatment status in the case of interaction instruments are summarised tables 4.10 and 4.11. When compared to the average outcomes by instrument and treatment status in the case of the first two instruments, it is clear that the interaction of these captures some intermediate effects: Simple (unconditional) Wald estimates on the effect of early job mobility on total income 1990-96 based on these interaction instruments suggest strongly positive returns of around 65% for those who only change job early because of both firm-specific risk of job loss and occupation-specific outside options compared to negative returns of around -13% for those who only change job early because of either firm-specific risk of job loss or occupation-specific outside options.

While instrumental variables estimation on the basis of both instruments simultaneously further has the advantage of allowing for tests of overidentifying restrictions, it is not clear how to interpret the assignment mechanism implied by such a multiple instrument, whether the results can still be interpreted as causal effects and to which subgroup of compliers they would apply. These reservations notwithstanding, results from instrumental variables estimation using

Table 4.11: Average outcome by instrument and treatment status (instr. Z4)

	D_{obs}=0	D_{obs}=1	Total	$\overline{D}$
$Z_{4,obs} = 0$	157.899,40	176.644,90	160.255,20	0.126
	(73.286,19)	(103.135,90)	(77.803,72)	(0.332)
	(N=487)	(N=70)	(N=557)	(N=557)
$Z_{4,obs} = 1$	149.483,60	179.178,60	157.203,40	0.260
	(68.741,45)	(85.254,13)	(74.479,82)	(0.439)
	(N=538)	(N=189)	(N=727)	(N=727)
Total	153.482,10	178.493,80	158.527,30	0.202
	(71.026,77)	(90.234.,73)	(75.924,73)	(0.401)
	(N=1025)	(N=259)	(N=1284)	(N=1284)

NOTES: Z_4 =max(Z_1,Z_2) where Z_1 = firm-level announced layoffs as of 1990 and Z_2 = occupation-specific labour market opportunities as of 1990; standard deviation and number of observations in parentheses; SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284)

both instruments simultaneously will be presented in the sequel for comparison.

4.5.2 Estimation results

The (conditional) instrumental variables estimates of the returns to early job mobility on the basis of the chosen instrument have been computed using the two-stage least squares procedure: in the first stage, the treatment indicator D is regressed on the whole list of exogenous variables augmented by the instrumental variable $Z_i, i = 1,2,3,4$ using a simple linear probability model; in the second stage, the predicted value of the dependent variable from the first stage regression, $\widehat{D_i}$, is then used as additional regressor in the outcome equation instead of the treatment status D itself. For one selected specification, the following tables 4.12 and 4.13 contain the detailed results of the two separate regressions for each of the four instruments discussed above.

There is a clear significant and positive effect of all instruments on the treatment status of early job mobility. Apart from gender, none of the other exogenous variables seems relevant for differences in early job mobility behaviour, however.

The estimated returns to early job mobility differ substantially according to the instrument

Table 4.12: IV first stage estimates for a selected specification

instrument	$i=1$ Z_1	$i=2$ Z_2	$i=3$ $\min(Z_1,Z_2)$	$i=4$ $\max(Z_1,Z_2)$
constant	0.940	0.888	0.893	0.920
	(0.720)	(0.719)	(0.711)	(0.726)
Z_i	0.108	0.128	0.176	0.135
	(0.023)	(0.034)	(0.054)	(0.022)
sex (female=1)	-0.091	-0.090	-0.093	-0.084
	(0.027)	(0.027)	(0.027)	(0.027)
age	-0.036	-0.031	-0.029	-0.037
	(0.058)	(0.058)	(0.057)	(0.058)
$age^2/100$	0.115	0.104	0.100	0.119
	(0.155)	(0.156)	(0.155)	(0.156)
$age^3/10.000$	-0.118	-0.107	-0.105	-0.119
	(0.134)	(0.135)	(0.134)	(0.135)
8 years schooling	-0.039	-0.035	-0.034	-0.039
	(0.033)	(0.033)	(0.033)	(0.033)
12 years schooling	0.062	0.077	-0.067	0.070
	(0.049)	(0.050)	(0.050)	(0.049)
no vocational degree	0.023	-0.009	0.000	0.023
	(0.065)	(0.066)	(0.065)	(0.064)
masters degree	-0.042	-0.034	-0.031	-0.041
	(0.046)	(0.046)	(0.046)	(0.046)
technical degree	-0.032	-0.027	-0.023	-0.032
	(0.038)	(0.038)	(0.038)	(0.038)
university degree	-0.015	-0.005	-0.003	-0.015
	(0.066)	(0.066)	(0.066)	(0.066)
small firm (<20)	0.004	-0.032	-0.018	-0.015
	(0.042)	(0.043)	(0.042)	(0.042)
large firm (>200)	-0.019	0.008	-0.008	-0.022
	(0.026)	(0.027)	(0.026)	(0.026)
monthly income 1989	-0.059	-0.005	-0.059	-0.058
	(0.033)	(0.026)	(0.033)	(0.033)

NOTES: dependent variable: indicator of early job-to-job mobility (1990-91) (treatment status D); Z_1 = firm-level announced layoffs as of 1990; Z_2 = occupation-specific labour market chances as of 1990; Z_3 = min(Z1,Z2); Z_4 = max(Z1,Z2); Z_5 = (Z1,Z2); results for model specification (3), including firm-size dummies, monthly income 1989 and occupation dummies among the regressors; robust standard error estimates in parentheses; SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284)

Table 4.13: IV second stage estimates for a selected specification

instrument	$i=1$ Z_1	$i=2$ Z_2	$i=3$ $\min(Z_1,Z_2)$	$i=4$ $\max(Z_1,Z_2)$
constant	13.617	12.326	12.671	13.248
	(1.106)	(0.720)	(0.714)	(0.884)
$\widehat{D}_i = IVE_i$	-0.937	0.435	0.069	-0.546
	(0.304)	(0.242)	(0.276)	(0.191)
sex (female=1)	-0.392	-0.256	-0.292	-0.353
	(0.050)	(0.038)	(0.038)	(0.038)
age	-0.262	-0.217	-0.229	-0.249
	(0.084)	(0.056)	(0.054)	(0.068)
$\text{age}^2/100$	0.804	0.654	0.694	0.761
	(0.226)	(0.153)	(0.147)	(0.183)
$\text{age}^3/10.000$	-0.792	-0.636	-0.678	-0.747
	(0.196)	(0.134)	(0.129)	(0.159)
8 years schooling	-0.117	-0.064	-0.078	-0.102
	(0.047)	(0.035)	(0.032)	(0.038)
12 years schooling	-0.068	-0.027	-0.002	0.041
	(0.075)	(0.052)	(0.052)	(0.060)
no vocational degree	-0.090	-0.082	-0.084	-0.088
	(0.093)	(0.061)	(0.058)	(0.074)
masters degree	-0.024	-0.039	0.022	-0.006
	(0.064)	(0.048)	(0.044)	(0.052)
technical degree	0.118	0.164	0.151	0.131
	(0.065)	(0.040)	(0.038)	(0.045)
university degree	0.185	0.206	0.200	0.191
	(0.096)	(0.071)	(0.068)	(0.080)
small firm (<20)	-0.136	-0.109	-0.116	-0.128
	(0.062)	(0.046)	(0.044)	(0.052)
large firm (>200)	0.000	-0.005	0.004	0.002
	(0.036)	(0.025)	(0.024)	(0.029)
monthly income 1989	0.193	0.269	0.249	0.215
	(0.056)	(0.041)	(0.041)	(0.045)

NOTES: dependent variable: log of the discounted sum of total income 1990-96; D: indicator of early job-to-job mobility (1990-91) (treatment status); Z_1 = firm-level announced layoffs as of 1990; Z_2 = occupation-specific labour market chances as of 1990; Z_3 = min(Z1,Z2); Z_4 = max(Z1,Z2); Z_5 = (Z1,Z2); results for model specification (3), including firm-size dummies, monthly income 1989 and occupation dummies among the regressors; robust standard error estimates in parentheses; SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284) used. They are summarised in the rows denoted IVE_1, IVE_2, IVE_3 and IVE_4 respectively, of table 4.14 below. For the purpose of comparison, the row denoted IVE_5 contains the IV estimation results when using both instruments simultaneously.[17] First-stage F-statistics and partial R^2 measures are reported as a diagnostic tool following the suggestions of Bound et al. (1995) and Staiger and Stock (1997). In both cases, the instrument quality seems reasonable as suggested by these measures.

The estimated returns to early job mobility contrast strongly with the OLS estimates discussed above. They give, moreover, qualitatively different results according to the instrument used: When using firm-level announced layoffs as an instrument, IV estimates of the returns to early job-to-job mobility are found to be significantly negative. When using occupation-specific labour market opportunities as instrument instead, the IV estimates of the returns to early job-to-job mobility are positive but only marginally significant. The estimation results for the alternative treatment variables are similar (Cf. tables 4.39 and 4.42 in the appendix).

4.5.3 Interpretation of results

The assumption that individuals do not decide on their job mobility behaviour during transition on the basis of some idiosyncratic gains from moving seems problematic. For this reason, the *average treatment effect on the treated* cannot be identified by instrumental variables methods, leading Heckman (1997) to refuse the interpretability of the IV estimate as a *causal effect*.[18] The obtained results from instrumental variables estimation should therefore be interpreted as estimates of *local average treatment effects* in the sense of Imbens and Angrist (1994): For those individuals who only change job early in transition due to a high firm/sector-specific risk of displacement, the returns to job mobility are found to be strongly negative. For those, on the contrary, who are induced to early job mobility only by the occupation-specific availability of

[17] In this latter case of two instruments, however, it is not clear how the estimates have to be interpreted in a quasi-experimental framework. As shown *inter alia* in Imbens and Angrist (1994), the instrumental variables estimate based on multiple instruments can be interpreted as a weighted average of local average treatment effects obtained from separate IV estimations for each single instrument. Nevertheless, the characterisation of the subgroups of interest - and in particular of the subgroup of compliers - remains problematic. In the case of the above instruments, the IV estimate IVE_5 represents the weighted average of the causal effects of *forced* job mobility, on the one hand, and *voluntary* job mobility, on the other. While informative about the "quality" of the two separate instruments Z_1 and Z_2, the magnitude of this weighted average is thus of limited interest.

[18] Cf. the discussion in chapter 3, section 3.3.

Table 4.14: Estimates of the returns to early (1990-91) job mobility: Overview

	(1)	(2)	(3)	(4)	estimated returns
OLSE	0.078	0.101	0.090	0.113	8.1 % - 12.0 %
	(0.026)	(0.026)	(0.026)	(0.026)	
R^2	0.382	0.441	0.452	0.454	
IVE_1	-0.886	-0.876	-0.937	-0.859	-60.8 % - -57.6 %
	(0.321)	(0.359)	(0.304)	(0.378)	
1st stage F	18.40	14.62	21.12	12.96	
partial R^2	0.0147	0.0121	0.0172	0.0107	
IVE_2	0.455	0.349	0.435	0.276	31.8 % - 57.6 %
	(0.270)	(0.233)	(0.242)	(0.239)	
1st stage F	12.28	14.48	14.18	13.56	
partial R^2	0.0119	0.0139	0.0136	0.0133	
IVE_3	0.148	0.057	0.069	0.107	5.9 % - 16.0 %
	(0.288)	(0.296)	(0.276)	(0.310)	
1st stage F	9.63	9.92	10.82	8.14	
partial R^2	0.0108	0.0110	0.0120	0.0094	
IVE_4	-0.525	-0.463	-0.546	-0.491	-42.1 % - - 37.1 %
	(0.210)	(0.212)	(0.191)	(0.223)	
1st stage F	32.16	27.49	36.44	25.58	
partial R^2	0.0230	0.0207	0.0267	0.0194	
IVE_5	-0.281	-0.222	-0.326	-0.232	-27.8 % - -19.9 %
	(0.184)	(0.181)	(0.169)	(0.190)	
1st stage F	17.55	15.98	20.04	14.41	
partial R^2					

NOTES: dependent variable: log of the discounted sum of total income 1990-96; treatment variable D: early job-to-job mobility (1990-91); Z_1 = firm-level announced layoffs as of 1990; Z_2 = occupation-specific labour market chances as of 1990; Z_3 = min(Z1,Z2); Z_4 = max(Z1,Z2); Z_5 = (Z1,Z2); model specifications (1) to (4) as defined above; robust standard error estimates in parentheses; SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284)

Table 4.15: Estimates of the fractions of the various subpopulations for the assignment mechanisms implied by the different instruments

instrument	AT lower bound	NT lower bound	C upper bound
Z_1	8.02 % (N=103)	34.66 % (N=445)	57.32 % (N=736)
Z_2	15.03 % (N=193)	11.45 % (N=147)	73.52 % (N=944)
Z_3	17.60 % (N=226)	4.21 % (N=54)	78.19 % (N=1004)
Z_4	5.45 % (N=70)	41.90 % (N=538)	52.65 % (N=676)

NOTES: treatment variable: early job-to-job mobility (1990-91); Z_1 = firm-level announced layoffs as of 1990; Z_2 = occupation-specific labour market chances as of 1990; Z_3 = min(Z1,Z2); Z_4 = max(Z1,Z2); SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284)

outside options, job mobility seems to pay off instead. Interpreting the relevant subpopulations of *compliers* in these two cases as forced movers and voluntary movers, respectively, allows to summarise the *causal effects* of early job-to-job mobility during transition as follows: Voluntary movers have gained from switching to a new job early in transition while forced movers have incurred substantial losses from doing so.

Although the respective subpopulations of compliers are unidentifiable by definition one can at least put lower bounds on the fractions of always-takers and never takers and consequently - since existence of defiers is ruled out by definition - an upper bound on the fraction of compliers in the sample. The lower bound for always-takers is given by the fraction of individuals with D=1 despite of Z=0, that for never-takers by the fraction of individuals with D=0 despite of Z=1. Table 4.15 shows these bounds and shows that the group of compliers is not unlikely to represent a considerable fraction of the population.

Another attempt in illustrating the above estimated returns to early job mobility for the unidentifiable group of compliers would be to make assumptions about the likely (observable) characteristics of the compliers and to re-run the above OLS regressions for this subgroup only. If one were able to identify the subpopulation of compliers, the OLS results should indeed be close to the obtained IVE results which - by definition - do measure the treatment effect for this subpopulation exclusively.

In a first approach, the above regressions were re-estimated on a sample which excluded the observed always-takers and never-takers. For the first instrument, the estimated coefficient on the treatment indicator decreased to 0.030 with a standard error of 0.034, suggesting that there were no returns to early job mobility. For the second instrument, the estimated returns increase to 0.165 (0.048). Both results, although no formal proof, are at least qualitatively in line with the results suggested by the LATE-IV estimation results.

A second approach to the identification of compliers is based on the theoretical results obtained in chapter 2 where it was shown that compliers in case of the first instrument are likely to be *LW-individuals* (individuals with low ability or weak networks) while compliers in the case of the second instrument tend to be *HS-individuals* (individuals with high ability or strong networks). It tries to identify compliers on the basis of a range of exogenous variables which might somehow be related to unobservable characteristics that are correlated to job mobility behaviour early in transition like ability and / or network access. One set of covariates that might be used in this context is self-reported job mobility intentions at the beginning of the transition process (cf. table 1.6). Individuals in firms that had announced mass layoffs in the beginning of 1990 were more likely to fear the loss their job than individuals in firms that had not yet announced mass layoffs (54% compared to 32%) but also more likely to quit their current job (30% compared to 15%). This might explain partially why the self-declared intentions to quit the current job were significantly higher for those individuals facing firm-level mass layoffs (but possibly without good outside options) than for those with good outside options, i.e. those who would not have had any difficulties in finding an equivalent job (25% compared to 15%).

Those movers who initially reported a low probability of quitting - possibly in combination with a low perceived probability of individual job loss - could be characterised as "forced movers". On the other hand, those movers who initially reported a high probability of quitting despite a low probability of job loss are more likely to be "voluntary movers". OLS regressions on various subsamples defined by different combinations of self-reported probabilities of job loss and quit at the beginning of transition yield indeed very different results: The estimated returns to early job mobility are found insignificant for the first group of "forced movers" declaring both a low probability of job loss and a low probability of quitting the job. On the contrary, the returns to "voluntary movers" as defined above amount to around 40% (cf. table 4.16). Again, while not providing any formal proof of the validity or relevance of the above instruments and IV estimates of the returns to early job mobility, these results seem to support the qualitative results based on the above LATE-IV estimation.

4.6 Sensitivity analyses

So far, the estimation results have been proven robust with respect to alternative regression specifications and information sets. This section discusses the robustness of the estimation results with respect to changes in the sample definitions and control groups, changes in the definition of the treatment variable and changes in the definition of the outcome variable. It also discusses some results for the restricted estimation of the full outcome distributions for the subpopulation of compliers when imposing nonnegativity.

4.6.1 Alternative sample definitions and control groups

Table 4.17 contains the estimated returns to early job mobility for various restrictions of the control group. Results in columns (3) of the table are based on samples further excluding all those individuals who were laid off in 1990 or 1991 from both treatment and control groups. As can be seen, the results seem robust with respect to the definition of the control group and generally confirm those obtained above, except probably for IV estimates based on weak instruments.[19]

Excluding further all migrants to the Western part of the country from the sample, however, does only lead to minor changes in the estimated returns.

[19] As discussed above, first-stage F-statistics and partial R^2 measures are considered diagnostic tools in assessing the instrument relevance. As is clear from table 4.17, the instrument relevance differs according to the chosen sample definition and control group. In particular when restricting the control group to "comparable stayers", the "quality" of the various instruments seems problematic.

