Essays in the Economics of Education and Microeconometrics

Matthias Parey

A Dissertation submitted to the Department of Economics
in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)
University College London (UCL)

London
January 2010

Declaration

I, Matthias Parey, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Matthias Parey

Abstract

This thesis employs microeconometric methods to understand determinants and effects of individual behavior relating to educational choice and consumer demand.

Chapter 2 studies the intergenerational effects of maternal education on a range of children's outcomes, including cognitive achievement and behavioral problems. Endogeneity of maternal schooling is addressed by instrumenting with schooling costs during the mother's adolescence. The results show substantial intergenerational returns to education. The chapter studies an array of potential channels which may transmit the effect to the child, including family environment and parental investments.

The following chapter 3 investigates the effect of studying abroad on international labor market mobility later in life for university graduates. As source of identifying variation, this work exploits the introduction and expansion of the European ERASMUS student exchange program. Studying abroad significantly increases the probability of working abroad, and the chapter provides evidence on the underlying mechanisms.

Chapter 4 compares labor market outcomes between firm-based apprenticeships and full-time vocational schooling alternatives, exploiting the idea that variation in apprenticeship availability affects the opportunities individuals have when they grow up. The chapter documents how variation in vacancies for apprenticeships affects educational choice. The results show that apprenticeship training leads to lower unemployment rates at ages 23 to 26, but there are no significant differences in wages.

Chapter 5 develops a new approach to the measurement of price responsiveness of gasoline demand and deadweight loss estimation. It uses shape restrictions derived from economic theory to match a desire for flexibility with the need for structure in the welfare analysis of consumer behavior. Using travel survey data, the chapter shows that these restrictions remove the erratic behavior of standard nonparametric approaches. Investigating price responsiveness across the income distribution, the middle income group is found to be the most responsive.

Contents

List of Figures

List of Tables

Acknowledgements

Funding Acknowledgements

Most of the work in this thesis has been made possible through a PhD scholarship at the Institute for Fiscal Studies (IFS), and I am grateful to the centre for the research funding and support.

The initial part of the dissertation was made possible and supported through a scholarship of the Hamburger Stiftung für Internationale Forschungs- und Studienvorhaben, awarded in cooperation with the German National Academic Foundation, and I am very grateful for this support.

Conjoint Work

Some of the chapters of this thesis are based on joint work:

1. The work for chapter 2 was conducted jointly with Pedro Carneiro and Costas Meghir.

2. The work for chapter 3 was conducted jointly with Fabian Waldinger.

3. The work for chapter 5 was conducted jointly with Richard Blundell and Joel Horowitz.

I thank these co-authors and collaborators for allowing me to draw in these works in the thesis. The remaining chapters (1, 4 and 6) are entirely my own, as of course is the responsibility for any errors.

Personal Thank Yous

First and most importantly, I would like to thank my supervisor, Costas Meghir, for his guidance and support throughout the work for this thesis. I have benefited immensely from the many discussions and suggestions, and the emphasis on framing key questions as tractable formal problems. His encouragement and enthusiasm have been very important to me.

I am also very grateful to Richard Blundell – his careful guidance and support throughout this work cannot be emphasized enough. The Institute for Fiscal Studies has been a unique and stimulating environment for doing research, and I thank my colleagues at the IFS for all their support and discussions.

I have experienced the Department of Economics at UCL as an inspiring place to discuss research. I am particularly grateful to Pedro Carneiro and Christian Dustmann for their ideas, suggestions, and support.

I would also like to thank my fellow PhD students at UCL and at other universities who I have met during my studies, for discussing and commenting on ongoing work on numerous occasions, especially Ben Etheridge, Josep Mestres, Andrew Shephard, Andreas Uthemann, Hans-Martin von Gaudecker, and Fabian Waldinger.

Finally, I would like to thank Miriam. Throughout this period, she has been a constant source of encouragement and support, for which I am very grateful.

Chapter 1

Introduction

This thesis employs microeconometric methods to understand determinants and effects of individual behavior relating to educational choices and consumer demand.