Table 4.16: OLS estimates of the returns to early (1990-91) job mobility on selected subsamples of "presumed compliers": Overview

self-reported probability of ...	(1)	(2)	(3)	(4)	estimated returns
job loss: high	0.187	0.207	0.205	0.231	20.6 % - -26.0 %
	(0.039)	(0.038)	(0.040)	(0.040)	
R^2	0.359	0.395	0.408	0.414	
N	545	545	545	545	
job loss: low	0.017	0.034	0.019	0.042	0 %
	(0.036)	(0.036)	(0.035)	(0.036)	
R^2	0.412	0.5	0.513	0.511	
N	728	728	728	728	
quitting job: high	0.275	0.281	0.277	0.298	31.7 % - 34.7 %
	(0.052)	(0.052)	(0.051)	(0.052)	
R^2	0.355	0.384	0.395	0.408	
N	294	294	294	294	
quitting job: low	0.030	0.056	0.042	0.067	0 % - 7.0 %
	(0.031)	(0.031)	(0.031)	(0.031)	
R^2	0.416	0.485	0.499	0.501	
N	981	981	981	981	
job loss low and quit high	0.337	0.307	0.363	na	35.9 % - 43.8 %
	(0.105)	(0.109)	(0.126)	(na)	
R^2	0.547	0.593	0.618	na	
N	67	67	67	67	
job loss low and quit low	-0.010	0.010	-0.014	na	0 %
	(0.039)	(0.039)	(0.038)	(na)	
R^2	0.416	0.509	0.525	na	
N	661	661	661	661	

NOTES: dependent variable: log of the discounted sum of labour income 1990-96; treatment variable: early job-to-job mobility (1990-91); results for model specification (3), including firm-size dummies, monthly income 1989 and occupation dummies among the regressors; robust standard error estimates in parentheses; SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284)

Table 4.17: Estimates of the returns to early (1990-91) job mobility with respect to various control groups: Overview

control group	$OLSE_1$	$OLSE_2$	IVE_1	IVE_2	IVE_3	IVE_4	IVE_5
early non-movers	0.090	0.124	-0.937	0.435	0.069	-0.546	-0.326
	(0.026)	(0.034)	(0.304)	(0.242)	(0.276)	(0.191)	(0.169)
R^2/ 1st stage F	0.452	0.503	21.12	14.18	10.82	36.44	20.04
N	1284	1019	1284	1284	1284	1284	1284
early stayers	0.028	0.106	-0.608	0.411	0.197	-0.370	-0.211
	(0.025)	(0.034)	(0.189)	(0.197)	(0.191)	(0.132)	(0.117)
R^2/ 1st stage F	0.450	0.480	29.69	18.26	17.78	43.06	26.07
N	1055	922	1055	1055	1055	1055	1055
"comparable stayers"	0.004	0.054	-0.003	0.660	0.729	0.048	0.325
	(0.046)	(0.056)	(0.426)	(0.560)	(0.517)	(0.353)	(0.310)
R^2/ 1st stage F	0.423	0.513	3.32	3.23	3.78	5.38	3.89
N	336	186	336	336	336	336	336
late or other movers	0.101	0.139	-0.578	0.363	0.323	-0.226	0.088
	(0.027)	(0.035)	(0.372)	(0.179)	(0.245)	(0.185)	(0.144)
R^2/ 1st stage F	0.447	0.501	7.45	19.42	10.57	19.21	14.46
N	860	619	860	860	860	860	860
late movers	0.132	0.162	-0.529	0.445	0.530	-0.024	0.355
	(0.030)	(0.037)	(0.871)	(0.187)	(0.277)	(0.258)	(0.169)
R^2/ 1st stage F	0.459	0.540	1.12	15.49	7.76	7.14	8.92
N	541	345	541	541	541	541	541
overall job stayers	-0.106	-0.030	-0.257	0.157	0.082	-0.204	-0.172
	(0.027)	(0.036)	(0.100)	(0.205)	(0.191)	(0.087)	(0.087)
R^2/ 1st stage F	0.436	0.480	43.38	9.87	13.02	51.78	28.64
N	615	482	615	615	615	615	615

NOTES: dependent variable: log of the discounted sum of total income 1990-96; sample restriction to treatment and control groups as defined in the table; treatment variable: early job-to-job mobility (1990-91); OLS estimation results in column (3),denoted $OLSE_2$, are based on samples excluding all those individuals laid off in 1990 or 1991; results for model specification (3), including firm-size dummies, monthly income 1989 and occupation dummies among the regressors; robust standard error estimates in parentheses; SOURCE: GSOEP, waves 7-14, subsample C (East Germany) (N=1284)

Table 4.18: Variable definitions and descriptive statistics: Labour income throughout transition

variable	description	mean	st. dev.	min	max
TYL90	labour income 1990	12.669,53	6.008,06	0	65.100
TYL91	labour income 1991	20.755,55	10.893,23	0	99.519
TYL92	labour income 1992	21.846,64	12.495,48	0	108.134
TYL93	labour income 1993	23.107,04	14.472,80	0	125.859
TYL94	labour income 1994	24.204,09	15.827,44	0	138.667
TYL95	labour income 1995	24.011,38	17.765,65	0	210.342
TYL96	labour income 1996	23.307,84	17.460,35	0	113.695
PDVTYL06	discounted sum 1990-96	149.902,10	81.753,93	900	735.310

NOTES: All income values gross in 1991 DM. The number of observations differs slightly from that with total income due to zero observations on labour income.; SOURCE: GSOEP, waves 7-14, subsample C (N=1280)

4.6.2 Alternative outcome measures

The above results are also qualitatively robust with respect to changes in the definition of the outcome variable. As alternative outcome variables, the following variables have been considered: (1) total income for the period 1992-96, (2) aggregate labour income throughout transition (cf. descriptive statistics in table 4.18) and (3) yearly total income measures.

Average total income "late" in transition (1992-96) amounts to 139.085,20 DM (with a standard deviation of 74.577,70 DM) for the treated compared to 120.777,40 DM (60.821,40 DM) for the control group. Average total labour income 1990-96 amounts to 173.102,90 DM (with a standard deviation of 88.449,50 DM) for the treated, compared to 141.644,20 DM (77.626,60) for the control group.

In both cases, the null of equal means or equal income distributions across treatment status is clearly rejected. Similar differences in the outcome variable by treatment status are obtained in the case of total yearly incomes as illustrated in table 4.19.

The estimation results on the basis of these different outcome measures (cf. tables 4.20 and 4.21) confirm the findings above. They suggest even bigger returns to early job mobility in terms of labour income and suggest, moreover, stronger income effects in early years that diminish over time. In the case of total labour income as outcome variable, it is important to

Table 4.19: Alternative outcome measures by treatment status: Yearly income measures

treatment group	E[Y|D=1]	E[Y|D=0]	t-test	KS-test
TYI91	24.902,23	20.254,67	-5.307	0.153
	(13.284,65)	(9.369,08)	(0.000)	(0.000)
TYI92	26.395,81	22.033,55	-4.829	0.134
	(13.516,08)	(10.659,18)	(0.000)	(0.001)
TYI94	28.508,08	24.894,74	-3.175	0.149
	(16.942,18)	(13.852,30)	(0.002)	(0.000)
TYI96	27.802,24	24.902,95	-2.591	0.136
	(16.258,03)	(15.411,72)	(0.010)	(0.001)

NOTES: D=early job-to-job movers (1990-91); standard deviations in parentheses; t-test refers to the one-sided two-sample t-test of equal means allowing for unequal variances with alternative $H_1 : E[Y|D = 1] > E[Y|D = 0$;KS-test refers to the two-sample Kolmogorov-Smirnov test statistic of equal outcome distributions for treatment and control group with alternative $H_1 : F_{Y|D=1} < F_{Y|D=0}$;for both tests p-values in parentheses; SOURCE: GSOEP public-use file, wave 7, subsample C, N=1284 note that it does not contain any transfer income which could help to mitigate the financial consequences of periods of non-employment. The results thus suggest that the potential losses or gains from job mobility might partly be due to differences in the incidence and duration of such spells of non-employment across treatment and control groups. This issue will be discussed in the section 4.7 below.

4.6.3 Alternative treatment variables

As alternative treatment variables, an indicator of job mobility in the first (1990) resp. the first three (1990-92) years of transition has been considered. Information on the treatment probabilities by assignment status for these alternative definitions is summarised in tables 4.38 and 4.41, and estimation results for these cases in tables 4.39 and 4.42 in the appendix. Furthermore, multivariate treatment variables have been considered which distinguish between (only) early movers, (only) late movers and frequent movers or between the time of the first job changes respectively. As control group, never-movers (including job losers) or job stayers (excluding job losers) were chosen.

Regression results indicate gains for early movers compared to never-movers (0.085 (0.032)), no gains for frequent movers (0.036 (0.042)), and losses for late movers (-0.070 (0.024)). These results are found to be qualitatively robust when focussing on mobility without layoffs by excluding those individuals from the sample who were laid off at some point in time between 1990 and 1996. When restricting the control group to job stayers (1990-96), any form of job mobility throughout transition is found to have a negative impact on earnings, less so for early movers (-0.096 (0.031)) and strongly for frequent movers (-0.148 (0.040)) and, in particular, for late movers (-0.247 (0.024)). Restricting the analysis further to individuals never laid off, we find that the effects remain strongly negative for late movers (-0.148 (0.036)) while becoming insignificantly different from 0 for either early or frequent movers.

Allowing for multivariate treatments according to the timing of the first job-to-job move, regression results indicate strong gains for those switching job in 1990 (0.19 (0.04)) as opposed to losses for those with first job-to-job shifts in 1993 or later. The results remain virtually unchanged when restricting the analysis to individuals never laid off. When restricting the control group to job stayers (1990-96), any form of job mobility throughout transition is found

Table 4.20: Estimates of the returns to early (1990-91) job mobility with labour income as dependent variable: Overview

	(1)	(2)	(3)	(4)	estimated returns
OLSE	0.162	0.194	0.177	0.210	17.6 % - 23.4 %
	(0.033)	(0.033)	(0.033)	(0.035)	
R^2	0.300	0.350	0.349	0.361	
IVE_1	-1.268	-1.131	-1.277	-1.082	-72.1 % - -66.1 %
	(0.480)	(0.511)	(0.447)	(0.531)	
1st stage F	18.22	14.47	21.04	12.80	
partial R^2	0.0146	0.0120	0.0172	0.0106	
IVE_2	0.837	0.664	0.785	0.601	82.4 % - 130.9 %
	(0.418)	(0.365)	(0.384)	(0.376)	
1st stage F	12.38	14.51	14.24	13.59	
partial R^2	0.0121	0.0140	0.0137	0.0134	
IVE_3	0.116	0.028	0.030	0.119	0.0 % - 12.6 %
	(0.480)	(0.487)	(0.461)	(0.516)	
1st stage F	9.59	9.90	10.81	8.11	
partial R^2	0.0107	0.0110	0.0120	0.0093	
IVE_4	-0.614	-0.442	-0.615	-0.448	-45.9 % - -35.7 %
	(0.301)	(0.295)	(0.279)	(0.307)	
1st stage F	32.08	27.34	36.42	25.40	
partial R^2	0.0230	0.0207	0.0267	0.0194	
IVE_5	-0.310	-0.167	-0.353	-0.147	-29.7 % - -13.7 %
	(0.274)	(0.265)	(0.256)	(0.279)	
1st stage F	17.50	15.92	20.02	14.33	
partial R^2					

NOTES: dependent variable: log of the discounted sum of labour income 1990-96; treatment variable: early job-to-job mobility (1990-91); model specifications (1) to (4) as defined above; robust standard error estimates in parentheses; SOURCE: GSOEP public-use file, wave 7, subsample C, N=1284

Table 4.21: Estimates of the returns to early (1990-91) job mobility with alternative outcome measures: Overview

	TYI91	TYI92	TYI94	TYI96	estimated returns
OLSE	0.149	0.121	0.059	0.018	0.0 % - 16.1 %
	(0.030)	(0.029)	(0.037)	(0.044)	
R^2	0.297	0.349	0.297	0.279	
IVE_1	-0.903	-0.998	-0.774	-0.585	-63.1 % - -44.3 %
	(0.330)	(0.347)	(0.317)	(0.325)	
IVE_2	0.531	0.365	0.760	0.053	0.1 % - 113.8 %
	(0.284)	(0.263)	(0.337)	(0.344)	
IVE_3	0.257	0.128	0.372	-0.095	-0.1 % - 45.1 %
	(0.300)	(0.300)	(0.303)	(0.355)	
IVE_4	-0.532	-0.649	-0.367	-0.411	-47,7 % - -30.7 %
	(0.211)	(0.225)	(0.226)	(0.256)	
IVE_5	-0.257	-0.382	-0.133	-0.295	-31.8 % - -12.5 %
	(0.177)	(0.191)	(0.198)	(0.224)	
N	1283	1277	1259	1223	

NOTES: dependent variable: log of the yearly total income of given year; specification 3 (occupation dummies); treatment variable: early job-to-job mobility (1990-91); results for model specification (3), including firm-size dummies, monthly income 1989 and occupation dummies among the regressors; robust standard error estimates in parentheses; SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284) to have a negative impact on earnings with the exception of job changes in 1990 with insignificant coefficient. The coefficients on mobility indicators for later job changes range from -0.10 to -0.35. Restricting the analysis further to individuals never laid off, the effect of job changes in 1990 becomes positive 0.12 (0.05) while remaining negative for later job changes.

4.6.4 Estimation of full outcome distributions for compliers

Following findings in Imbens and Rubin (1997) that IV estimation results of causal effects might be flawed due to the negativity of the implicitly estimated outcome distributions for compliers, this section presents some results based on an explicit estimation of these outcome distributions imposing non-negativity. As discussed in Imbens and Rubin (1997), the estimates of local average treatment effects may alter substantially when imposing nonnegativity of the underlying density estimate. In this application, however, the qualitative nature of the above IV estimation results is confirmed, although the estimated causal effects turn out to be slightly different from the simple (unconditional) Wald estimates as well as the standard IV estimates conditional on observables.

In the case of the first instrument, firm-level announced layoffs, the nonnegative IV point estimate of the returns to early job mobility remains negative, but smaller in absolute value than either the simple (unconditional) Wald estimate or the standard IV estimate conditional on observables. In the case of the second instrument, occupation-specific job opportunities, and the third instrument, the nonnegative IV point estimate of the returns to early job mobility is found to be even bigger than suggested by either the simple (unconditional) Wald estimate or the standard IVE conditional on observables. When imposing nonnegativity of the outcome distributions, in the case of all instruments the mean of the respective outcome distribution $Y(0)$ for compliers is found to be smaller than the unrestricted estimate.

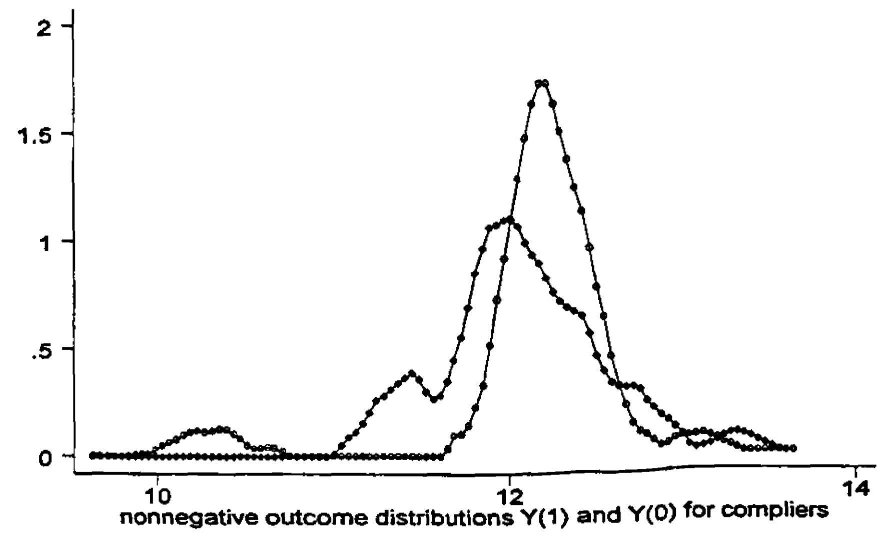
As can be seen from table 4.22, only in the case of the second instrument the mean of the outcome distribution $Y(0)$ for compliers is found to be smaller for compliers than for never-takers. In the case of the first instrument, the mean of the outcome distribution $Y(0)$ for compliers is found to exceed not only the mean of the outcome distribution $Y(0)$ for never-takers but also that of the outcome distribution $Y(1)$ for compliers, thus leading to a negative causal effect. Figures 4-2 to 4-5 below display the estimated nonnegative outcome distributions

Table 4.22: Mean estimates of potential outcomes by subpopulation

	NT $Y(0)$	AT $Y(1)$	C $Y(0)$	C $Y(1)$	C $Y(1)-Y(0)$
Z_1					
simple IVE	11.78	11.94	12.39	12.13	-0.26
nonnegative IVE	11.78	11.94	12.14	12.07	-0.07
Z_2					
simple IVE	11.86	11.97	11.61	12.17	0.56
nonnegative IVE	11.86	11.97	11.49	12.12	0.63
Z_3					
simple IVE	11.78	11.96	11.99	12.39	0.40
nonnegative IVE	11.78	11.96	11.86	12.32	0.46
Z_4					
simple IVE	11.80	11.95	12.11	12.05	-0.06
nonnegative IVE	11.80	11.95	12.02	12.03	0.01

NOTES: dependent variable: log of the total income 1990-96; treatment variable: early job-to-job mobility (1990-91); results for simple model without covariates; SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284)

Figure 4-2: Nonnegative outcome distributions Y(0) and Y(1) for compliers (instr. Z1)

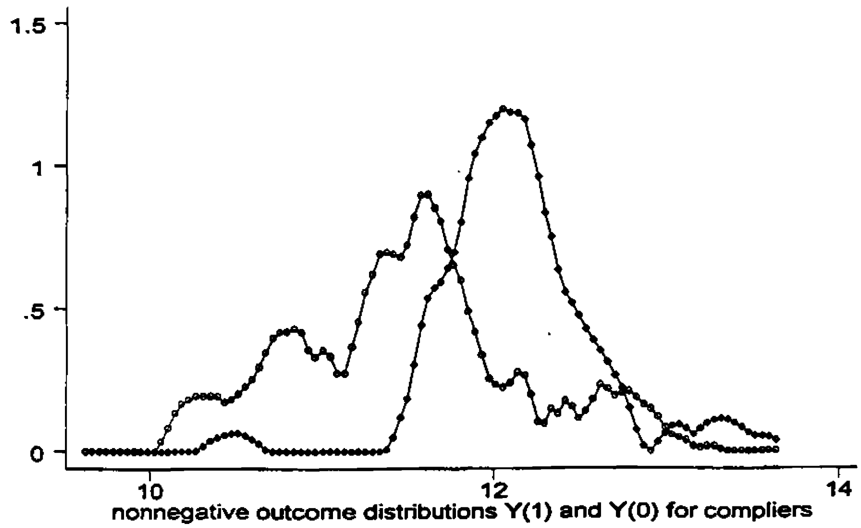


$\circ\, Y(0); \diamond\, Y(1)$

for compliers when imposing nonnegativity.

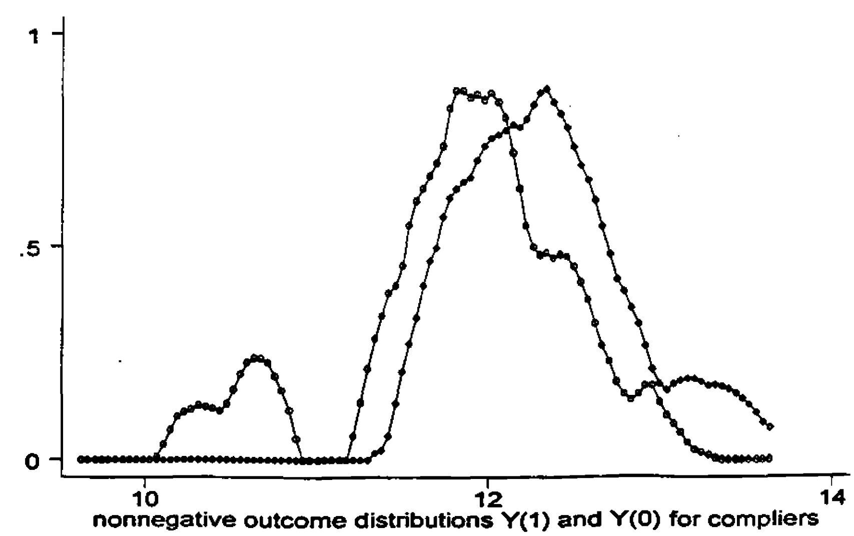
While this section has only illustrated the simple case without covariates, the above qualitative IV estimates of the returns to early job mobility - negative in the case of "forced mobility" and positive in the case of "voluntary mobility" - are corroborated when imposing nonnegativity of the implicit estimates of the full outcome distributions for compliers. In order to integrate covariates into this approach, distributional assumptions are needed for the estimation of the observable outcome distributions by assignment and treatment status. Further research in this direction is necessary to assess the impact of the nonnegativity restriction in this case with covariates.

Figure 4-3: Nonnegative outcome distributions Y(0) and Y(1) for compliers (instr. Z2)



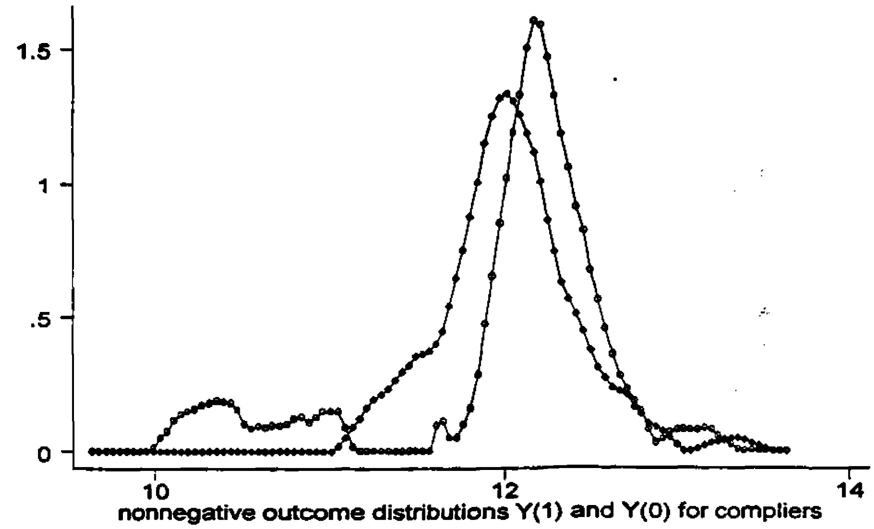
○ $Y(0)$; ◇ $Y(1)$

Figure 4-4: Nonnegative outcome distributions Y(0) and Y(1) for compliers (instr. Z3)



○ $Y(0)$; ◇ $Y(1)$

Figure 4-5: Nonnegative outcome distributions Y(0) and Y(1) for compliers (instr. Z4)



∘ $Y(0)$; ⋄ $Y(1)$

4.7 Employment effects of early job mobility

In the previous sections, it has been shown that the returns to early job mobility in transition are heterogeneous across different subpopulations. The returns to early job mobility for voluntary movers were seen to be positive and marginally significant while those for forced movers were found to be strongly negative. Given that aggregate income information for the whole transition period 1990-96 has been used as outcome measure in the previous analysis, it is clear that the individuals' employment histories themselves are an important determinant of total income and hence of the returns to job mobility in terms of this income.

The differences in the returns to job mobility with respect to total income and labour income that were estimated above suggest indeed that the potential losses or gains from job mobility might partly be due to differences in the incidence and duration of spells of employment and non-employment across treatment and control groups. It is therefore important to recognise that these estimated causal effects each are a combined effect of job mobility on income, on the one hand, and on employment probabilities, on the other, and that, consequently, the "estimates of treatment effects for (income) conditional on employment status do not necessarily have a causal interpretation, even when the unconditional estimates do" (Angrist (1998)). The effect of job mobility on aggregate income throughout the transition process should therefore be decomposed into, first, a direct effect on the employment history throughout transition and, second, an indirect effect on the level of income conditional on the labour market status.

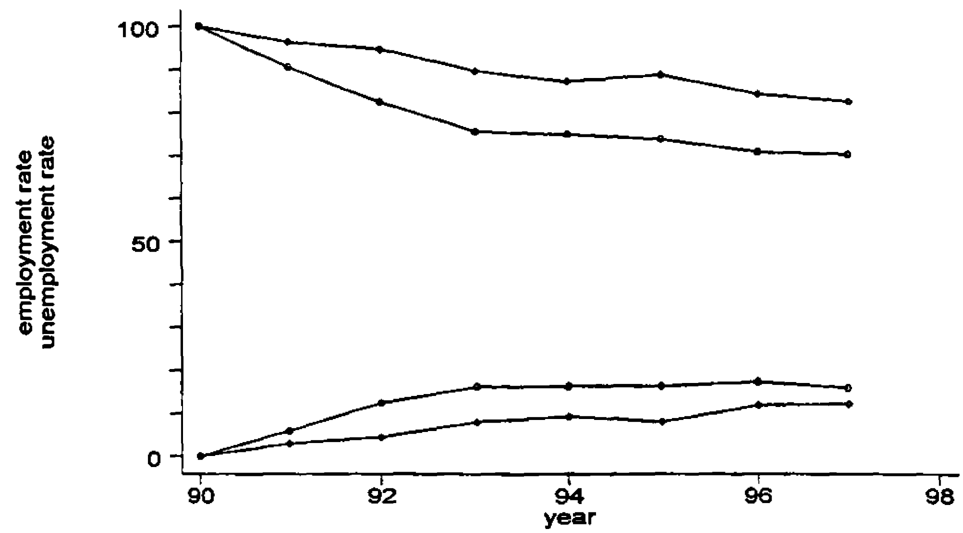
This section tries to separate the unconditional effect of early job mobility on employment histories and employment probabilities throughout transition from the effect of job mobility on income conditional on employment. To this aim, after describing the overall employment histories and the evolution of employment states, it has a closer look at the incidence and duration of employment and non-employment spells by job mobility status. It presents the estimation results for the effect of early job-to-job mobility on employment probabilities on the basis of, first, simple treatment control comparisons, second, exact match comparisons and nonlinear probability models following Card and Sullivan (1988) and Angrist (1998,2000), and third, instrumental variable estimates. The latter are based on the method applied in the previous chapter which also allows the identification of the causal effects of a bivariate treatment variable on the incidence (bivariate outcome) of employment and/or non-employment spells as outcome variable. Finally, potential extensions to the analysis of the returns to job mobility are discussed.

4.7.1 Incidence and duration of employment and non-employment spells by treatment status

This section provides a descriptive analysis of employment probabilities and labour market transitions by treatment status. As above, attention is restricted to early job mobility throughout the first two years of transition (1990-91) as treatment, and complementary evidence for the case of early (1990-92) job mobility as treatment is presented in the annex.

Tables 4.33 to 4.36 in the appendix present the employment evolution during transition by treatment status, and figure 4-6 below compares in particular the evolution of employment and unemployment probabilities during transition by treatment status. Both show that, as conjectured informally by Hunt (1999), "being flexible enough to change (jobs) may have played an important role in remaining employed at all." The results of tests for differences in future

Figure 4-6: Employment and unemployment rates 1990-96 by treatment status



◇treatment group; ○ control group

employment probabilities, transition rates and average durations by treatment status presented in tables 4.23 to 4.26 below seem to confirm this conjecture.

Employment rates, unemployment rates and transition (retention or accession) rates clearly differ by treatment status. Both transition rates and employment rates are found to be significantly higher for early (1990-91) job movers. While the average employment rates in the treatment group exceeded that of the control group by around 10% in the period 1991-97, rates of unemployment and inactivity were significantly lower in the treatment group. Differences in retention and accession probabilities, however, were only significant in 1991/92 and 1994/95. In other years, there do not seem to be any significant differences between the treatment and control group in labour market transitions.[20]

While more individuals in the control group were actually subject to unemployment when looking at the labour market state at the time of the survey, it is interesting to note that there

[20]These differences in employment, unemployment or transition rates can be roughly interpreted as difference-in-difference estimates of the effect of early job mobility, given that all individuals in the sample were employed at the beginning of transition and that labour market dynamics was relatively weak before 1990.

Table 4.23: Employment rates 1990-97 by treatment status

	1991	1992	1993	1994	1995	1996
Pr(E|D=1) : Pr(E|D=0)	0.97:0.91**	0.96:0.82**	0.88:0.77**	0.85:0.76**	0.88:0.75**	0.84:0.72**
Pr(U|D=1) : Pr(U|D=0)	0.03:0.06*	0.02:0.12**	0.09:0.15**	0.10:0.16**	0.08:0.16**	0.12:0.17*
Pr(O|D=1) : Pr(O|D=0)	0.00:0.03**	0.01:0.05**	0.03:0.08**	0.04:0.08**	0.04:0.09**	0.05:0.11**

NOTES: D=early (1990-91) job-to-job mobility; Pr(E|D=.), Pr(U|D=.), and Pr(O|D=.) denote the probability of being in employment, in unemployment and out of the labour force respectively conditional on the treatment status D; * denotes significant t-statistic at the 5%-level for a one-sided two-sample t-test of equal means allowing for unequal variances; ** denotes significant t-statistic at the 1%-level for a one-sided two-sample t-test of equal means allowing for unequal variances; SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284)

Table 4.24: Transition rates 1990-97 by treatment status

	1991/92	1992/93	1993/94	1994/95	1995/96	1996/97
Pr(E|E;D=1) : Pr(E|E;D=0)	0.96:0.87**	0.90:0.87	0.93:0.91	0.95:0.90**	0.91:0.89	0.93:0.92
Pr(U|E;D=1) : Pr(U|E;D=0)	0.02:0.10**	0.08:0.10	0.07:0.07	0.04:0.08*	0.07:0.10	0.06:0.07
Pr(O|E;D=1) : Pr(O|E;D=0)	0.01:0.03**	0.03:0.04	0.00:0.02*	0.00:0.02**	0.02:0.02	0.01:0.01
Pr(E|U;D=1) : Pr(E|U;D=0)	1.00:0.38**	0.33:0.39	0.39:0.35	0.60:0.35*	0.38:0.28	0.23:0.28
Pr(E|O;D=1) : Pr(E|O;D=0)		0.50:0.15	0.13:0.51	0.18:0.11	0.10:0.09	0.33:0.08

NOTES: D=early (1990-91) job-to-job mobility; Pr(E|E;D=.), Pr(U|E;D=.), and Pr(O|E;D=.) denote the probability of staying in employment, moving into unemployment and moving into inactivity respectively conditional on the treatment status D; Pr(E|U;D=.) and Pr(E|O;D=.) denote the probability of moving into employment out of unemployment and inactivity, respectively, conditional on the treatment status D; * denotes significant t-statistic at the 5%-level for a one-sided two-sample t-test of equal means allowing for unequal variances; ** denotes significant t-statistic at the 1%-level for a one-sided two-sample t-test of equal means allowing for unequal variances; SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284)

Table 4.25: Cumulative non-employment experience by treatment status

	1991	1992	1993	1994	1995	1996
not employed						
total	0.03:0.09**	0.06:0.21**	0.15:0.32**	0.22:0.36**	0.25:0.41**	0.31:0.46**
unemployed	0.03:0.06*	0.05:0.15**	0.12:0.24**	0.18:0.28**	0.22:0.32**	0.26:0.38**
out of the LF	0.00:0.03**	0.01:0.07**	0.03:0.11**	0.05:0.13**	0.06:0.16**	0.08:0.18**
calendar inform.						
unemployed	0.10:0.07	0.17:0.12*	0.23:0.23	0.30:0.30	0.32:0.33	0.37:0.38
out of the LF	0.11:0.07*	0.13:0.11	0.16:0.15	0.17:0.18	0.19:0.21	0.22:0.23

NOTES: in % of restricted sample (N=1284); percentage of individuals who were in the respective non-employment state at least at one survey date up to the respective year; calendar information reports the percentage of individuals who were either in unemployment or out of the labour force (vocational training) for at least one month up to the respective year and is based on monthly calendar information; SOURCE: GSOEP public-use file, waves 7-14, subsample C

Table 4.26: Average duration of employment and non-employment spells by treatment status

	1990	1991	1992	1993	1994	1995	1996
av. duration in							
employment	11.2:11.2	10.9:10.5**	11.4:10.9**	11.2:10.8**	11.4:11.1**	11.4:11.1*	11.2:11.2
unemployment	2.4:3.0	2.6:5.7**	4.2:6.0**	5.3:6.7*	6.0:6.6	5.2:6.8*	5.5:6.5
out of the LF	4.1:3.2	2.8:5.1*	3.7:6.3**	4.4:6.1	6.1:5.7	6.7:6.0	6.6:6.7

NOTES: in % of restricted sample (N=1284); average duration (in months) in the respective labour market state; SOURCE: GSOEP public-use file, waves 7-14, subsample C were no differences in the incidence of intervening unemployment spells (on the basis of the calendar information) between the treatment and control group. In 1992, interestingly, those in the treatment group were more likely to have experienced unemployment than those in the control group, while those in the control group were initially slightly more likely to leave the labour force. Furthermore, those in the control group had on average shorter employment and longer unemployment spells up to 1993: After that, differences in the average duration of unemployment spells between treatment and control group vanished.

4.7.2 Selection on covariates

As in the case of income effects of early job mobility, however, given the potential non-randomness of the treatment group, one has to ask again whether these differences in the employment histories by treatment status are *caused* by the respective mobility behaviour. Since the transition and employment probabilities themselves are potentially causal outcomes of job mobility, neither controlling for intervening unemployment spells in the estimation procedure employed in the previous chapter nor comparing employment outcomes by nature of job separations (layoffs or quits) would be valid approaches to estimate the causal effects of job mobility on either income or employment.

Card and Sullivan (1988) propose two approaches to the estimation of the effect of treatment - training in their case, early job mobility here - on employment and transition probabilities: first, exact-match comparisons of employment outcomes for individuals with identical pre-treatment employment histories (so-called "exact matches" or "controlled contrasts"), and second, an estimation of non-linear probability models with employment as bivariate.outcome variable. Results from these approaches are summarised below and will be contrasted with instrumental variables estimates using the instruments suggested in section 4.5.

Comparison of exact matches

In the context of this study, "pre-treatment" employment histories - i.e. employment histories before 1990 - are unknown. Given the sample selection, all individuals were further employed at the time of the survey in 1990. The employment histories as from 1990, finally, are already part of the outcome. The proposal to estimate the employment effects for the period after 1991 by conditioning on the employment history up to that time does therefore not seem promising. Instead, similar to the approach taken in section 4.6, one could argue that individuals might be matched according to their job mobility intentions at the beginning of transition, namely their perceptions regarding the probability of losing their job, quitting their job, searching actively a new job or becoming self-employed. Individuals with similar self-declared job mobility intentions can be considered more comparable than all individuals. After "matching" individuals by their job mobility intentions at the beginning of transition, the average employment and unemployment rates are calculated for each type of match for both treated and controls. Overall estimates of the effect of job mobility on employment and transition rates are obtained by weighting the results for the single matches by the sample fractions of early job-to-job movers - the treated - for each type of match.[21]

As can be seen from tables 4.27 and 4.28 below, early job movers seem to benefit from both higher employment and lower unemployment probabilities throughout transition. While the effect on employment probability seems to last over the whole period 1991-97, the effect on unemployment probabilities is found to decline over time. These results from exact match comparisons are thus similar to those obtained when comparing mean employment or unemployment rates by treatment status. Interestingly, when analyzing the group-specific controlled contrasts, the effects of early job-to-job mobility on both employment and unemployment probabilities are strongest for those individuals who had announced a high likelihood of either job loss or job search, or both, at the beginning of the transition process. They are insignificant, on the other hand, for those who had announced that they were unlikely to lose their job, the more so when they were also unlikely to search for a new job. This finding corroborates the above hypothesis that "forced movers" - those who had to change job although they neither intended nor expected to have to do so - seemingly had to incur substantial losses in terms of employment probabilities throughout transition.

[21] Propensity score estimation represents an extension to this simple but intuitive method, defining "matches" on the model-based estimates of the individual-specific treatment probabilities. Self-declared job mobility intentions might be included in the underlying model to estimate these propensities.

Probability models

Estimation results for a simple (bivariate probit) employment probability model show further that men and younger people, individuals with vocational training or an engineering or university diploma, those with relatively high monthly earnings already in 1989, white collar workers and those working in the public sector in 1990 benefitted from significantly higher employment probabilities in the period 1990-97. On the contrary, women, older people, those without vocational training and those in low hierarchical positions in 1990 had significantly lower employment probabilities in the same period. People working in the energy, trade, transport or public sectors in 1990 had further significantly higher employment rates throughout 1990-97 than those in manufacturing, mining or agriculture. The estimation results suggest returns to early job-to-job mobility in terms of an increase in the average employment probability by 7-8%.

Estimation results from a similar model for the probability of unemployment show that, similarly, unemployment rates throughout 1990-97 were significantly higher for women, individuals without vocational training and those in low hierarchical positions in 1990, while significantly lower for those with engineering or university diploma and those working in the transport or public sector in 1990. The estimation results further suggest returns to early job-to-job mobility through a reduction in the probability of unemployment of around 3%. The estimation results on the effect of the treatment on unemployment probabilities year by year, some doubts arise as to whether early job mobility was effective in reducing the probability of becoming unemployed later in transition. In fact, the estimated returns for the years 1994 and 1996-97 are found to be insignificant.

4.7.3 Instrumental variables estimation

The results presented above cannot necessarily be interpreted as causal effects of early job mobility on employment outcomes throughout transition.[22] Instead of aiming at an improved

[22] As Card and Sullivan (1988) mention in the context of estimating the employment effects of training: "If the probability of remaining employed from one year to the next (the retention probability) and the probability of moving from unemployment to employment (the accession probability) are the same for a given individual, then a simple comparison of relative changes in employment probabilities among the trainees and controls provides a consistent estimate of the training effect in the context of a linear probability model. With state dependence

panel data model of employment probabilities, allowing explicitly for state dependence and nonlinearities, as suggested by Card and Sullivan (1988), a different solution is chosen here following Angrist (2000). Instrumental variables estimates are obtained by standard two-stage least squares methods. The estimation results of the first stage are equivalent to those in the previous chapter and are therefore not repeated here. The second stage results are based on a simple linear probability model and are presented in the last two columns of tables 4.27 and 4.28. Second stage estimation based on nonlinear probability models led to almost identical results which are not reported here.

Not surprisingly, instrumental variables estimates of the returns to early job mobility are also found to differ substantially from the results above. In the case of the first instrument, firm-level announced layoffs, early job-to-job mobility is found to have a strong negative impact on employment probabilities only in the beginning of the transition process and at the same time a strong positive impact on the probability of becoming unemployed throughout the whole transition process. The effects were generally strongest in 1992/93 and have since then declined. On the contrary, in the case of the second instrument, occupation-specific labour market opportunities, early job mobility is found to have a positive impact on the probability of employment and a negative impact on the probability of unemployment in particular in the years 1993-95.

The interpretation of the IV estimates of the returns to early job-to-job mobility in terms of employment probabilities follows that of section 4.5: While "forced movers" seem to have lost in particular due to the fact that early job mobility has in many cases not led to job stability later in transition, "voluntary movers" seem to have benefitted from an increased employment probability throughout the whole transition period. This difference is most obvious in the years 1994-95 that are marked by a slowdown in the economic activity in Germany in general and in East Germany in particular. Contrary to "voluntary movers", "forced movers" seem disproportionately affected by this slowdown.