Chapter 2 investigates intergenerational effects of education. The objective of this work is to provide a detailed analysis of how maternal education affects the human capital accumulation and skill formation of children. Beyond studying outcomes relating to cognitive skills, this work highlights potential transmission channels which may act as mechanism in transmitting the effect of maternal education. In particular, this work studies parental investment behavior and family environment. Identification is achieved through an instrumental variable strategy, using variation in schooling cost during the mother's adolescence. The results indicate that maternal education significantly increases the child's performance in standardized tests for mathematics and reading at ages 7-8, but this effect is not seen at older ages. Although mother's education has a strong effect on maternal labor supply, the results indicate that, nonetheless, more educated mothers invest more into their children along a number of different dimensions, for example by reading to the child. Conceptually, this work is closely related to Currie and Moretti (2003), but using the National Longitudinal Survey of Youth allows to study outcomes at different ages of the child, and to investigate the effect of maternal education on a range of parental investments into their children. This work is also closely related to questions of transmission of intergenerational inequality.

Chapter 3 investigates the long-run effect of international mobility decisions during higher education, motivated by the ongoing internationalization of higher education and increased incidence of study abroad spells (see e.g. Freeman (2009)). Specifically, this chapter studies how international labor market mobility is affected by earlier study abroad experiences during tertiary education. Identification is achieved through exploiting the differential introduction of the large-scale European ERASMUS scholarship program, which strongly increases the incidence of study abroad spells. Studying abroad is found to have substantial positive effects on the probability of working abroad after graduation, and this effect is found to be robust in a range of sensitivity checks. In terms of policy implications, these results indicate that an opportunity to attract talented graduates is through student exchange opportunities. The chapter also sheds light on the underlying mechanism for the effect. Descriptive evidence indicates that location choices are sticky, i.e. that graduates tend to return to work where they previously studied abroad. Furthermore, students with study abroad experience are more likely to state that they work abroad because of an interest in foreign cultures, a career abroad, or because of their partner.

Chapter 4 investigates training choices of young school leavers from lower-track schools in Germany, who decide between apprenticeship training and a full-time vocational schooling alternative. Although there is a large descriptive literature comparing these two forms of vocational training, little is known about the causal effect of these alternatives (Ryan, 2001). This chapter argues that differential availability of training opportunities affects the choice of young people between these alternatives. The chapter presents an open-economy framework in which aggregate price shocks affect local training choices, but have no differential effect on factor rewards; this provides an economic framework which motivates an exclusion restriction. The chapter then exploits local shocks to training availability to study how alternative training forms affect labor market outcomes at ages 23 to 26 in the main labor market. The results indicate that the main benefit of (former) apprentices is through a lower initial probability of being unemployed, while productivity (as measured by wages) is not statistically different. Evidence from a firm closure experiment suggests that this attachment effect has a strong firm-specific component, which is lost when the firm closes down. Overall, the evidence indicates that the (former) apprentices have a transitory advantage in the form of lower unemployment rates.

All of these chapters share a concern for appropriately addressing endogeneity of the schooling choice and the possibility of a correlation between schooling choice and unobservable characteristics. This selection problem is addressed by exploiting differences in the cost and availability of educational opportunities as exogenous shifters in the educational choice, which can then be used to identify the treatment effect. While not always statistically significant, the differences between a straightforward comparison of means and the use of instrumental variables highlight the importance of accounting for selection effects. A focus in this work is to understand heterogeneity in effects, where the estimand captures the effect on the sub-population which is affected by the instrument, following the work by Imbens and Angrist (1994).

Chapter 5 investigates how shape restrictions which arise from economic theory can be imposed in nonparametric Kernel regression, in an application to consumer demand for gasoline. This is motivated by the work of Hausman and Newey (1995), who emphasize the usefulness of nonparametric estimation in understanding gasoline demand patterns. As has been noted previously, see e.g. Schmalensee and Stoker (1999), the resulting demand estimates appear erratic in certain regions and are difficult to reconcile with the properties we would expect the demand function to have. In this work, these patterns are interpreted as resulting from sampling variation. Instead of choosing a particular functional form, this work imposes structure based on economic theory. For this purpose, the Slutsky constraint is imposed on the estimated demand function. This is implemented by making use of a re-weighting procedure suggested by Hall and Huang (2001), which has a number of favorable properties relative to alternative methods of imposing constraints. Imposing the economic restriction leads to well-behaved estimates of the demand function without the need for often arbitrary functional form assumptions. This procedure appears to be attractive because it brings closer together the estimation procedure and the underlying economic theory. The implications for resulting Deadweight Loss estimates of taxation are explored. A substantial focus of this chapter is on understanding how the price responsiveness of gasoline demand varies across the income distribution. From a policy perspective, this is of great importance because it informs us about how different parts of the income distribution are affected by gasoline taxation. The results indicate that price-responsiveness differs in a non-monotonic fashion across income, with the middle income group being most responsive to price changes. — The last chapter concludes.