For "voluntary movers", the employment effect of early job mobility seems to explain entirely the positive effect of early job mobility on total income. For "forced movers", too, the effect on

or with nonlinear probability specifications, this technique will not necessarily eliminate permanent differences between trainees and controls."

Table 4.27: Estimated effects of early (1990-91) job mobility on employment: Overview

	Difference in means	Exact matches	Probit	IVE_1	IVE_2
1991	0.059	0.072	0.047	-0.274	0.108
	(0.014)	(0.011)	(0.014)	(0.166)	(0.153)
1992	0.138	0.124	0.110	-0.525	0.191
	(0.017)	(0.021)	(0.017)	(0.234)	(0.192)
1993	0.112	0.095	0.075	-0.478	0.422
	(0.024)	(0.028)	(0.026)	(0.237)	(0.223)
1994	0.094	0.081	0.048	-0.311	0.535
	(0.026)	(0.029)	(0.028)	(0.224)	(0.242)
1995	0.133	0.122	0.104	-0.420	0.572
	(0.025)	(0.028)	(0.026)	(0.234)	(0.237)
1996	0.123	0.117	0.082	-0.404	0.292
	(0.027)	(0.030)	(0.030)	(0.239)	(0.225)
1997	0.108	0.090	0.061	-0.432	0.482
	(0.028)	(0.031)	(0.031)	(0.237)	(0.237)

NOTES: dependent variable: employment rate at the time of the survey; treatment variable: early job-to-job mobility (1990-91); results for model specification (3), including firm-size dummies, monthly income 1989 and occupation dummies among the regressors; standard error estimates in parentheses; SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284)

Table 4.28: Estimated effects of early (1990-91) job mobility on unemployment: Overview

	Difference in means	Exact matches	Probit	IVE_1	IVE_2
1991	-0.026	-0.039	-0.022	0.345	-0.223
	(0.013)	(0.009)	(0.013)	(0.149)	(0.135)
1992	-0.100	-0.090	-0.084	0.527	-0.259
	(0.014)	(0.017)	(0.014)	(0.216)	(0.176)
1993	-0.064	-0.054	-0.043	0.759	-0.517
	(0.021)	(0.024)	(0.022)	(0.261)	(0.222)
1994	-0.061	-0.056	-0.035	0.562	-0.471
	(0.022)	(0.024)	(0.023)	(0.235)	(0.221)
1995	-0.076	-0.075	-0.055	0.599	-0.615
	(0.021)	(0.023)	(0.021)	(0.239)	(0.233)
1996	-0.053	-0.055	-0.031	0.380	-0.118
	(0.024)	(0.026)	(0.024)	(0.220)	(0.206)
1997	-0.042	-0.027	-0.015	0.449	-0.114
	(0.023)	(0.027)	(0.024)	(0.221)	(0.201)

NOTES: dependent variable: unemployment rate at the time of the survey; treatment variable: early job-to-job mobility (1990-91); results for model specification (3), including firm-size dummies, monthly income 1989 and occupation dummies among the regressors; standard error estimates in parentheses; SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284) employment probabilities explains a substantial part of the negative impact of early job mobility on total income throughout transition. Given the difference in the size of the estimated effects on total income, on the one hand, and on employment probabilities, on the other, however, there seems further a negative income effect of early job mobility conditional on the employment status. Future work will be necessary to model explicitly the level of income conditional on employment, allowing for endogeneity of the job mobility status.

4.8 Some frequently asked questions

Given that the assumptions underlying instrumental variables methods are generally untestable and that, moreover, the subgroup of compliers is not identifiable by definition, results based on these methods are usually greeted with some scepticism in the economics profession. This section tries to assuage some of the most prominent doubts arising in the discussion of the above presented results.

4.8.1 Lack of theory

Wouldn't it be more adequate to estimate the returns to job mobility on the basis of some explicit latent index formulation of the underlying decision-making problem?

As shown in Heckman and Vytlacil (1999), such an approach is not necessarily opposite to the one used in this chapter since the LATE has an interpretation within a latent variable model as some appropriately weighted average of the so-called local instrumental variable.

4.8.2 Lack of model evaluation

Wouldn't it be more adequate to employ more than one instrument and test for overidentifying restrictions instead of using only one instrument?

It is certainly possible to use more than one instrument and test for overidentifying restrictions. It is, however, not at all clear whether the estimates could still be interpreted in a quasi-experimental framework. In addition, the most important tool for evaluation in this kind of study is the detailed discussion of the inherently untestable assumptions underlying AIR(1996)'s theorem of the interpretability of instrumental variables estimates as local average treatment effects. Hence, for a final answer to the above question, one has to weigh the advantages of interpretability of the estimates as causal effects against the lack of formal tests of overidentifying restrictions.

4.8.3 Implausibility of the estimates

Aren't the estimated returns just imprecisely estimated "unbelievable numbers"?

It is certainly fair to say that the estimates presented above are rather imprecise. Given the counterfactual interpretation of the estimated returns as well as the consideration of potential unemployment spells in the aggregate outcome measure, however, the estimated numbers are not necessarily implausible. Assuming for simplicity that monthly labour income is the only invariable potential source of income over the whole period 1990-96, one easily calculates that an individual who has changed to a new job before the end of 1992, who is laid off from this new job after only a short time, and who subsequently remains without job would incur (counterfactual) losses of more than 50% of his potential income for the whole period 1990-96 in case he would have remained employed at his initial firm over the whole period had he not changed to a new job. Clearly, these losses could even be higher when taking into account income growth over time, tenure effects, seniority pay etc..

4.8.4 Restriction to compliers

Given that the LATE applies exclusively to the group of compliers, why should the LATE be considered an economically interesting parameter?

The economic interest in the *LATE* for a subgroup of compliers is based on recognising that, first, treatment effects are heterogeneous in the population, and second, that the compliers are considered a theoretically interesting subgroup. Not only in the context of a transitional labour market there are strong theoretical reasons for heterogeneity of the returns to job mobility in the population. Moreover, the distinction between the returns to forced job mobility and those to voluntary job mobility, as suggested by the instruments used in the above empirical analysis, is obviously of interest. Consequently, the *LATE* should be considered more interesting than the treatment effect on the treated also from a policy point of view.

4.8.5 Unidentifiability of compliers

Given that one cannot identify the group of compliers, why should the LATE be considered an economically interesting parameter?

In the case of endogenous treatment status of (early) job mobility, the *LATE* is actually the only causal effect that can be identified at all by instrumental variables techniques. Additionally, it has been shown in Siebern (1999) that the compliers can actually be characterised theoretically in terms of their unobservable characteristics on the basis of a simple economic model of individual decision-making. Therefore, the *LATE* is an important structural parameter of economic interest in the evaluation of the returns to job mobility during transition.

4.8.6 Characterisation of compliers

Isn't the characterisation of compliers on the basis of the instrument as forced movers and voluntary movers just arbitrary?

Maybe. But the alternative definition of these two types of movers on the basis of survey information on the nature of previous job losses has to be considered as at least as arbitrary, the more so as e.g. individuals who have recently quit their job due to a perceived high probability of layoff cannot necessarily be considered as voluntary movers. Compliance with one of the above suggested instruments should capture these subtleties much better.

4.8.7 Negative average returns

How can returns on average be negative if individuals act rationally?

Noting that the *LATE* estimates reported above are estimates of the realised (*ex post*) returns to job mobility and that the average is calculated for the subgroup of compliers only, there are obviously a wealth of possible reasons why it can be negative. Recall that, in the case of firm-level announced layoffs, the compliers can be shown theoretically to be individuals with either low ability or weak networks. Given that these individuals might actually not improve upon the quality of their job match or their employment prospects by switching to a new employer early, it should not come as a surprise that they might fail to benefit from their decision. Moreover, one should bear in mind that e.g. positive average returns to job mobility for voluntary movers, i.e. for those individuals who only switch job early due to the availability of equivalent outside options, imply, by definition, negative returns to staying on the job for all those individuals who would have changed job if there had been equivalent outside options. Hence, given the counterfactual definition of the *LATE*, the motivations for the above question remain unclear.

4.8.8 Bad assignment luck

Isn't the LATE just measuring the effect of bad luck, namely of being in a firm that decides to announce mass layoffs early?

Considering individuals in "bad" firms or "bad" occupations as unlucky has no bearing at all on the interpretation of the above presented IV estimates as *causal effects* of (early) job mobility. Otherwise, one would logically also have to interpret the ineffectiveness (or, in other words, the negative *causal effect*) of some medicament in a clinical trial as bad luck instead, the more so as it might result e.g. in premature death of the treated as opposed to the individuals in the control group.

4.9 Summary and conclusions

A simple look at correlation measures between job mobility indicators and earnings - usually found to be significantly positive of a magnitude around 10% - gives the misleading impression that job mobility early in the transition process was advantageous. While this result might hold true on average in the whole population there is sufficient theoretical reason to believe that returns to early job mobility do actually vary considerably across individuals. The use of interaction terms of the job mobility indicator and the survey information on the reason for leaving the previous job allowed some researchers to recover such heterogeneity. However, in general these estimates cannot be interpreted as *causal effects* of job mobility on earnings. This chapter has tried to overcome this problem by analysing the returns to job mobility in a quasi-experimental framework. The resulting instrumental variables estimates of the *causal effects* to (early) job-to-job mobility during the transition from a centrally planned economy to a market economy in this framework suggest that returns do actually vary enormously in the population: while forced movers had to incur substantial losses from changing their job of up to -60% in terms of earnings, voluntary movers might have gained up to +60% from moving.

A substantial part of the differences in the returns to early job mobility across different subgroups of compliers can be explained by differences in employment probabilities throughout transition. These differences do not only occur at the beginning of transition but are, in some cases, equally strong in later years. In particular "forced movers" have probably in many cases not been able to find a new job which guarantees job stability throughout the transition process and were adversely affected by either individual layoff or increased difficulties of new companies later in the transition process.

Several extensions to the analysis presented in this chapter would be useful. Besides modelling the effect of early job mobility on income conditional on employment, the above instrumental variables estimates based on standard 2SLS should be improved by applying more appropriate models such as the causal-IV methods proposed by Abadie (1999). Furthermore, the analysis could be extended to other outcome variables such as the count or the duration of employment and unemployment spells. Finally, the analysis could be extended to the estimation of distributional effects of job mobility such as quantile treatment effects.

4.10 References

Abadie, A. (1999), Semiparametric estimation of instrumental variable models for causal effects, Massachusetts Institute of Technology, mimeo

Angrist, J.D. (1998), Estimating the labor market impact of voluntary military service using social security data on military applicants, Econometrica, 66(2), 249-288

Angrist, J.D. (2000), Estimation of limited-dependent variable models with dummy endogenous regressors: Simple strategies for empirical practice, NBER Technical Working Paper No. 248

Angrist, J., G.W. Imbens and D.W. Rubin (1996), Identification of causal effects using instrumental variables, Journal of the American Statistical Association, 91(434), 444-472

Angrist, J.D. and A.B. Krueger (1998), Empirical strategies in labor economics, Massachusetts Institute of Technology Working Paper No. 98-7, Cambridge, forthcoming in: O. Ashenfelter and D. Card (eds) (1999), Handbook of Labor Economics, Vol. 3A, Elsevier Science

Publishers

Bound, J., D.A. Jaeger and R.M. Baker (1995), Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak, Journal of the American Statistical Association, 90(430), 443-450

Card, D. and D. Sullivan (1988), Measuring the effect of subsidised training programs on movements in and out of employment, Econometrica, 56(3), 497-530

DIW (ed.) (1998), The German Socio-Economic Panel (GSOEP) Desktop Companion, Berlin

Heckman, J.J. (1997), Instrumental variables, A study of implicit behavioral assumptions used in making program evaluations, Journal of Human Resources, 441-462

Heilemann, U. (1991), The economics of German unification: A first appraisal, Konjunkturpolitik, 37(3), 127-155

Hunt, J. (1999), Post-unification wage growth in East Germany, NBER Working Paper No. 6878

Imbens, G.W. and J.D. Angrist (1994), Identification and estimation of local average treatment effects, Econometrica, 62(2), 467-475

Imbens, G.W. and D.W. Rubin (1997), Estimating outcome distributions for compliers in instrumental variables models, Review of Economic Studies, 64, 555-574

Mincer, Jacob (1986), Wage changes in job changes, Research in Labor Economics, JAI Press, Volume 8, Part A, 171-197

Simmonet, V. (1998), The role of unobservables in the measurement of return to job mobility: Evidence from Germany and the United States, Lamia, Université Paris I Panthéon-Sorbonne, mimeo

Staiger, D. and J.H. Stock (1997), Instrumental variables regression with weak instruments, Econometrica, 65, 557-586

Wagner, G., R.V. Burkhauser and F. Behringer (1993), The English language public use file of the German socio-economic panel, Journal of Human Resources, 28, 429-433

4.11 Appendix

Table 4.29: Descriptive statistics: Exogenous variables

Variable	mean	st. dev.	min	max
Individual characteristics				
sex (female=1)	0.48	0.50	0	1
age (in years)	38.7	9.5	20	55
8 years schooling	0.30	0.46	0	1
10 years schooling	0.53	0.50	0	1
12 years schooling	0.17	0.38	0	1
no vocational degree	0.03	0.17	0	1
vocational degree	0.61	0.49	0	1
masters degree	0.06	0.24	0	1
technical degree	0.19	0.39	0	1
university degree	0.11	0.32	0	1
Job / firm characteristics				
tenure (in months)	154	116	0	489
small firm (<20)	0.09	0.29	0	1
medium firm	0.30	0.46	0	1
large firm (>200)	0.61	0.49	0	1
public service	0.37	0.48	0	1
monthly income 1989	1092	436	150	4300
Occupation dummies				
agriculture	0.10	0.29	0	1
blue collar	0.37	0.48	0	1
white collar	0.54	0.50	0	1
low position	0.18	0.38	0	1
medium position	0.56	0.50	0	1
leading position	0.27	0.44	0	1
Industry dummies				
agriculture	0.11	0.32	0	1
energy	0.03	0.16	0	1
mining	0.03	0.17	0	1
manufacturing	0.28	0.45	0	1
construction	0.07	0.26	0	1
trade and retail	0.08	0.27	0	1
transport	0.08	0.27	0	1
finance/other services	0.05	0.22	0	1
public service	0.27	0.44	0	1

SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284)

Table 4.30: Descriptive statistics: Exogenous variables by treatment status

Variable	D=0	D=1	t-statistic (p-value)
Individual characteristics			
sex (female=1)	0.51 (0.50)	0.40 (0.49)	3.15 (0.002)
age (in years)	39.0 (9.6)	37.5 (9.0)	2.30 (0.022)
8 years schooling	0.32 (0.47)	0.22 (0.42)	3.16 (0.002)
10 years schooling	0.52 (0.50)	0.54 (0.50)	-0.56 (0.574)
12 years schooling	0.16 (0.36)	0.23 (0.42)	-2.61 (0.010)
no vocational degree	0.03 (0.17)	0.02 (0.14)	1.17 (0.241)
vocational degree	0.61 (0.49)	0.60 (0.49)	0.36 (0.719)
masters degree	0.06 (0.25)	0.05 (0.23)	0.64 (0.520)
technical degree	0.19 (0.39)	0.18 (0.38)	0.44 (0.663)
university degree	0.10 (0.31)	0.15 (0.36)	-1.91 (0.057)
Job / firm characteristics			
tenure (in months)	161 (118)	125 (104)	4.84 (0.000)
small firm (<20)	0.09 (0.29)	0.08 (0.27)	0.87 (0.386)
medium firm	0.29 (0.46)	0.31 (0.46)	-0.50 (0.614)
large firm (>200)	0.61 (0.49)	0.61 (0.49)	0.05 (0.962)
public service	0.38 (0.49)	0.29 (0.45)	2.96 (0.003)
monthly income 1989	1091 (442)	1098 (415)	-0.21 (0.834)
Occupation dummies			
agriculture	0.09 (0.28)	0.14 (0.34)	-2.14 (0.033)
blue collar	0.38 (0.48)	0.32 (0.47)	1.71 (0.088)
white collar	0.54 (0.50)	0.54 (0.50)	-0.23 (0.822)
low position	0.19 (0.39)	0.14 (0.35)	1.99 (0.047)
medium position	0.55 (0.50)	0.58 (0.50)	-0.73 (0.468)
leading position	0.26 (0.44)	0.29 (0.45)	-0.81 (0.421)
Industry dummies			
agriculture	0.10 (0.30)	0.17 (0.38)	-2.75 (0.006)
energy	0.03 (0.17)	0.01 (0.11)	2.28 (0.023)
mining	0.03 (0.17)	0.03 (0.16)	0.28 (0.778)
manufacturing	0.28 (0.45)	0.30 (0.46)	-0.82 (0.413)
construction	0.07 (0.25)	0.08 (0.27)	-0.68 (0.492)
trade and retail	0.07 (0.26)	0.10 (0.30)	-1.28 (0.200)
transport	0.09 (0.29)	0.03 (0.17)	4.45 (0.000)
finance/other services	0.06 (0.23)	0.04 (0.19)	1.22 (0.224)
public service	0.27 (0.44)	0.24 (0.43)	1.06 (0.289)
N	1025	259	1284

NOTES: D=early job-to-job mobility (1990-91); standard deviation and p-values in parentheses;
SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284)

Table 4.31: Descriptive statistics: Exogenous variables by assignment status (instr. Z1)

Variable	Z_1=0	Z_1=1	t-statistic (p-value)
Individual characteristics			
sex (female=1)	0.55 (0.50)	0.41 (0.49)	4.93 (0.000)
age (in years)	38.8 (9.5)	38.6 (9.4)	0.45 (0.650)
8 years schooling	0.30 (0.46)	0.30 (0.46)	0.21 (0.832)
10 years schooling	0.54 (0.50)	0.51 (0.50)	0.99 (0.324)
12 years schooling	0.16 (0.36)	0.19 (0.39)	-1.56 (0.120)
no vocational degree	0.04 (0.21)	0.01 (0.11)	3.59 (0.000)
vocational degree	0.58 (0.49)	0.64 (0.48)	-1.89 (0.059)
masters degree	0.06 (0.24)	0.06 (0.25)	-0.36 (0.720)
technical degree	0.21 (0.41)	0.16 (0.37)	1.93 (0.054)
university degree	0.11 (0.31)	0.12 (0.33)	-0.99 (0.321)
Job / firm characteristics			
tenure (in months)	153 (115)	155 (118)	-0.34 (0.738)
small firm (<20)	0.13 (0.34)	0.04 (0.20)	5.75 (0.000)
medium firm	0.34 (0.47)	0.25 (0.43)	3.70 (0.000)
large firm (>200)	0.52 (0.50)	0.71 (0.45)	-7.15 (0.000)
public service	0.50 (0.50)	0.21 (0.41)	11.54 (0.000)
monthly income 1989	1049 (399)	1142 (470)	-3.77 (0.000)
Occupation dummies			
agriculture	0.10 (0.30)	0.09 (0.29)	0.30 (0.765)
blue collar	0.32 (0.47)	0.42 (0.49)	-3.78 (0.000)
white collar	0.58 (0.49)	0.49 (0.50)	3.54 (0.000)
low position	0.19 (0.39)	0.16 (0.37)	1.35 (0.178)
medium position	0.54 (0.50)	0.57 (0.50)	-0.82 (0.414)
leading position	0.26 (0.44)	0.27 (0.44)	-0.18 (0.861)
Industry dummies			
agriculture	0.10 (0.30)	0.13 (0.33)	-1.43 (0.153)
energy	0.03 (0.18)	0.02 (0.14)	1.53 (0.126)
mining	0.02 (0.16)	0.03 (0.18)	-1.05 (0.294)
manufacturing	0.18 (0.38)	0.39 (0.49)	-8.64 (0.000)
construction	0.06 (0.23)	0.09 (0.28)	-2.24 (0.025)
trade and retail	0.07 (0.26)	0.08 (0.28)	-0.67 (0.503)
transport	0.12 (0.32)	0.04 (0.20)	5.26 (0.000)
finance/other services	0.04 (0.20)	0.06 (0.24)	-1.40 (0.161)
public service	0.37 (0.48)	0.15 (0.36)	9.35 (0.000)
N	683	601	1284