Chapter 2

Maternal Education, Home Environments and the Development of Children and Adolescents

2.1 Introduction

"... the forces that are driving the transition are leading to two different trajectories for women - with different implications for children. One trajectory - the one associated with delays in childbearing and increases in maternal employment - reflect gains in resources, while the other - the one associated with divorce and nonmarital childbearing - reflects losses. Moreover, the women with the most opportunities and resources are following the first trajectory, whereas the women with the fewest opportunities and resources are following the second." (McLanahan, 2004)

The above quote is from Sara McLanahan's presidential address to the Population Association of America, in which she documents a striking increase in inequality in children's home environments across families where mothers have different levels of education.[1] The trends documented in these and other papers, starting with Coleman et al. (1966), are cause for great concern because the home environment is probably the best candidate for explaining inequality in child development.[2]

To address this problem, McLanahan (2004) ends her paper by proposing a set of changes to the welfare system. The effectiveness of such proposals is still to be assessed. However, given that home environments are rooted in the experiences of each family, they are probably difficult to change if we rely only on the welfare system. Furthermore, more direct interventions require invading family autonomy and privacy and are notoriously difficult to enforce. Therefore, one possible alternative is to target future parents in their youth, by affecting their education, before they start forming a family. In this work we assess the potential for such a policy, by estimating the impact of maternal education on home environments and on child outcomes.

Our analysis is based on the Children of the National Longitudinal Survey of Youth of 1979, a data set with very detailed information on maternal characteristics, home environments, and child outcomes. Since the data covers mothers and children over several years it allows a unified treatment of different aspects of child development across ages, including cognitive, noncognitive, and health outcomes.[3] Furthermore, using this single data set it is possible to estimate the impact of maternal education not only on parental characteristics like employment, income, marital status, spouse's education, age at first birth, but also on several aspects of parenting practices. This chapter provides a detailed analysis of the possible mechanisms mediating the rela-

[1] She examines trends in six dimensions of home environments over the last 50 years: age of mothers of young children (below 5), maternal employment, single motherhood, divorce during the first 10 years of marriage, father's involvement, and family income. In this work we consider a more detailed set of measures.

[2] For example, Jencks and Phillips (1998), Cameron and Heckman (2001), Fryer and Levitt (2004, 2006, 2007), Carneiro, Heckman, and Masterov (2005), Todd and Wolpin (2006) and others show how differences in home environments account for a large share of the black-white test score gap.

[3] The dynamic aspect of cognitive and noncognitive skill formation is emphasized in the recent literature on child development, such as Carneiro and Heckman (2003), Cunha, Heckman, Lochner, and Masterov (2005), Cunha and Heckman (2007), and Todd and Wolpin (2003).

tionship between parental education and child outcomes. The novelty of our work is in the systematic treatment of a very large range of inputs and outputs to the child development process, at different ages of the child, in a unified framework and data set. We also compare the relative roles of maternal education and ability,[4] and we show how the role of maternal education varies with the gender and race of the child, and with the cognitive ability of the mother.

We show that maternal education has positive impacts both on cognitive skills and behavioral problems of children, but the latter are more sustained than the former. This is perhaps because behavior is more malleable than cognition (e.g., Carneiro and Heckman (2003)). Especially among whites, there is considerable heterogeneity in these impacts, which are larger for girls, and for mothers with higher cognition.