NOTES: Z_1= firm-level announced layoffs as of 1990; standard deviation and p-values in parentheses;
SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284)

Table 4.32: Descriptive statistics: Exogenous variables by assignment status (instr. Z2)

Variable	Z_2=0	Z_2=1	t-statistic (p-value)
Individual characteristics			
sex (female=1)	0.50 (0.50)	0.38 (0.49)	3.38 (0.001)
age (in years)	39.5 (9.5)	35.0 (8.6)	6.87 (0.000)
8 years schooling	0.31 (0.46)	0.24 (0.43)	2.20 (0.029)
10 years schooling	0.50 (0.50)	0.67 (0.47)	-4.62 (0.000)
12 years schooling	0.19 (0.39)	0.09 (0.29)	4.02 (0.000)
no vocational degree	0.03 (0.17)	0.03 (0,17)	0.06 (0.950)
vocational degree	0.58 (0.49)	0.75 (0.44)	-4.95 (0.000)
masters degree	0.07 (0.25)	0.04 (0.19)	1.96 (0.051)
technical degree	0.20 (0.40)	0.14 (0.35)	2.06 (0.040)
university degree	0.13 (0.33)	0.05 (0.21)	4.51 (0.000)
Job / firm characteristics			
tenure (in months)	162 (118)	117 (97)	5.83 (0.000)
small firm (<20)	0.08 (0.27)	0.14 (0.35)	-2.39 (0.017)
medium firm	0.30 (0.46)	0.25 (0.43)	1.72 (0.087)
large firm (>200)	0.61 (0.49)	0.60 (0.49)	0.37 (0.715)
public service	0.36 (0.48)	0.40 (0.49)	-1.26 (0.210)
monthly income 1989			0.88 (0.377)
Occupation dummies			
agriculture	0.10 (0.30)	0.08 (0.26)	1.22 (0.223)
blue collar	0.34 (0.47)	0.51 (0.50)	-4.71 (0.000)
white collar	0.56 (0.50)	0.41 (0.49)	4.04 (0.000)
low position	0.18 (0.38)	0.19 (0.39)	-0.39 (0.699)
medium position	0.54 (0.50)	0.64 (0.48)	-2.74 (0.006)
leading position	0.28 (0.45)	0.17 (0.38)	3.74 (0.000)
Industry dummies			
agriculture	0.12 (0.32)	0.09 (0.29)	1.38 (0.168)
energy	0.03 (0.16)	0.03 (0.17)	-0.09 (0.930)
mining	0.03 (0.17)	0.02 (0.15)	0.63 (0.530)
manufacturing	0.30 (0.46)	0.19 (0.39)	3.67 (0.000)
construction	0.06 (0.23)	0.14 (0.34)	-3.18 (0.002)
trade and retail	0.07 (0.26)	0.11 (0.31)	-1.50 (0.134)
transport	0.08 (0.27)	0.08 (0.26)	0.35 (0.724)
finance/other services	0.05 (0.22)	0.06 (0.24)	-0.60 (0.550)
public service	0.26 (0.44)	0.29 (0.46)	-0.93 (0.354)
N	1071	213	1284

NOTES: Z_2 = occupation-specific labour market opportunities as of 1990; standard deviation and p-values in parentheses; SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284)

Table 4.33: Labour market status 1990-97 in case of non-treatment

	D=0						
LM status	1990	1991	1992	1993	1994	1995	1996
employed	100.00	91.02	82.34	76.88	75.90	74.73	71.51
full-time	88.10	82.44	75.90	70.05	67.90	66.83	63.12
part-time	11.90	8.00	5.85	5.56	6.54	6.93	7.41
(re)training		0.39	0.20	1.17	0.98	0.39	0.49
marginally		0.20	0.39	0.10	0.49	0.59	0.49
not empl.		8.98	17.66	23.12	24.10	25.27	28.49
unemployed		5.66	12.29	15.32	15.71	15.71	16.88
out of the LF		3.32	5.37	7.80	8.39	9.56	11.61
calendar inform.							
employed	100.00	90.05	85.85	82.34	81.37	80.20	77.46
unemployed	6.73	10.63	18.05	21.17	19.22	19.51	20.78
out of the LF	6.73	6.93	7.71	7.51	6.63	6.44	5.27

NOTES: in % of restricted sample (N=1284); labour market status as of time of survey for all categories with the exception of the calendar information which reports the percentage of individuals who were in the respective labour market state for at least one month up to the respective year and which is based on monthly calendar information; SOURCE: GSOEP, waves 7-14, subsample C

Table 4.34: Labour market status 1990-97 in case of treatment

	D=1						
LM status	1990	1991	1992	1993	1994	1995	1996
employed	100.00	96.91	96.14	88.03	85.33	88.03	83.78
full-time	88.80	89.58	90.73	82.63	80.69	79.15	73.36
part-time	11.20	6.95	5.02	4.63	4.63	7.72	8.49
(re)training		0.39	0.39	0.39	0.00	1.16	1.93
marginally		0.00	0.00	0.39	0.30	0.00	0.00
not empl.		3.09	3.86	11.97	14.67	11.97	16.22
unemployed		3.09	2.32	8.88	9.65	8.11	11.58
out of the LF		0.00	1.54	3.09	5.02	3.86	4.63
calendar inform.							
employed	100.00	99.23	97.68	91.89	91.51	91.12	89.58
unemployed	9.65	14.29	10.04	14.67	11.58	14.67	13.95
out of the LF	9.65	3.47	3.47	3.86	3.47	5.02	6.18

NOTES: in % of restricted sample (N=1284); labour market status as of time of survey for all categories with the exception of the calendar information which reports the percentage of individuals who were in the respective labour market state for at least one month up to the respective year and which is based on monthly calendar information; SOURCE: GSOEP, waves 7-14, subsample C

Table 4.35: Labour market transitions in case of non-treatment

	D=0					
LM transitions	90/91	91/92	92/93	93/94	94/95	95/96
employment to:						
employment	91.02	86.60	86.61	91.24	90.10	88.64
unemployment	5.66	9.97	9.83	7.23	7.71	9.53
out of the labour force	3.32	3.43	3.55	1.52	2.19	1.83
unemployment.to:						
employment		37.93	38.89	35.03	34.78	27.95
unemployment		50.00	49.21	54.14	54.66	53.42
out of the labour force		12.07	11.90	10.83	10.56	18.63
out of the LF to:						
employment		41.18	14.55	5.00	10.47	9.18
unemployment		11.76	21.82	23.75	15.12	14.29
out of the labour force		47.06	63.64	71.25	74.42	76.53

NOTES: one-year transition probabilities between labour market states (N=1284); labour market states as of time of survey; SOURCE: GSOEP, waves 7-14, subsample C

Table 4.36: Labour market transitions in case of treatment

	D=1					
LM transitions	90/91	91/92	92/93	93/94	94/95	95/96
employment to:						
employment	96.91	96.02	89.56	92.54	95.48	91.23
unemployment	3.09	2.39	7.63	7.02	4.07	6.58
out of the labour force	0.00	1.59	2.81	0.44	0.45	2.19
unemployment.to:						
employment		100.00	33.33	39.13	60.00	38.10
unemployment		0.00	66.67	39.13	32.00	57.14
out of the labour force		0.00	0.00	21.74	8.00	4.76
out of the LF to:						
employment			75.00	12.50	15.38	10.00
unemployment			0.00	0.00	30.77	30.00
out of the labour force			25.00	87.50	53.85	60.00

NOTES: one-year transition probabilities between labour market states (N=1284); labour market states as of time of survey; SOURCE: GSOEP, waves 7-14, subsample C

Table 4.37: OLS estimates of the returns to early (1990) job mobility following Mincer (1986): Overview

	(1)	(2)	(3)	(4)	estimated returns
control: movers in 1991-92					
$G(m) = b_t - b_{t+1}$	0.244	0.286	0.264	0.302	27.6 % - 35.3 %
coefficient on D	0.208	0.236	0.224	0.245	
	(0.036)	(0.035)	(0.035)	(0.036)	
coefficient on $\tilde{D}$	0.036	0.050	0.040	0.057	
	(0.027)	(0.026)	(0.026)	(0.026)	
R^2	0.391	0.451	0.461	0.463	
control: late movers in 1991-96					
$G(m) = b_t - b_{t+1}$	0.232	0.246	0.244	0.251	26.1 % - 28.5 %
coefficient on D	0.177	0.210	0.195	0.216	
	(0.037)	(0.036)	(0.036)	(0.038)	
coefficient on $\tilde{D}$	-0.055	-0.036	-0.049	-0.035	
	(0.023)	(0.023)	(0.022)	(0.023)	
R^2	0.393	0.451	0.462	0.463	

NOTES: dependent variable: log of the discounted sum of labour income 1990-96; D: early job-to-job mobility (1990); model specifications (1) to (4) as defined above;standard error estimates in parentheses; SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284)

Table 4.38: Correlation measures and treatment probabilities

	Corr(Z_i,D)	E[D|Z_i=0]	E[D|Z_i=1]	t-test
i=1	0.0929	5.86 (0.90)	10.98 (1.28)	-3.283 (0.001)
i=2	0.0945	7.10 (0.78)	14.08 (2.39)	-2.779 (0.006)
i=3	0.0880	7.60 (0.77)	17.24 (4.07)	-2.353 (0.022)
i=4	0.1198	4.49 (0.88)	11.14 (1.17)	-4.554 (0.000)

NOTES: D=early (1990) job-to-job mobility; Z_1 = firm-level announced layoffs as of 1990; Z_2 = occupation-specific labour market chances as of 1990; Z_3 = min(Z1,Z2); Z_4 = max(Z1,Z2); average treatment effects in %; standard errors in parentheses; t-values refer to two-sided test of equality of treatment probabilities for different assignment status allowing for unequal variance (p-values in parentheses); SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284)

Table 4.39: Estimates of the returns to early (1990) job mobility: Overview

	(1)	(2)	(3)	(4)	estimated returns
OLSE	0.200	0.225	0.215	0.232	18.1 % - 20.7 %
	(0.035)	(0.034)	(0.034)	(0.036)	
R^2	0.390	0.450	0.461	0.462	
IVE_1	-1.967	-1.967	-1.882	-1.794	-86.0 % - -83.4 %
	(0.918)	(1.061)	(0.764)	(1.000)	
1st stage F	7.93	5.76	10.48	5.71	
partial R^2	0.0064	0.0051	0.0091	0.0052	
IVE_2	0.872	0.685	0.862	0.536	70.9 % - 139.2%
	(0.553)	(0.476)	(0.509)	(0.466)	
1st stage F	6.03	6.73	6.45	6.41	
partial R^2	0.0070	0.0077	0.0074	0.0075	
IVE_3	0.307	0.118	0.141	0.212	12.5 % - 35.9 %
	(0.599)	(0.608)	(0.563)	(0.614)	
1st stage F	3.73	3.82	4.31	3.35	
partial R^2	0.0053	0.0055	0.0061	0.0050	
IVE_4	-1.099	-0.983	-1.080	-0.998	-66.7 % - -62.6 %
	(0.497)	(0.509)	(0.430)	(0.512)	
1st stage F	15.95	11.25	18.96	10.94	
partial R^2	0.0112	0.0098	0.0146	0.0100	
IVE_5	-0.491	-0.384	-0.643	-0.429	-47.4 % - -31.9 %
	(0.394)	(0.388)	(0.366)	(0.400)	
1st stage F	7.76	6.65	9.21	6.25	
partial R^2					

NOTES: dependent variable: log of the discounted sum of total income 1990-96; treatment variable: early job-to-job mobility (1990); model specifications (1) to (4) as defined above; robust standard error estimates in parentheses; SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284)

Table 4.40: OLS estimates of the returns to early (1990-92) job mobility following Mincer (1986): Overview

	(1)	(2)	(3)	(4)	estimated returns
control: late movers in 1993-96					
$G(m) = b_t - b_{t+1}$	0.188	0.188	0.188	0.198	20.7 % - 21.9 %
coefficient on D	0.059	0.080	0.068	0.087	
	(0.025)	(0.025)	(0.025)	(0.025)	
coefficient on $\tilde{D}$	-0.127	-0.108	-0.120	-0.111	
	(0.029)	(0.028)	(0.027)	(0.027)	
R^2	0.392	0.449	0.461	0.461	

NOTES: dependent variable: log of the discounted sum of labour income 1990-96; D: early job-to-job mobility (1990-92); model specifications (1) to (4) as defined above; standard error estimates in parentheses; SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284)

Table 4.41: Correlation measures and treatment probabilities

	Corr(Z_i,D)	E[D|Z_i=0]	E[D|Z_i=1]	t-test)
i=1	0.1304	20.79 (1.55)	32.38 (1.91)	-4.668 (0.000)
i=2	0.1203	23.81 (1.30)	38.03 (3.33)	-3.972 (0.000)
i=3	0.1074	24.90 (1.25)	43.68 (5.35)	-3.420 (0.001)
i=4	0.1672	17.77 (1.62)	32.60 (1.74)	-6.235 (0.000)

NOTES: D=early (1990-92) job-to-job mobility; Z_1 = firm-level announced layoffs as of 1990; Z_2 = occupation-specific labour market chances as of 1990; Z_3 = min(Z1,Z2); Z_4 = max(Z1,Z2); average treatment effects in %; standard errors in parentheses; t-values refer to two-sided test of equality of treatment probabilities for different assignment status allowing for unequal variance (p-values in parentheses); SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284)

Table 4.42: Estimates of the returns to early (1990-92) job mobility: Overview

	1	2	3	4	estimated returns
OLSE	0.089	0.107	0.097	0.114	9.3% - 12.1%
	(0.024)	(0.023)	(0.023)	(0.024)	
R^2	0.385	0.443	0.454	0.456	
IVE_1	-0.864	-0.878	-0.906	-0.883	-59.6% - -57.8%
	(0.324)	(0.443)	(0.302)	(0.456)	
1st stage F	16.83	12.99	19.97	10.87	
partial R^2	0.0131	0.0101	0.0155	0.0085	
IVE_2	0.468	0.361	0.447	0.304	35.5% - 59.7%
	(0.277)	(0.240)	(0.248)	(0.273)	
1st stage F	12.22	14.10	13.94	11.82	
partial R^2	0.0095	0.0110	0.0109	0.0093	
IVE_3	0.167	0.066	0.077	0.129	6.8 % - 18.2 %
	(0.323)	(0.336)	(0.309)	(0.372)	
1st stage F	7.28	7.32	8.17	5.34	
partial R^2	0.0071	0.0071	0.0080	0.0054	
IVE_4	-0.503	-0.451	-0.519	-0.499	-40.5 % - -36.3%
	(0.204)	(0.210)	(0.184)	(0.233)	
1st stage F	28.94	23.82	34.10	20.46	
partial R^2	0.0211	0.0183	0.0249	0.0159	
IVE_5	-0.296	-0.235	-0.341	-0.266	-28.9 % - -20.9 %
	(0.186)	(0.187)	(0.171)	(0.207)	
1st stage F	15.36	13.50	18.06	11.11	
partial R^2					

NOTES: dependent variable: log of the discounted sum of total income 1990-96; treatment variable: early job-to-job mobility (1990-92); model specifications (1) to (4) as defined above; robust standard error estimates in parentheses; SOURCE: GSOEP, waves 7-14, subsample C (East Germany, N=1284)

Chapter 5

Instrumental variables estimation of the returns to education in Germany: A variable treatment intensity approach[1]

5.1 Introduction

The presence of heterogeneity in returns to schooling is by now well established. Building on Gary Becker's (1967) model of optimal schooling according to which individuals choose their optimal schooling level by equating marginal benefits from continuing in education with the related marginal costs, recent theoretical contributions by *inter alia* Card (1995a,1995b) and Lang (1993) argue that individuals with different unobservable characteristics like ability, liquidity constraints or discount rates are likely to incur different marginal costs and benefits of further education and hence self-select into specific schooling levels. Such differences in the marginal costs and benefits of schooling imply different returns to schooling at different optimal schooling levels.

This in turn suggests the estimation of the returns to schooling on the basis of adequate

[1]joint work with Sascha O. Becker

instrumental variables and an interpretation of these estimates as *local average treatment effects* (LATE) along the lines of Angrist, Imbens and Rubin (1996): The estimated returns apply only to those individuals who are affected by the underlying instrument, i.e. those who only continue in school one more year because of their being induced to it by the instrument. The instrument is interpreted as an assignment to treatment and one year more spent in schooling as treatment. Moreover, different instruments will naturally affect different subgroups and hence lead to varying estimates of *the* returns to schooling.[2]

Several empirical studies on the returns to schooling in the US seem to corroborate the LATE interpretation of instrumental variables estimates.[3] Card (1995b) and Kling (2000) e.g. use an indicator of the presence of a college in the county of residence at schooling age as instrument. They argue that this "college proximity" might allow individuals from low income (and probably liquidity constrained) families to attend college who otherwise (i.e. if they would have had to move to another county in order to go to college) would not have done so.[4]

For Germany, Ichino and Winter-Ebmer (1999,2000)[5] are the only authors we know of that provide LATE estimates of the returns to schooling.[6] IWE (1999) contrast estimates obtained on the basis of two different instruments: first, parental educational background, and second, an indicator of the father's serving in the military during World War II. Since parental education as assignment mechanism is likely to affect less able children from well-off families (IWE (1999) call them the "stupid rich"), the corresponding IV estimate is interpreted as a lower bound of the returns to schooling in Germany. On the other hand, "childhood during war" and particularly "father in war" are considered an extreme form of liquidity constraints that might hinder highly talented children from poor families (the "smart poor") to continue schooling. For this reason, the authors interpret the IV estimate based on the "war instrument" as an upper bound of the

[2] For authorative overviews of the recent literature on the identification and estimation of causal effects in economics cf. Angrist and Krueger (1999) and Card (1999).

[3] Further empirical studies on the returns to schooling in a LATE framework are e.g. Angrist (1990), Angrist and Krueger (1991), Angrist and Imbens (1995), and Kane and Rouse (1993).

[4] Cf. overview of IVE results by Card (1999).

[5] IWE (1999) draws on IWE (2000) where in addition to the instrument 'father in war' an indicator of the individual's having been in the age group 9-15 during the Second World War is used as an instrument. The latter paper is more specifically concerned with the long-run educational cost of World War II, while the first paper is more methodological and aims at providing evidence for heterogeneity in the returns to schooling.

[6] Lauer and Steiner (2000), for Germany, and Pons and Gonzalo (2001), for Spain, do actually seem to follow a similar approach but they refrain from interpreting their estimates as local average treatment effects in a heterogenous treatment effect model.

returns to schooling in Germany.

In this paper, we extend the IWE (1999)-study in several ways: First, we replicate the Card- and Kling-studies for Germany, making use of an instrument similar to Card's "college proximity". Second, following the discussion in chapter 3 above, we allow for a variable treatment intensity and try to characterise both the affected subgroups as well as the response functions; and third, we compare results for 1985 with those for 1995, thus testing indirectly for changes in the returns to schooling, in the instrument effectiveness, and in the response functions over time.

The results obtained on the basis of GSOEP data suggest that, similar to the US results by Card and Kling, IV estimates of the returns to schooling are substantially higher than corresponding OLS estimates. We show that individuals from disadvantaged family backgrounds profit most from a better schooling infrastructure prevalent in urban areas.

The remainder of this chapter is organised as follows: section 5.2 presents a brief overview of Becker's (1964) well-known model of optimal schooling. Section 5.3 discusses some basic evidence on the relationship between educational attainment and college proximity using regional and GSOEP data. After a description of the underlying data and the sample definitions in section 5.4, section 5.5 presents the instruments used in the empirical analysis and summarises the empirical results. Section 5.6 concludes.

5.2 Theoretical considerations

In this section we shortly recall Becker's (1964) model of endogenous schooling in the version laid out by Card (1995b). It provides both the rationale for heterogeneous returns to schooling and the basis for the LATE interpretation of the final estimation results.

An individual maximises

$$U(y, S) = \log y - \phi(S) \tag{5.1}$$

where y is average earnings per year, S is years of schooling and $\phi(\bullet)$ is the cost of schooling.

An individual's opportunities are represented by $y = g(S)$.[7] The first order condition of the optimisation problem is

$$\frac{g'(S)}{g(S)} = \phi'(S) \tag{5.2}$$

Now, assume for simplicity that

$$\frac{g'(S)}{g(S)} = \beta_i(S) = b_i - k_1 S \qquad (k_1 \geq 0) \tag{5.3}$$

and

$$\phi'(S) = \delta_i(S) = r_i + k_2 S \qquad (k_2 \geq 0) \tag{5.4}$$

The optimal schooling level is then given by $S_i^* = (b_i - r_i)/k$, where $k = k_1 + k_2$. Integrating out (5.3) yields

$$\log y = b_i S - 0.5 k_1 S^2 \tag{5.5}$$

Equations (5.3) and (5.4) clearly state the reason for heterogeneous returns to schooling: Individuals are likely to differ in either marginal costs r_i or marginal benefits b_i and are therefore likely to choose different optimal schooling levels.

This is exactly what is exploited by the LATE-IV approach. A given instrument will affect different margins, i.e. different subpopulations at different schooling levels. As explained in detail in Angrist and Imbens (1995) we can hope to estimate only the average marginal return

[7] There is considerable discussion in the literature as to which variable best describes the theoretical concept of human capital. Griliches (1977) points out that years of schooling is rather one of the inputs of the human capital production process than its outcome. To the extent that output measures are unavailable, years of schooling as a proxy for human capital is the best variable we can get to describe what is valued in the labor market.

to schooling for a well-defined subgroup which is affected by the instrument. In the presence of heterogeneity, the notion of a unique return to schooling is hence nonsensical.

Estimation of the returns to schooling will be based on the following system of equations:

$$y = X\beta + S\gamma + \varepsilon \quad (5.6)$$

$$S = X\delta + Z\alpha + \eta \quad (5.7)$$

where Z is an instrument or set of instruments. For the LATE interpretation of IV to apply to the estimate of γ in (5.6), the conditions in Imbens and Angrist (1994) have to apply.[8] This approach thus makes a good out of the two main problems faced in a simple OLS regression of (5.6): the problem of self-selection into schooling and heterogeneity in returns to schooling. The main problem in empirical applications is, of course, to find an adequate instrument as an exogenous source of variation in education choices.

5.3 Educational outcomes and returns to schooling in Germany

In this section, we present descriptive evidence based on regional data for some recent years (1996-1998).[9] We collected data about school completion rates and school infrastructure as well as some information about the state of the labor market at the level of counties (*Kreise*). These data show, in particular, a huge variation in completion rates across counties as well as a positive correlation between completion rates and schooling infrastructure.

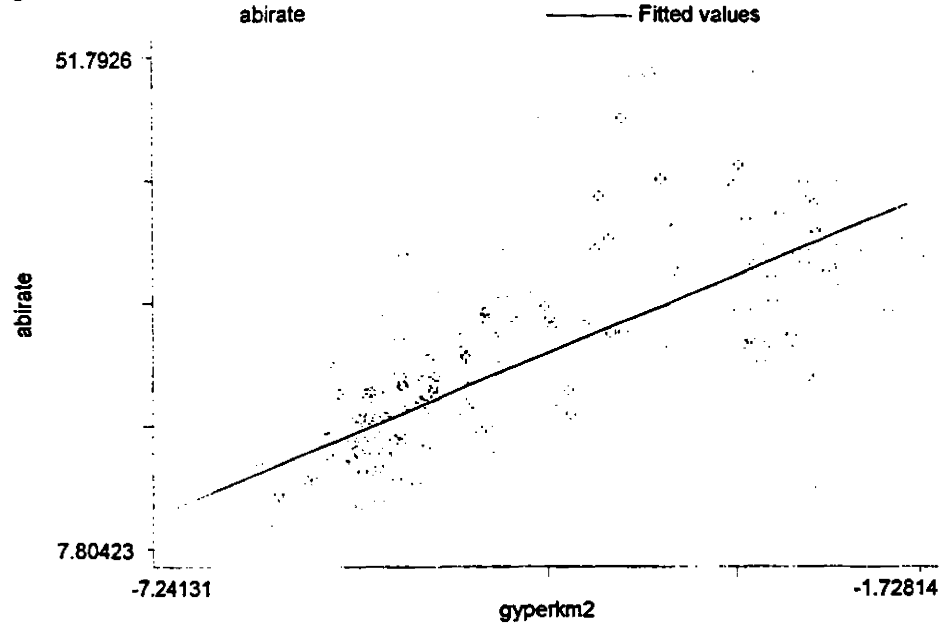
5.3.1 Some background information using regional data

High school completion rates (*Abitur*) in Germany range from roughly 8% (in *Südwestpfalz)* to 52% (in *Darmstadt*) of all school leavers across counties and hence show astonishingly strong

[8]Further assumptions implicit in equations (5.6) and (5.7) are log-linearity of earnings in schooling and the absence of degree effects (sheepskin effects). On the basis of the GSOEP data, the latter assumption cannot be tested since years of schooling are calculated from survey responses to questions about educational attainment levels. See Card (1999), however, for empirical evidence on the absence of sheepskin effects in the US.

[9]The data were obtained from the various regional statistical offices (Statistische Landesämter). To our knowledge, no consistent educational data base exists at the national level.

Figure 5-1: Educational attainment as a function of schooling infrastructure



regional variation. To see whether there is any systematic relationship between these high school completion rates, on the one hand, and the schooling infrastructure, on the other, we plotted the percentage of school leavers having *Abitur* against the log of the number of high schools per square kilometer as a measure of schooling infrastructure. As can be seen from figure 5-1, the availability of high schools is in fact seen to be highly correlated with high school completion rates.[10]

A higher average distance to the nearest high school is likely to increase the costs of education. Apart from the (time) opportunity costs of having to travel more, direct costs involve additional transport costs. All other costs do a priori not differ by distance to school. They might differ, however, across the various German regions (*Länder*) which are solely responsible for educational matters. Although there are generally no school fees neither for primary and secondary schools nor for universities, regulations regarding the public provision of books and other material used by students or subsidies for book purchases to low income families as well as regarding transportation subsidies for students do actually differ significantly across the various *Länder*. In many regions subsidies to either transport or book purchase are limited to students

[10]Of course, this is not necessarily a causal relationship driven by the supply of high schools. It could also be that lower demand for higher education causes less supply by the state.

Table 5.1: High school completion rates and average years of schooling by type of agglomeration

	completion rates		av. years of schooling	
agglomeration type	1985	1995	1985	1995
city	18.49	25.00	12.01	12.42
big town	15.76	20.12	12.00	12.37
small town	10.94	18.70	11.28	11.97
in the countryside	8.58	12.97	11.10	11.63

NOTES: type of agglomeration defined following classification in GSOEP; SOURCE: GSOEP (100% version), waves 2 (1985) and 12 (1995), subsamples A and B (West Germany; N=4617 in 1985 and N=3457 in 1995)

up to compulsory school age (i.e. 18 years old) or some other specific age (15 or 16 years old) and have to be borne fully by older students. Last but not least, the schooling years necessary for high school completion amount to 13 years in the West German *Länder* and Brandenburg as opposed to only 12 years in the remaining new German *Länder*. At university, the only fee to pay is for social security and health contributions.[11]

To sum up, using regional data we find lower high school completion rates in rural, less densely populated regions with a poorer schooling infrastructure. In addition, using microdata (GSOEP) we find lower high school completion rates for individuals who grew up in rural as opposed to urban areas. Average years of schooling by type of agglomeration show a similar pattern. (cf. table 5.1).[12]

Do these differences tell us something about regional variations in the quality of schools and/or high school degrees (as often suggested in the political debate) or are they indicative of regionally varying opportunity costs related to longer schooling? Our conjecture is that higher costs of education in regions with 'poor schooling infrastructure' reduce private investments in schooling, at least among children from relatively low-income/high discount rate families. This is also suggested by existing empirical studies on the returns to schooling based on instrumental

[11] In the later regressions, we try to capture differences in regulations across states by including a set of state dummies.

[12] Using regional data and defining agglomerations by quartiles of population density - which obviously do not coincide with the GSOEP classification - we observe a similar pattern. Going from the most densely to the least densely populated quartile, high schol completion rates in 1997 are 30.9%, 23.1%, 18.9%, and 19.4%, respectively.

variable estimation (Card (1995b), Kling (2000)). Card finds that the IV estimates of the earnings gain per year of additional schooling (10-14%) are substantially above the earnings gains estimated by a conventional OLS procedure (7.3%). Kling (2000), using Card's data, confirms Card's results and further characterises the group of students affected by differences in place of childhood.

5.3.2 Previous studies

Most previous results on returns to education for Germany are based on OLS regressions of earnings on schooling. Using data for the years 1984 and 1985 of the German Socioeconomic Panel (GSOEP), Wagner and Lorenz (1989) estimate returns to schooling of 6.5%. In a further study, Lorenz and Wagner (1993) give a range of 6.2-7.0% based on the Luxemburg Income Study (LIS 1981) and of 4.0-4.9% using data of the International Social Survey Program (ISSP 1987).

To our knowledge, the only studies applying IV estimation to infer returns to education in Germany are Ichino and Winter-Ebmer (1999,2000) and Lauer and Steiner (2000). The former authors exploit three different instruments: first, an indicator of father's education, second, an indicator of whether an individual was 10 years old during World War II, and third, an indicator of whether their father was in war during this period. Based on GSOEP data for 1986, the authors conclude that there is a lower bound of 4.8% and an upper bound of 14% to the returns to schooling for those subpopulations that are affected by the respective instruments. Lauer and Steiner (2000), on the other hand, not only estimate the returns to schooling using various estimation methods but also employ IV estimators on the basis of a long list of various instruments. They are mainly interested in an analysis of the robustness of the estimated returns to schooling with respect to the various estimation methods and do not provide any LATE interpretation of the obtained IV estimation results. Moreover, the authors conclude that there is no statistical evidence for heterogeneous returns to schooling with respect to unobservable characteristics.

5.4 Database and sample selection

As in the previous chapter, the empirical analysis is based on data from the richest German micro data set, the German Socioeconomic Panel. The sample is chosen to comprise all full-time employed in 1985 or 1995 who have no missing information on the variables of interest, in particular data on current earnings at the time of the survey and information on the educational background. Years of schooling is chosen as mulivalued treatment variable. Further information on - schooling or vocational - degrees obtained is also presented. Exogenous variables to be included into the analysis are gender, age, experience and tenure, as well as information on changes in the place of residence since childhood. Tables 5.8 and 5.9 in the appendix, contain the descriptive statistics for these variables.

5.5 Instrumental variables estimation of the returns to education

5.5.1 Choice of instrument

Previous studies have used a broad range of instruments to establish causality in the returns to schooling (see Card (1999) and the references therein). The choice of an instrument has several important aspects. First, econometrically speaking the instrument should fulfill the exclusion restriction, i.e. have an effect on earnings only via the schooling channel but no direct effect on earnings. Second, heterogeneity in marginal costs and benefits of schooling and therefore the absence of a unique return to schooling for the population as a whole can be exploited by choosing an instrument which describes a quasi-experiment of important policy interest. So, IV estimation is not just a solution to the econometric problem of endogeneity bias but also allows to analyze interesting policy questions. On the basis of these two considerations, we choose our instrument 'place of childhood' which is similar to Card's (1995b) college proximity indicator. It has not yet been used in studies for Germany and allows us to address the question as to *who* profits *how* from differences in schooling infrastructure across different places of childhood.

On place of childhood, the GSOEP questionnaire contains the question: *"Did you spend the major portion of your childhood up to age 15 in a) a city, b) a big town, c) a small town, or d)*

Table 5.2: Percentage of sample with given instrument status

	individual grew up in ...	1985	1995
pc1	... a city	21.90	19.09
pc2	... a city or a big town	36.28	33.27
pc3	... some urban area	58.74	54.12

SOURCE: GSOEP (100% version), waves 2 (1985) and 12 (1995), subsamples A and B (West Germany; N=4617 in 1985 and N=3457 in 1995)

in the countryside ?" In the sequel, we are going to use three different binary indicators based on this question: 'spent childhood in a city'(pc1), 'spent childhood in a city or big town' (pc2), and 'spent childhood in an urban area' (pc3), i.e. in a city, or in a small or big town. Table 5.2 shows the percentage of the sample with given instrument status.

5.5.2 Variable treatment intensity approach to the estimation of returns to schooling

The natural experiment implied by the above instrument is described by place of childhood as assignment to treatment (Z), the schooling level as treatment (S), and log(monthly earnings) as outcome (Y). In this framework, individual-level potential schooling levels and potential outcomes can be defined - at least conceptually - for all values of the instrument (e.g. "grown up in the countryside", "grown up in a small town" or "grown up in a big city").

Given that schooling is a multivalued treatment variable, it follows from theorem 13 in chapter 3 that the IV estimate of the returns to schooling based on 'place of childhood' as an instrument identifies a causal effect for well-defined subpopulations and schooling levels if the conditions "independence" and "strong monotonicity" as stated in that theorem are met. In this case, IV generates an estimate of the average causal effect among individuals with different marginal benefits from schooling. This estimate is equal to a weighted average causal response, allowing for several particularities: first, different subgroups are affected by different instruments; second, individuals in these subgroups are affected by the respective instrument in different ways; and third, the instrument may induce changes of behavior at different levels of schooling.

In the empirical part, we present both the weighting function and the response function for the given choice of instrument and thereby try to characterise the affected subgroups and schooling levels. The plausibility of the assumptions underlying theorem 13 will be discussed below.

5.5.3 IV Estimation results

We started by estimating an OLS regression of earnings on years of schooling controlling for sex, experience and tenure on the job polynomials, yielding estimates in the usual range of 6.5% to 6.7% for both years, 1985 and 1995.

For the reasons given above, these estimates are probably not amenable to an interpretation as the causal effect of schooling on earnings. We therefore performed an IV estimation of the returns to education on the basis of the instruments suggested above. The instrumental variables estimates of the returns to schooling on the basis of the chosen instrument have been computed using the two-stage least squares procedure: in the first stage, the years of schooling are regressed on the whole list of exogenous variables augmented by the respective instrumental variable using a simple linear probability model; in the second stage, the predicted value of the dependent variable from the first stage regression is then used as additional regressor in the outcome equation instead of the schooling years itself. Table 5.3 contains the IV estimation results for the various chosen instrumental variables. Further, first-stage t-statistics and partial R^2 measures are reported as a diagnostic tool following the suggestions of Bound et al. (1995) and Staiger and Stock (1997). In all cases, the instrument quality seems reasonable as suggested by these measures.

The returns estimated using either of these instruments are considerably higher than the OLS estimates. In 1985, the point estimates are 12.6%, 12.5% and 13.3% for the binary instruments 'spent childhood in a city', 'spent childhood in a city or big town', and 'spent childhood in an urban area', respectively. A similar picture arises for the 1995 data. Throughout, the IV estimates are nearly double the size of the OLS estimates. In the light of the LATE framework, these results can be interpreted as the returns to education for those who acquired more education *because* they are living in an area with a good schooling infrastructure.

Table 5.3: Estimates of the returns to education: Overview

	1985	1995
OLSE	6.71	6.54
	(6.31;7.11)	(6.12;6.96)
IVE (pc1)	12.63	12.58
	(7.89;17.39)	(8.45;16.70)
1st stage t	6.737	7.087
partial R^2	0.0098	0.0141
IVE (pc2)	12.49	9.67
	(9.00;15.98)	(6.91;12.45)
1st stage t	9.247	9.691
partial R^2	0.0183	0.0265
IVE (pc3)	13.28	9.22
	(7.94;18.63)	(6.95;11.48)
1st stage t	6.131	11.387
partial R^2	0.0081	0.0362
IVE (pc1*(poor fbq))	10.65	11.17
	(6.76;14.55)	(7.57;14.77)
1st stage t	-7.848	-7.361
partial R^2	0.0075	0.0105
IVE (pc2*(poor fbq))	9.86	11.29
	(7.44;12.28)	(8.34;14.25)
1st stage t	-11.721	-9.512
partial R^2	0.0142	0.0176
IVE (pc3*(poor fbq))	9.33	9.68
	(7.58;11.08)	(7.77;11.60)
1st stage t	-15.795	-13.845
partial R^2	0.0219	0.0273

NOTES: The estimates denoted IVE (pc_i), i=1,2,3, denote the IV estimates when using place of childhood as instrument (pc1=city; pc2=city or big town; pc3=urban); The estimates denoted IVE (pc_i*(poor fbq)), i=1,2,3, denote the IV estimates when using place of childhood interacted with poor family background as instrument; confidence intervals in parentheses SOURCE: GSOEP (100% version), waves 2 (1985) and 12 (1995), subsamples A and B (West Germany; N=4617 in 1985 and N=3457 in 1995)

5.5.4 Internal validity of the instruments

To check the internal validity of the instrument for identification of the LATE parameter, the assumptions underlying theorem 12 in chapter 3 have to be verified:

We can be quite confident that the *SUTVA assumption* is satisfied in our sample. It requires potential earnings to be unrelated to the amount of schooling taken by *other* individuals in the sample. This assumption is more likely to be violated in clustered samples.