More educated mothers are more likely to work and work for longer hours, especially among blacks. This is true independently of the child being in its infancy, childhood, and adolescence. Nevertheless, there is no evidence that more educated mothers do less breastfeeding, spend much less time reading to their children, or even taking them on outings. This is important because some studies suggest that maternal employment may be detrimental for child outcomes if it leads to reduced (quality) time with children.

Due to the nature of the data, this work focuses on the effect of maternal, but not paternal, schooling. Because of assortative mating, part of the effects we find may be driven by the father's schooling through a mating effect. However, unless the effect of partner's schooling is incredibly large, assortative mating cannot fully explain our main results, as suggested in some of the literature.

The key empirical problem we face is controlling for the endogeneity of mother's schooling: factors that influence the mother's decision to obtain schooling may also affect her ability to bring up children or may relate to other environmental and genetic

[4]Maternal cognitive ability is a central determinant of child's cognitive achievement. According to Todd and Wolpin (2006), racial differences in mother's cognition account for half of the minority-white test score gap among children.

factors relevant to child outcomes. To deal with this issue we exploit differential changes in the direct and opportunity costs of schooling across counties and cohorts of mothers, while controlling both for permanent differences and aggregate trends as well as numerous observed characteristics such as mother's ability. The variables we use to measure the costs of education include local labor market conditions, the presence of a four year college, and college tuition at age 17, in the county where the mother resided when she was 14 years of age. These variables have previously been used as instruments for schooling by Card (1993), Kane and Rouse (1993), Currie and Moretti (2003), Cameron and Taber (2004), and Carneiro, Heckman, and Vytlacil (2006), among others. We also control for county fixed effects, to allow for permanent differences in area characteristics and in the quality of offered education, as well as for mother's cohort effects, to allow for common trends, thus leaving only the differential changes in local costs of education between counties and cohorts to drive the results. To provide evidence in favor of our exclusion restrictions we show that our instruments cannot predict early measures of mother's personality and health limitations.

One potential problem is that our instruments may be weak. We study the importance of this problem in the context of a fixed coefficient model, since not much is known about the effects of weak instruments in the estimation of a random coefficient model. In particular, we estimate some of our models by limited information maximum likelihood (LIML), as suggested by Staiger and Stock (1997). The resulting estimates are larger in absolute value than our original two stage least squares estimates and further away from the OLS coefficients, but also have larger standard errors (as predicted by Blomquist and Dahlberg (1999)).

Recently, several papers have appeared on this topic dealing with the endogeneity issue in different ways. Behrman and Rosenzweig (2002) compare the schooling attainment of children of twin mothers and twin fathers (with different levels of schooling). They find that the effect of father's education is strong and large in magnitude, but the effect of maternal education on child schooling is insignificant (see also Antonovics and Goldberger (2005); Behrman and Rosenzweig (2005)).

Black, Devereux, and Salvanes (2005), Oreopoulos, Page, and Stevens (2003), Chevalier (2004), Chevalier, Harmon, O'Sullivan, and Walker (2005), Maurin and McNally (2005), and Galindo-Rueda (2003) use an instrumental variables strategies to estimate the effect of parental education on child outcomes, exploring changes in compulsory schooling or in examination standards. Each paper focuses on different outcomes, but child's education is common across papers. Their findings are quite diverse.

Currie and Moretti (2003) find that maternal education has significant effects on birth-weight and gestational age. Maternal education also affects potential channels by which birth outcomes are improved such as maternal smoking, the use of prenatal care, marital status, and spouse's education. Related studies by Plug (2004), Sacerdote (2002) and Bjoerklund, Lindahl, and Plug (2006), which are based on adoptions data, compare the correlation between parental schooling and the outcomes of biological children, with the correlation between foster parents' schooling and adopted children's schooling. Adoption studies inform the debate by separating the effect of environmental and genetic factors (although their standard design can be problematic if there are substantial interactions between genes and environments), but they do not tell us directly about the causal effect of parental schooling on child outcomes. These studies cannot distinguish between the role of parental schooling and ability in the provision of better environments. Plug (2004) finds weak effects of adoptive mother's schooling on child's schooling but large effects of father's schooling, and Bjoerklund, Lindahl, and Plug (2006) find strong effects of both adoptive father and mother's schooling. Sacerdote (2002) argues that a college educated adoptive mother is associated with a 7% increase in the probability that the adopted child graduates from college. The general sense we get from the whole literature is that the results are quite disparate and a consensus has not formed yet (see Holmlund, Lindahl, and Plug (2006)).[5]

[5] Holmlund, Lindahl, and Plug (2006) replicate the differing findings based on twin studies, adop-

The plan of the chapter is as follows. In the next section we describe the data, followed by an explanation of our empirical strategy. Then we discuss our results on the impact of mother's schooling on child outcomes, followed by results on the possible mechanisms through which schooling may operate. Finally, we present a sensitivity analysis and a concluding section.