Strongly ignorable assignment to treatment requires that after controlling for observable characteristics, unobservables like ability should be randomly distributed across different places of childhood. This assumption could be violated if parents endogenously choose to live in an urban area because of better schooling infrastructure. Most of this potential selection into places of living is probably controlled for by observables. In any case, geographical mobility in Germany is quite low by international standards. While Germany has 16 states and about 80 million inhabitants, the US have 51 states and about 250 million inhabitants, so average population per state is relatively similar, the US states being bigger in size, however. While in the US, 3% of the population move across state borders every year, in Germany only 1% of the population move across state borders.[13] Not only are mobility rates low anyway, but the reasons for moving are very unlikely to be related to schooling infrastructure as well. The GSOEP data contain a question on reasons for move. In 1997, respondents could give a maximum of three (out of a list of 15) possible reasons. Overall, 8.6% of the movers give "other family reasons" (i.e. family reasons other than divorce, marriage and leaving parent's home) as the reason for moving. If at all, families that move to give their children access to a better schooling infrastructure might show up in this group. For families with children under age 18 (i.e. those families for whom schooling infrastructure might play a roll), the percentage moving for "other family reasons" is even lower (5.2%), thus making "better schooling infrastructure" an even more unlikely reason for moving. We conclude that our estimates are very unlikely to suffer from violation of the *strongly ignorable assignment to treatment* assumption.

Strong monotonicity compares again two counterfactual situations: an individual growing

[13]Data come from the websites of the US Census Bureau and the German National Statistical Office, respectively:
http://www.census.gov/population/socdemo/migration/tab-a-1.txt
http://www.statistik-bund.de/jahrbuch/jahrtab5.htm

up in a city (i.e. in region with good schooling infrastructure) takes at least as much schooling as if he had grown up in the countryside (i.e. in a region with a worse infrastructure). This assumption rules out *defiers*, i.e. individuals who, if growing up in a city, take less schooling than if growing up in the countryside. In theory, there might be individuals who take less schooling growing up in an urban area due to various "outside attractions" such as drugs and delinquency, but would have obtained more schooling if growing up in a rural area. In a similar way, labor demand in cities might be higher and therefore students might have more outside options in a city as compared to an urban area and for some individuals these outside options might lead to a lower schooling level. While we cannot really rule out that there are some cases like this, for the reliability and interpretability of our estimates it is important that the fraction of defiers is nevertheless very small. One testable implication of strong monotonicity is that the cumulative density functions of schooling by instrument status do not cross. As we will show, this holds in our data and makes us confident that violation of the strong monotonicity assumption is not a serious issue here.

The *exclusion restriction*, finally, would be violated if there existed a direct effect of the suggested instrument on earnings, e.g. in the form of an 'urban wage premium'. We are in the fortunate situation to have some information about the current place of living. The GSOEP data contain both current state (*Bundesland*) of residence as well as the so-called Boustedt regions.[14] We find that by including these as additional controls, in 1985 the estimated returns to schooling do not change and in 1995 they even go slightly up. When controlling for state dummies, the coefficients on the Boustedt dummies are found to be statistically insignificant. We might therefore conclude that there is no violation of the exclusion restriction through an urban wage premium.

Another reason why the exclusion restriction might be violated is that school quality might vary by place of childhood. In this case, controlling for characteristics of the current place of living is not sufficient because people might have moved and the decision to take further schooling depended on their place of childhood and not on their current place of living. To see if this is a valid objection, we follow an idea similar to Card (1995b) and Kling (2000). They

[14] Boustedt (1970) classifies urban regions into seven categories, assigns the neighbouring communities of an urban center to four different sub-categories from "rural" to "urban center".

propose to define family background quartiles across which the returns to schooling will vary. In order to test whether college proximity is a legitimate instrument, they use the interaction of college proximity with an indicator for low parental background as an instrument and control for the main effect of college proximity. Translated to our setup, the idea is that our instrument is unlikely to affect individuals from higher family background quartiles because they have the necessary support by their family to pursue further education even if the respective schools are not nearby. So, using the instrument as such or using the instrument interacted with an indicator of low family background is the same, and gives us one more degree of freedom by allowing us to control for the main effect of the 'place of childhood' indicator. We will further discuss the construction of the family background quartiles in the following section. There, we also use them to characterise the subgroup of compliers, so they serve a double purpose.

Let us shortly summarise the results of the estimation using the interacted instruments. We find that indeed the main effect of 'growing up in an urban area' is small in size and statistically insignificant.[15] The lower panel of table 5.3 shows that the point estimates are lower than the ones where we do not control for the main effect of 'growing up in an urban area', but that they are still considerably higher than the OLS estimates. On the basis of this evidence in favor of both the absence of urban wage premia and the validity of the exclusion restriction, we conclude that the returns to education for the subgroups of compliers, i.e. those individuals who only acquire more schooling when enjoying a good schooling infrastructure, are significantly and substantially higher than the simple OLS estimates. In the following section, we turn to the characterisation of the subgroups affected by our instrument.

5.5.5 External validity of the instruments

If we want to generalise our estimates to some larger populations ("external to the sample"), we have to characterise as closely as possible the subgroups affected by our instrument and the size of the effect on them. We suggested above that the effect of schooling infrastructure is more important for children from less advantaged family backgrounds. We follow Card (1995b) and Kling (2000) in defining family background quartiles in the following way: First, we perform a

[15]The coefficients on the main effect pc1 is 0.012 with a standard error of 0.019 in 1985, and 0.011 with a s.e. of 0.020 in 1995.

regression of years of schooling on the subgroup of people who spent their childhood in a rural area. Then, based on the parameter estimates obtained, we predict - for all individuals - their 'counterfactual schooling level if they had grown up in a rural area' and split the sample into four quartiles, from the lowest (fbq1) to highest (fbq4).

Table 5.4 presents some summary statistics on average years of schooling by instrument status and family background quartile for the years 1985 and 1995. Apart from the fact that average years of schooling are higher for those who grew up in urban areas, the table clearly shows that for those who have a higher predicted (counterfactual) schooling level, also actual schooling attainment is higher.

Table 5.5 further shows the distribution of family background and individual variables across these 'counterfactual schooling quartiles'.[16] There is no single individual in the lowest three family background quartiles whose father has a university degree. Conversely, there is virtually no individual in the two highest background quartiles who has a father without a schooling degree. We also see that a higher percentage of those in the upper family background quartiles did actually grow up in a city.

The IV estimate of the returns to schooling can be interpreted as a weighted average of the potentially differing treatment effects across the four background quartiles, γ_q, with the weight given to each quartile q by the product of the proportion of the population in that subgroup (w_q) and the impact on schooling for that subgroup (ΔS_q). This allows us to write

$$\gamma = \sum_{q=1}^{4} \frac{w_q \Delta S_q \gamma_q}{\Delta S}$$

The observed weights w_q by family background are given in table 5.6.

Table 5.7 shows the differences in schooling levels by instrument status for the population as a whole (ΔS).

Figures 5-2 and 5-3 further split up the information of table 5.6 by family background quartiles for 1985 and 1995 respectively. In 1985, the actual average education difference by

[16]It is interesting to note that in the lowest background quartile, none of individuals report that either their father or mother graduated from high school.

Table 5.4: Average years of schooling by instrument status and family background quartile

	fbq1	fbq2	fbq3	fbq4
1985				
City	10.79	11.45	12.23	13.36
City or big town	10.66	11.36	12.19	13.54
Urban area	10.44	11.05	12.11	13.28
1995				
City	11.46	11.64	12.48	14.13
City or big town	11.59	11.57	12.49	13.97
Urban area	11.18	11.44	12.49	13.88

NOTES: fbq_j, j=1,2,3,4, denotes the family background; definition of quartiles based on regression of schooling level on family background variables (and age) for individuals from rural background and subsequent predictions for all observations as 'counterfactual schooling level if individual had grown up in a rural area';
SOURCE: GSOEP (100% version), waves 2 (1985) and 12 (1995), subsamples A and B (West Germany; N=4617 in 1985 and N=3457 in 1995)

instrumental status is much larger for the two lower background quartiles, supporting the suggestion that instead of our indicator for 'growing up in an urban area' we can equally well use this indicator interacted with poor family background. This allows us to use the main effect of 'growing up in an urban area' in the estimation and thereby control for there being an urban wage premium. We already reported the results of this exercise in the previous subsection.

5.5.6 Characterizing the response function

The response function can be estimated from the cumulative distribution functions (CDF) of schooling at different values of the instrument. The difference in the CDFs is equivalent to the fraction of the population who received at least one more year of schooling due to the instrument. Figure 5-4 shows the difference in the CDFs for the 1985 sample using pc1 as an instrument.[17] It indicates that schooling infrastructure has its largest effect at 11 years of schooling. More specifically we interpret the estimates to indicate that around 10 percent of individuals with similar demographics are induced to obtain more years of schooling due to

[17] Figures based on the instruments pc2 and pc3 show a similar pattern and are therefore not shown here.

Table 5.5: Distribution of family background and individual variables across 'counterfactual schooling quartiles'

variable	fbq_1	fbq_2	fbq_3	fbq_4	mean
father's education					
high school degree	0.00	0.26	2.43	20.85	5.76
vocational degree	0.25	2.35	4.59	23.34	7.59
university degree	0.00	0.00	0.00	16.68	4.07
no schooling degree	35.30	23.67	0.09	0.09	15.01
Mother's education					
high school degree	0.00	0.00	0.35	6.83	1.75
vocational degree	0.25	1.65	3.47	12.87	4.48
university degree	0.00	0.09	0.09	3.11	0.80
no schooling degree	42.29	30.72	0.87	0.80	18.93
parental presence	0.00	30.64	96.01	95.21	54.86
place of childhood					
city	19.63	20.19	22.36	25.55	21.90
city or big town	33.87	33.86	35.44	42.15	36.28
urban	56.11	59.18	56.93	62.91	58.74
change of place	55.46	61.74	62.02	58.86	60.14
female	48.19	25.76	30.59	11.36	29.22
mean age	40.59	35.90	38.04	34.00	37.18

NOTES: frequency of respective characteristic by family background quartile; definition of quartiles based on regression of schooling level on family background variables (and age) for individuals from rural background and subsequent predictions for all observations as 'counterfactual schooling level if individual had grown up in a rural area'; SOURCE: GSOEP (100% version), waves 2 (1985) and 12 (1995), subsamples A and B (West Germany; N=4617 in 1985 and N=3457 in 1995)

Table 5.6: Covariate weights

	fbq_1	fbq_2	fbq_3	fbq_4
1985				
city	23.05	22.95	25.52	28.49
city or big town	24.00	23.22	24.42	28.36
urban area	24.56	25.07	24.23	26.14
1995				
city	16.21	26.06	27.88	29.85
city or big town	18.00	25.91	26.61	29.48
urban area	19.99	25.28	26.19	28.54

NOTES: w_q is the fraction in each family background quartile; definition of quartiles based on regression of schooling level on family background variables (and age) for individuals from rural background and subsequent predictions for all observations as 'counterfactual schooling level if individual had grown up in a rural area';
SOURCE: GSOEP (100% version), waves 2 (1985) and 12 (1995), subsamples A and B (West Germany; N=4617 in 1985 and N=3457 in 1995)

Table 5.7: Differences in schooling by instrument status

	1985		1995	
$Z_i = pc_i = ...$	0	1	0	1
city	11.32	12.04	11.85	12.59
city or big town	11.17	12.01	11.73	12.53
urban area	11.10	11.74	11.56	12.36

SOURCE: GSOEP (100% version), waves 2 (1985) and 12 (1995), subsamples A and B (West Germany; N=4617 in 1985 and N=3457 in 1995)

Figure 5-2: Actual average education difference by instrumental status (pcl) 1985

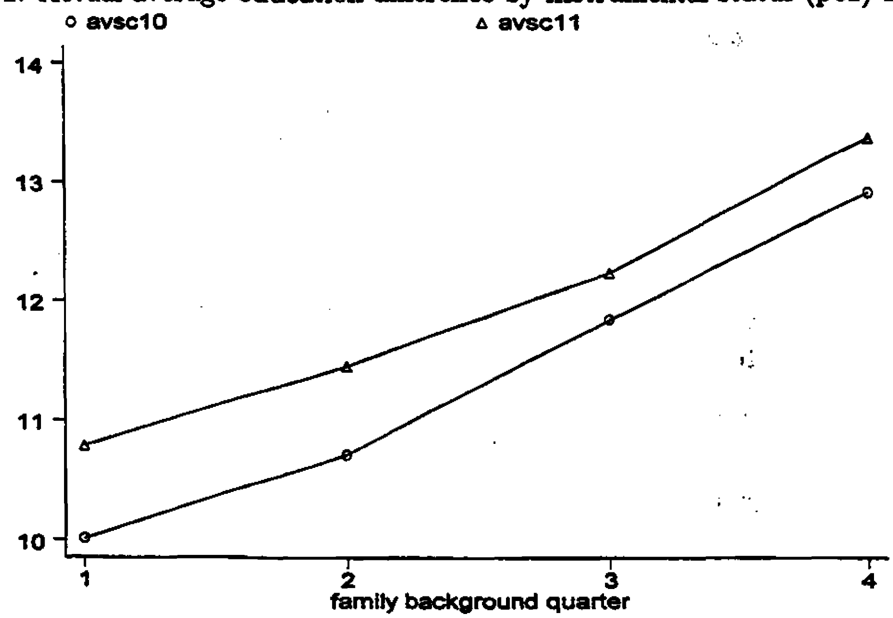


Figure 5-3: Actual average education difference by instrumental status (pcl) 1995

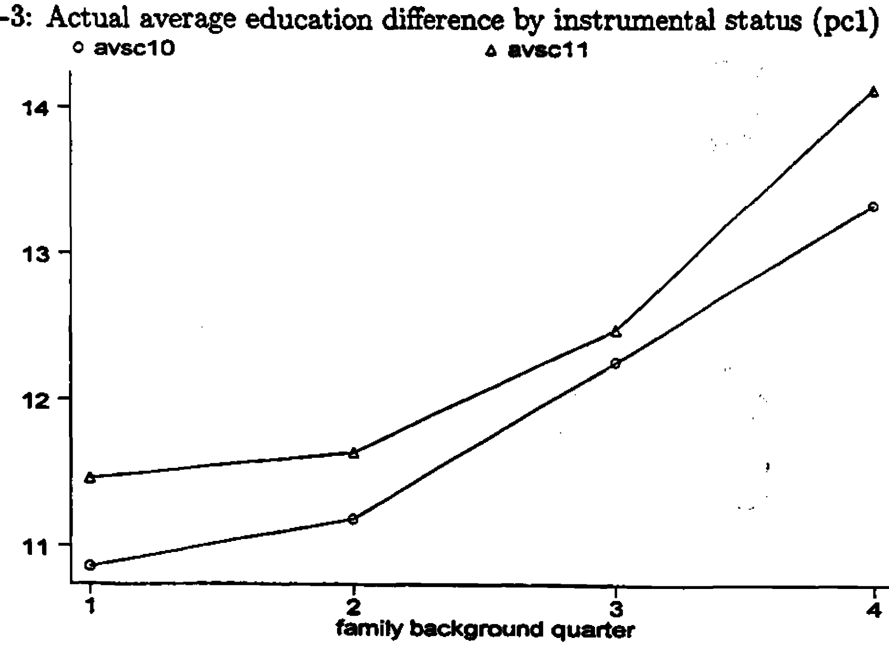
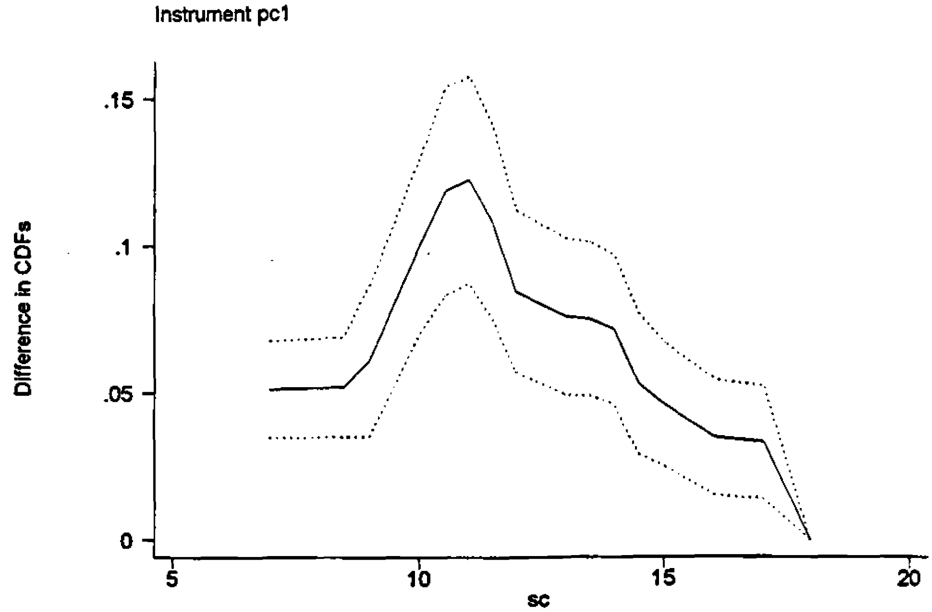


Figure 5-4: CDF difference 1985 using pc1 as instrument



better schooling infrastructure.

It is even more interesting to break down the response function by background quartiles. Figure 5-5 shows that the response function of the two lower background quartiles peaks at 10 years of schooling while the response of the two upper quartiles is concentrated among those with 13 or more years of schooling. Furthermore, the fraction of 'compliers' in the two upper quartiles is overall much lower, again showing that the instrument affects mainly the two lower family background quartiles.

From a policy point of view, this result suggests that the provision of schools beyond 10th grade, i.e. basically the provision of (senior) high schools (*Gymnasien*), can considerably increase the fraction of youths from disadvantaged backgrounds who obtain more schooling.

For 1995, the picture is slightly different. First, figure 5-6 suggests that for this later cohort, schooling infrastructure increased educational attainment at a later stage in educational careers.