2.2 Data

We use data from the National Longitudinal Survey of Youth (NLSY79). This is a panel which follows 12,686 young men and women, aged between 15 and 22 years old in the first survey year of 1979. Surveys are conducted annually from 1979 until 1994, and every two years from 1994 onwards. We use data up to 2002.

Apart from the main cross-sectional sample representative of the population, the NLSY79 contains an over-sample representative of blacks and hispanics, an over-sample of economically disadvantaged whites, and a sample of members of the military. In our analysis we exclude the over-sample of economically disadvantaged whites and the sample of the military. This ensures that our sample is drawn according to pre-determined characteristics. Attrition rates are very low (see CHRR (2002)). As we describe below, for our purpose only the females of the NLSY79 are of interest.

We measure mother's schooling as completed years of schooling. Since we observe mothers over a number of years, we have multiple observations of years of schooling. We are interested in the mother's schooling at the time when the outcome is measured.[6]

The data contains detailed information on family background of the mother, namely her parents' schooling, and whether she was raised by both her biological parents. Furthermore, we know the mother's score in the Armed Forces Qualification Test

tions, and instrumental variables within one Swedish data set, suggesting that the differences cannot be fully explained by country specifics or sample characteristics.

[6]Occasionally, sample members do not answer this question in the year of interest. In order to include these observations, we take as the measure of schooling the maximum number of completed years reported up to the year of interest.

(AFQT), administered in 1980, which we use as a measure of mother's cognitive ability. The original AFQT score may be influenced by the amount of schooling taking up to the test date, but it is possible to estimate the effect of schooling on the test score (see Hansen, Heckman, and Mullen (2004)), and then construct a separate measure of ability (we apply the same procedure as in Carneiro, Heckman, and Vytlacil (2005)). Throughout the chapter, we refer to the AFQT score as this schooling-corrected ability measure, normalized to have mean zero and standard deviation one.

In 1986, when the females of the NLSY79 were between 22 and 29 years old, another data set, the Children of the NLSY79, was initiated. It follows the children of the female members of the NLSY79 over time and surveys each child throughout childhood and adolescence. Questionnaires are tailored to the age of the child, and information is collected from both the mother and the child. We match the information on each child of the NLSY79 to the data of the mother. Even though the NLSY79 surveys a random sample of potential mothers, the design of the children's sample leads to an initial oversample of children of younger mothers, until all women are old enough and have completed their child-bearing period. In 2000, the women of the NLSY79 have completed an average of 90% of their expected childbearing (CHRR, 2002).

Table 2.1 presents an overview of the different outcomes for reference. In order to measure the child's cognitive ability we use the Peabody Individual Achievement Tests (PIAT) in math and reading, which are widely used in the literature. Behavior problems are measured using the Behavior Problems Index (BPI).[7] We also construct grade repetition[8] and child obesity indicators.

In addition, we examine potential transmission channels: mother's age at birth, an indicator variable for whether the mother is married, years of schooling of the mother's

[7]Based on data from the UK National Child Development Survey, Currie and Thomas (2001) and Carneiro, Crawford, and Goodman (2007) show that early test scores and early measures of behavioral problems are strongly associated with adolescent and adult labor market outcomes, health, and engagement in risky behaviors.

[8]In the NLSY79, mothers are asked whether their child ever repeated a grade in school and which grade the child repeated. We set observations to missing if the mother's set of answers to grade repetition is not consistent.