Overall, in 1995 the response function is flatter and takes on lower values than in 1985. Second, breaking down by background quartiles, we find that the point of maximum response has moved to the right for all subgroups. Also has the fraction of the population in all subgroups who respond to our instrument decreased (cf. figure 5-7)

Figure 5-5: CDF difference 1985 by family background quartile using pc1 as instrument

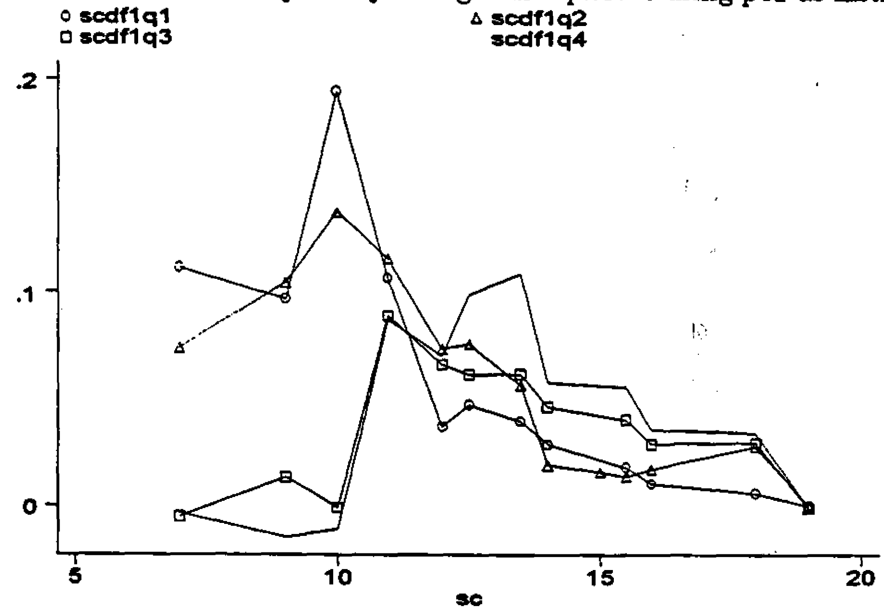


Figure 5-6: CDF difference 1995 using pc1 as instrument

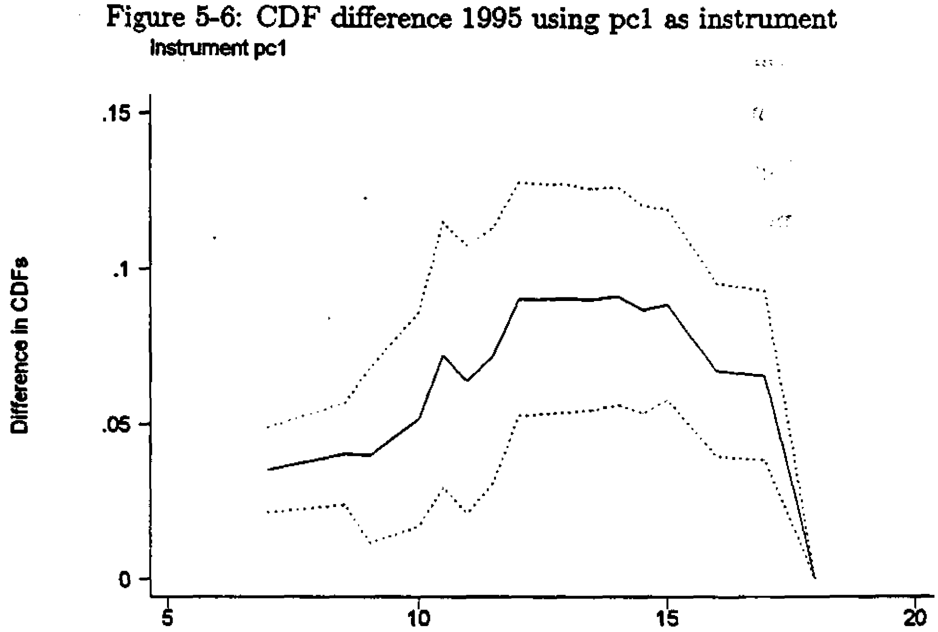
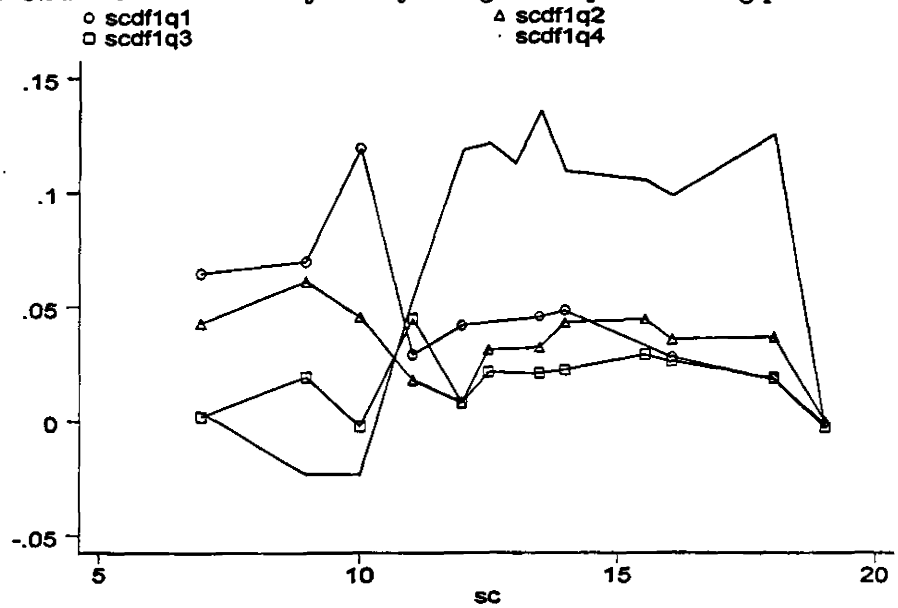


Figure 5-7: CDF difference 1995 by family background quartile using pc1 as instrument



The fact that figures 5-4 and 5-6 display only non-negative values is equivalent to saying that the CDFs for $Z = 1$ and $Z = 0$ don't cross, a finding that supports the strong monotonicity assumption laid out in theorem 13 of chapter 3.

To sum up, there seems to be a decreasing effect of our instrument on lower schooling levels and/or an increasing effect of the instrument on higher schooling levels. This also explains why returns to education seem to have decreased between 1985 and 1995.

5.6 Summary and conclusions

This study corroborates the general finding of other studies based on IV estimation that OLS estimates are downward biased. It confirms the empirical evidence that different instruments lead to different estimates of the schooling coefficient, underlining the fact that returns to schooling are heterogeneous. Our estimates remain within the bounds given by IWE (1999). We find that individuals from 'poor family background' respond most strongly to the instrument 'place of childhood'. Their response is further most pronounced at low schooling levels whereas the response of individuals with 'rich family background' is most pronounced at higher schooling levels. Finally, this approach allows us to detect changes in the response function over time.

The temporal variation of returns to schooling operates through two different channels. First, temporal variation in the covariate weights leads to a reweighting of the returns for different subgroups. We conjecture that there is a decreasing fraction of compliers from a poor family background and/or an increasing fraction of compliers from a rich family background. Second, temporal variation of returns to schooling is also due to temporal variation in the response functions. There seems to be a decreasing effect of our instrument on lower schooling levels and/or an increasing effect of instrument on higher schooling levels.

The finding that educational attainment crucially depends on the provision of post-compulsory schooling in proximity to the place of living, has important policy implications. Consider the case of a regional government that has decided to devote a certain amount of money to the improvement of upper secondary schooling infrastructure.[18] It then faces the decision *where* to build the school, in an urban area or in a rural area, or similarly whether to build one big school in a city or some smaller schools in the countryside. If the per student cost of providing further places at school is constant independent of where schools are built, our results clearly indicate that students living in areas with a less favourable schooling infrastructure would probably benefit most from such an investment because of their above average marginal returns to education. To the extent that schooling infrastructure is correlated with the degree of urbanisation, providing a better schooling infrastructure especially in rural areas could thus considerably increase the incentives for individuals from disadvantaged family background to acquire more education and thus improve their long-run prospects in the labor market.

It is important to note, though, that the policy implications might be quite different for the case in which the federal government increases schooling infrastructure in the country as a whole. In this case there might be general equilibrium effects that decrease the return to education in the long run due to an overall higher supply of better-educated individuals (see Heckman et al., 1999). The policy implications of this paper do therefore refer to the optimal allocation of schools but not necessarily to the optimal overall spending on schooling infrastructure.

[18] We do not address the cost-benefit issue here, i.e. we do not ask whether for the region as a whole investing in schooling infrastructure is beneficial. In contrast, we take an individual-level perspective and take the provision of funds by the government as given in this thought experiment.

5.7 References

Angrist, J.D. (1990), Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records, American Economic Review, 80(3), 313-36

Angrist, J.D. and G.W. Imbens (1995), Two-Stage Least Squares Estimation of Average Causal Effects in Models With Variable Treatment Intensity, Journal of the American Statistical Association, 90(430), 431-442

Angrist, J.D. and A.B. Krueger (1991), Does Compulsory School Attendance Affect Schooling and Earnings, Quarterly Journal of Economics 106, 979-1014

Angrist, J.D. and A.B. Krueger (1999), Empirical strategies in labor economics, in: O. Ashenfelter and D. Card (eds.), Handbook of Labor Economics, Vol. 3, chapter 23, North-Holland

Angrist, J.D., G.W. Imbens and D.B. Rubin (1996), Identification of Causal Effects Using Instrumental Variables, Journal of the American Statistical Association, 91(434), 444-455

Becker, G.S. (1964), Human Capital, 1st ed. New York: Columbia University Press

Becker, G.S. (1967), Human Capital and the Personal Distribution of Income, Ann Arbor, University of Michigan Press

Bound, J., D.A. Jaeger and R.M. Baker (1995), Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak, Journal of the American Statistical Association, 90(430), 443-450

Boustedt, Olaf (1970): Pendlerverkehr, in: Handwörterbuch der Raumforschung und Raumordnung, 2. Aufl., Bd. II, Sp. 2284. Hannover

Card, D. (1995a), Earnings, Schooling and Ability Revisited, in: S. Polachek (ed.), Research in Labor Economics, 14 , 23-48

Card, D. (1995b), Using Geographic Variation in College Proximity to Estimate the Return to Schooling, in L.N. Christofides, E.K. Grant, and R. Swidinsky (eds), Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp, University of Toronto Press, Toronto, 201-222

Card, D..(1999), The Causal Effect of Education on Earnings, in: O. Ashenfelter and D. Card (eds), Handbook of Labor Economics, Vol. 3A, 1801-63

Griliches, Z. (1977), Estimating the returns to schooling: Some econometric problems, Econometrica, 45(1), 1-22

Heckman, J.J., L. Lochner, and C. Taber, Human Capital Formation and General Equilibrium Treatment Effects: A Study of Tax and Tuition Policy, Fiscal Studies, 20(1), 25-40

Ichino, A. and R. Winter-Ebmer.(1999), Lower and Upper Bounds of Returns to Schooling: An Exercise in IV Estimation with Different Instruments, European Economic Review, 43(4-6), 889-901

Ichino, A. and R. Winter-Ebmer (2000), The Long-Run Educational Cost of World War Two, Euroean University Institute, mimeo

Imbens, G.W. and J.D. Angrist (1994),.Identification and Estimation of Local Average Treatment Effects, Econometrica, 62(2), 467-475

Kane, T.J. and C.E. Rouse (1993),.Labor Market Returns to Two- and Four-Year Colleges: Is a Credit a Credit and Do Degrees Matter?, NBER Working Paper No. W4268

Kling, J. (2001), Interpreting Instrumental Variables Estimates of the Returns to Schooling, Journal of Business and Economic Statistics (2001), forthcoming

Lang, K. (1993), Ability Bias, Discount Rate Bias and the Returns to Education, Boston University, mimeo

Lauer, C.and V. Steiner (2000),.Returns to Education in West Germany, An Empirical Assessment, Center for European Economic Research (ZEW), Mannheim, mimeo

Lorenz, W., and J. Wagner (1993), A Note on Returns to Human Capital in the Eighties: Evidence from Twelve Countries, Jahrbücher für Nationalökonomik und Statistik, 211(1-2): 60-71

Pons, E. and M.T. Gonzalo (2001), Returns to schooling in Spain: How reliable are IV estimates?, Working Paper No. 446, Queen Mary University of London, Department of Economics

Staiger, D. and J.H. Stock (1997), Instrumental variables regression with weak instruments, Econometrica, 65, 557-586

Wagner, J., and W. Lorenz (1989), Einkommensfunktionsschätzungen mit Längsschnittdaten für vollzeiterwerbstätige deutsche Männer (Estimates of Earnings Functions for Full-Time Working German Men Using Longitudinal Data), Konjunkturpolitik, 35(1-2), 99-109

5.8 Appendix

Table 5.8: Summary statistics on education and outcome variables

variable	mean	st. dev	min	max
1985				
years of schooling	11.47	2.78	7	19.5
Hauptschule	0.43	0.50	0	1
Realschule	0.34	0.48	0	1
Fachhochschulreife	0.03	0.18	0	1
Abitur	0.09	0.29	0	1
Apprenticeship	0.64	0.48	0	1
University degree	0.10	0.30	0	1
gross monthly income	2983.33	1382.73	0	19000
net monthly income	2029.74	959.23	0	13000
1995				
years of schooling	12.00	2.87	7	19.5
Hauptschule	0.40	0.49	0	1
Realschule	0.34	0.47	0	1
Fachhochschulreife	0.05	0.23	0	1
Abitur	0.13	0.34	0	1
Apprenticeship	0.69	0.46	0	1
University degree	0.14	0.34	0	1
gross monthly income	4524.49	3038.21	0	99999
net monthly income	2984.71	1845.25	0	50000

SOURCE: GSOEP (100% version), waves 2 (1985) and 12 (1995), subsamples A and B (West Germany; N=4617 in 1985 and N=3457 in 1995)

Table 5.9: Summary statistics on exogenous variables

Variable	mean	dt. dev	min	max	N
1985					
sex	0.29	0.45	0	1	4617
age	37.18	10.06	20	55	4617
experience	20.70	10.54	0	43	4617
tenure	9.74	8.06	0	56.6	4606
changed place since childhood	0.60	0.49	0	1	3181
1995					
sex	0.31	0.46	0	1	3457
age	37.00	9.70	20	55	3457
experience	20.00	9.99	1	43	3457
tenure	9.74	8.77	0	41.3	3457
changed place since childhood	0.64	0.48	0	1	2274

SOURCE: GSOEP (100% version), waves 2 (1985) and 12 (1995), subsamples A and B (West Germany; N=4617 in 1985 and N=3457 in 1995)

Chapter 6

Summary and concluding remarks

As discussed in detail in the introductory **chapter 1**, the assessment of the individual-level costs and benefits of job mobility behaviour throughout the transition from a centrally planned to a market economy in terms of both income streams and employment histories in the transition process is above all an empirical problem. Solving this empirical problem requires addressing two prominent issues: First, individuals self-select into different job mobility patterns on the basis of their unobservable characteristics. Second, returns to job mobility are likely to be heterogeneous across different subgroups of the population. Solving these issues, in turn, calls for adequate though possibly unconventional identification strategies and econometric methods. This thesis presented: in chapter 2 a theoretical model of individual job mobility decision-making in a transitional labour market; in chapter 3 a quasi-experimental framework appropriate for the identification and estimation of causal effects of individual behaviour in which the two issues mentioned above can be addressed explicitly; in chapter 4 the empirical results of the returns to early job mobility in transition; and in chapter 5 empirical results on the returns to education in Germany, equally obtained a quasi-experimental framework.

The model in **chapter 2** helped to clarify the role of unobservable characteristics in job mobility decision making and rationalised the endogeneity status of job mobility. Under the model assumptions, it was shown that mainly individuals with either relatively low levels of ability and network access or with relatively high levels of ability and good access to networks do self-select into early job mobility. The model was further shown to be amenable to an interpretation in a quasi-experimental framework in which the subgroups of interest - *always-*

takers, *never-takers* and *compliers* - could be characterised in terms of their unobservable characteristics. On the basis of appropriate assignment to treatment mechanisms, the compliers were interpreted as the subgroup of either forced early job movers or voluntary early job movers. Moreover, the returns to early job mobility were seen to differ significantly across the subgroups of always-takers and compliers.

Chapter 3 presented *Rubin's Causal Model* as a framework and instrumental variables estimation as an appropriate method for the identification and estimation of causal effects of individual decisions such as the returns to early job mobility during transition and the returns to education. Based on the main identification results for causal inference reviewed in this is chapter, **chapter 4** presented empirical results of the *causal* returns to early job mobility on the basis of two different instruments: first, firm-level announced layoffs and, second, occupation-specific labour market opportunities. The resulting instrumental variables estimates suggested that returns do actually vary enormously in the population: while forced movers had to incur substantial losses from changing their job of up to -60% in terms of earnings, voluntary movers might have gained up to +60% from moving. A substantial part of the differences in the returns to early job mobility across different subgroups of compliers could be explained by differences in employment probabilities throughout transition. In particular forced movers had difficulties in finding a new stable job and were often adversely affected by either individual layoff or increased difficulties of new companies later in the transition process.

Based on the observation that educational attainment levels depend crucially on the provision of post-compulsory schooling in proximity to the place of living and on an application of instrumental variables methods to an extension of Rubin's Causal Model, **chapter 5** has corroborated two by now widely accepted findings of several US studies in the case of Germany: first, simple OLS estimates of the returns to education tend to underestimate these returns, and second, the returns to education differ substantially across various subgroups in the population as well as across time. Individuals from 'poor family background' are found to respond most strongly to the provision of post-compulsory schooling in proximity to the place of living. Their response is further most pronounced at low schooling levels whereas the response of individuals with 'rich family background' is most pronounced at higher schooling levels. Changes in the returns to education over time reflect the fact that, first, the fraction of compliers from a poor family background is decreasing while that of compliers from a rich family background is increasing, and second, the effect of school infrastructure on lower schooling levels is decreasing over time while that on higher schooling levels is increasing.

Both of the above presented empirical results have important policy implications. The importance of the schooling infrastructure for educational attainment and the strong heterogeneity in returns to education have a clear bearing on the optimal allocation of schools. Given that the average marginal returns to education are highest among individuals from a disadvantaged family background, improvements in the schooling infrastructure especially in rural areas could prove beneficial by incentivating these individuals to acquire more education and thus improve their long-run prospects in the labor market. It has to be noted, however, that policy conclusions regarding the optimal level of overall spending on 'schooling infrastructure' would require a full cost-benefit analysis and cannot be drawn from the above results.

Finally, the empirical evidence on strong income losses due to early job mobility in transition among forced movers as opposed to large gains among voluntary movers sheds light on the mobility incentives during transition and their impact on the speed of labour reallocation and human capital investment in transition. They also help in identifying 'winners' and 'losers' during transition and in understanding both individual and public opinions about the transition process. Future labour market policies need to recognise the imbalance in mobility incentives during transition as well as the starkly diverging experiences between the various subgroups on the labour market. Similarly to the note of caution above, the estimated 'microeconomic' returns to job mobility do not provide any assessment of the overall 'macroeconomic' effects of labour mobility throughout transition. Contrary to many straightforward theoretical results on these latter in terms of employment or the speed of convergence in wages and living standards, empirical assessments of the 'macroeconomic' *causal* effects of labour mobility in transition are still lacking. Indeed, the transferability of notions and concepts central to the analysis in this thesis - *counterfactuals*, *quasi-experiments*, *compliance with assignment status*, etc. - to a macroeconomic setting is not obvious. While clearly beyond the scope of this study, other - in particular time-series related - concepts of *causality* might be more appropriate for their assessment.