Table 2.1: Outcome variables

Name	Definition
	Child outcomes (ages 7-8 and 12-14)
PIAT math	Peabody Individual Achievement Test Mathematics. Age-specific score with population mean 0 and variance 1.
PIAT read.	Peabody Individual Achievement Test Reading Comprehension. Age-specific score with population mean 0 and variance 1.
BPI	Behavior Problem Index. Gender-age specific score with population mean 0 and variance 1.
Grade repetition	Indicator for whether child has ever repeated a grade
Overweight	Indicator for whether child is overweight: Takes value 1 if child's Body Mass Index (BMI) is larger than the 95th percentile of age-gender specific distribution.
	Family environment (ages 7-8)
Maternal age*	Age of the mother at birth of the child (in years)
Number of children*	Total number of children ever reported by the mother.
Marital status	Indicator for whether the mother is married
Spouse's schooling	Years of schooling of mother's spouse.
Hours worked	Number of hours mother worked in past year
Log family income	Log of total annual family income
Maternal aspirations	Indicator for whether mother believes that child will go to college
	Parental investment measures (ages 7-8 and 12-14)
Museum	Indicator for whether child is taken to museum several times or more in last year
Musical instrument	Indicator for whether there is a musical instrument child can use at home
Special lessons	Indicator for whether child gets special lessons
Mother reads	Indicator for whether mother reads to child at least three times a week
Newspaper	Indicator for whether family gets a daily newspaper
Computer	Indicator for whether child has a computer in his/her home
Adult home	Indicator: takes the value 1 if adult is present when child comes home after school, and 0 if no adult is present or if child goes somewhere else.
Joint meals	Indicator for whether child eats with both parents at least once per day.
	Early child outcomes (ages 0-1)
Low birthweight	Indicator for whether child's birthweight is 5.5 lbs or less
Motor skills	Motor and social development scale (MSD), gender-age specific score standardized to mean 0 and variance 1.
	Early investments (ages 0-1)
Smoking during pregnancy*	Indicator for whether mother smoked in the year prior the child's birth
Weeks breastfeeding*	Number of weeks mother was breastfeeding
Formal child care	Indicator for whether formal childcare arrangements were in place for at least six months over past year
Hours worked	Number of hours mother worked in past year
Mother reads	Indicator for whether mother reads at least three times a week to the child
Books	Number of books child has
Soft toys	Number of cuddly, soft or role-playing toys child has
Outings	Indicator for whether the child gets out of the house at least four times a week
	Adolescent outcomes (ages 18-19)
Enrollment	Indicator for enrollment status of the young adult
Conviction	Indicator for whether the young adult has been convicted up to the age of interest
Number of own children	Total number of own children born to the young adult up to the age of interest
	Falsification exercise (ages 7-8)
Mother's sociability*	Indicator for maternal sociability at age 6.
Mother's early health problems*	Indicator for whether the mother had health limitations before age 5

Note: Age ranges (in italics) refer to the child and define at which child age this outcome is included in the outcome regression. Not all variables vary across time, but we follow the same sample selection principle for consistency. Variables which do not vary across time are indicated by a star (*).

spouse, log of total family income (for couples, it includes both husband's and wife's incomes), number of hours the mother worked in a year, maternal aspirations of the child's educational achievement, and number of children. We take the child's age as the relevant reference point for observing the measures of interest.

One unusual feature of the data set we use is that it contains direct measures of parenting behaviors, which can also be studied as mediating channels. In particular, we look at whether: the child is taken to the museum; there is a musical instrument at home; the child gets special lessons; the mother reads to the child; newspaper and computer are available; there is adult supervision after school; and there are joint meals with both parents (Table 2.1).

Finally, we study children's outcomes very early in life and in adolescent years. Early measures include an indicator function for low birth-weight, and the standardized score on the Motor and Social Development scale (MSD), an assessment of early motor, social and cognitive developments. We focus on ages 0 to 2. As early investments, we study smoking during pregnancy, weeks breastfeeding, use of formal child care and hours worked, and indicators for whether the mother reads to the child, how many books and soft toys the child has, and an indicator for whether the child gets out of the house regularly. Adolescent outcomes are measured at ages 18-19 and include school enrolment, criminal convictions and number of own children.

In the next section we discuss in detail our instrumental variable strategy, its justification and validity. Before we do so, we explain how the instruments are constructed. The instruments for mother's schooling are average tuition in public four-year colleges (in 1993 prices), distance to four-year colleges (an indicator whether there is a college in the county of residence), local log wage and local unemployment rate. When assigning the instruments to mothers, our general approach is the following: we assign values that correspond to the year when the mother was 17, in order to be relevant for educational choices towards the end of highschool; in order to avoid any potentially endogenous re-location around that period, we use maternal location at age 14. The local wage variable is local log wages in the county of residence where the mother resided at 14, but measured in the year when the mother is aged 17 (based on county data from the Bureau of Economic Analysis, Regional Economic Accounts, and adjusted to 2000 prices using the CPI). The state unemployment rate data comes from the BLS.[9] The unemployment variable is again assigned to state of residence at 14, and measured at age 17. The distance variable, which is from Kling (2001), is an indicator variable whether in 1977 there is a four-year college in the county of residence. Annual records on tuition, enrollment, and location of all public two- and four year colleges in the United States were constructed from the Department of Education's annual Higher Education General Information Survey and Integrated Postsecondary Education Data System 'Institutional Characteristics' surveys. By matching location with county of residence, we determined the presence of two-year and four-year colleges. Tuition measures are enrollment weighted averages of all public four-year colleges in a person's county of residence, or at the state level if there is no college in the county.

The data set, limited to the subsamples of interest for which all maternal variables are observed, contains information on a total of 4,379 white children from 1,948 white mothers, and 3,051 children from 1,211 black mothers. For some children, we observe the outcome more than once during the age range of interest. To increase precision of our estimates, we pool all available observations within the specific age range. We cluster all standard errors by cohort and county of mother's residence at age 14, thus allowing for arbitrary dependence between repeat observations from a particular child, and between outcomes of several children from one mother, and more generally for arbitrary dependence within county-cohort cells.

To give a sense of what our sample looks like, the following Table 2.2 shows summary statistics for the covariates based on the sample from our PIAT math regression.

[9] State unemployment data is available for all states from 1976 on, and it is available for 29 states for 1973, 1974 and 1975, and therefore for some of the individuals we have to use the unemployment rate in the state of residence in 1976 (which will correspond to age 19 for those born in 1957 and age 18 for those born in 1958).

There are some strong differences between the black and the white sample. Average

Table 2.2: Descriptive sample statistics

	Whites (1)	Blacks (2)
Mother's yrs. of schooling	13.236	12.670
	[2.185]	[1.919]
Mother's AFQT (corrected)	0.367	-0.458
	[0.882]	[0.774]
Grandmother's yrs. of schooling	11.719	10.541
	[2.278]	[2.677]
Grandfather's yrs. of schooling	11.813	9.798
	[3.114]	[3.612]
'Broken home' status	0.207	0.437
	[0.406]	[0.496]
Child age (months)	95.166	95.821
	[6.979]	[6.937]
Child female	0.495	0.498
	[0.500]	[0.500]
College availability	0.519	0.598
	[0.500]	[0.491]
Local tuition	2.133	1.964
	[0.851]	[0.830]
Local unemployment	7.161	6.928
	[1.752]	[1.521]
Local wages	10.270	10.245
	[0.186]	[0.213]
Observations	2492	1271

Note: The table reports sample means and (in brackets) standard deviations for covariates and instruments, based on the sample of our PIAT math outcome regression for children aged 7 to 8 (see Tables 2.5 and 2.7).

years of schooling are 0.6 years higher for whites. Also, note the strong difference in the corrected AFQT score: since this variable is normed to have a standard deviation of 1 in the population, the means of these two groups are more than 0.8 of a standard deviation apart. The 'broken home' status is an indicator for whether the mother grew up with both biological parents status; it is more than twice as prevalent in the black sample compared to the white.

2.3 Empirical Strategy

We assume that child outcomes (y_i) are determined by mother's years of schooling (S_i) as well as a set of observable (X_i) and unobservable factors. Schooling is determined by the same factors as child outcomes, and by a set of instruments (Z_i) that reflect the measured direct and indirect costs of schooling. In interpreting the results we assume that the effects of schooling on outcomes depends on unobservables and that the IV estimates will represent Local Average Treatment Effects (LATE).[10]

We also allow the coefficient on maternal schooling to depend on observable characteristics. We define four groups depending on the sex of the child and on whether the mother is characterized by high or low ability based on her AFQT score. These four group indicators will be denoted by D_{ij}, and take the value 1 if observation i belongs to group j ($j = 1...4$). A_i denotes child age. Thus our estimating equation is

$$\begin{aligned} y_i \;=\; & \sum_j \beta_j D_{ij}\, S_i \;+\; \sum_j \gamma_{1j} D_{ij}\, X_{mi} \;+\; \sum_j \gamma_{2j} D_{ij} \;+\; \sum_j \gamma_{3j} D_{ij}\, A_i \\ & +\, \gamma_4\,(\text{county FE}) \;+\; \gamma_5\,(\text{cohort FE}) \;+\; u_i \end{aligned} \tag{2.1}$$

where X_{mi} (indexed by m for maternal characteristics) include corrected AFQT score, grandmother's schooling, grandfather's schooling, and an indicator for mother's broken home status. The corresponding first stage regressions ($k = 1...4$) are:

$$\begin{aligned} S_i D_{ik} \;=\; & \sum_j \delta_{1j} D_{ij}\, Z_i \;+\; \sum_j \delta_{2j} D_{ij}\,(X_{mi} * Z_i) \;+\; \sum_j \delta_{3j} D_{ij}((\text{cohort FE}) * Z_i) \\ & +\, \sum_j \delta_{4j} D_{ij}\, X_{mi} \;+\; \sum_j \delta_{5j} D_{ij} \;+\; \sum_j \gamma_{6j} D_{ij}\, A_i \\ & +\, \delta_7\,(\text{county FE}) \;+\; \delta_8\,(\text{cohort FE}) \;+\; \epsilon_i \end{aligned} \tag{2.2}$$

where the asterisk ($*$) denotes the Kronecker product. Note that in the first term we leave out the variable 'distance to college', because in our data set this variable does not

[10] see Imbens and Angrist (1994).

vary over time (since it is only measured in 1977). To estimate average effects across groups, we apply the Minimum Distance procedure (Rothenberg, 1971; Chamberlain, 1984) using as weights the covariance matrix of the unrestricted coefficients.

One part of the direct cost of schooling is the amount of tuition fees a student faces and how far she has to travel to attend college. These variables have frequently been used as instruments (e.g. Kane and Rouse (1993), Card (1993), Currie and Moretti (2003), Cameron and Taber (2004), Carneiro, Heckman, and Vytlacil (2006)). Another major cost of acquiring higher education is foregone earnings. We proxy these variables by using the local unemployment rate, reflecting the speed with which someone can find work, and the local wages, as a direct measure of foregone earnings and as a determinant of expectations about future conditions. Both these variable also capture temporary shocks to family income. Therefore, it is not possible to determine a priori whether these variables have a positive or negative effect on maternal schooling, and the effect may well vary across individuals.[11] A key element of our approach is that we include both cohort and county fixed effects, thus relying on the way the instruments change across counties and cohorts to identify our effects.

Our instruments must be correlated with mother's schooling, but must not have an independent effect on the outcome equation except through mother's schooling. We discuss these conditions in turn.

Underlying the use of geographical variation in schooling costs is the presumption that *local* variables matter for the schooling choice of the individual. In principle, individuals might move to a different location for their studies, e.g. in order to avoid high tuition costs. Still, it seems reasonable to believe that local variation matters: Moving is costly for a variety of reasons: the student is prevented from the option of living at home. Furthermore, movers may be disadvantaged in the form of higher out-of-state tuition. Currie and Moretti (2002) report evidence that the majority of students do not move to a different state to go to college (see also Hoxby (1997)).

---

[11] See Cameron and Taber (2004) and Arkes (2005).

Table 2.3 shows the effect of schooling cost variables on maternal schooling, where for consistency the sample of interest are white children aged 7 and 8. Similar re-

Table 2.3: Maternal schooling choices and schooling costs

Dependent variable: Mother's years of schooling	
Mother's AFQT (corrected)	0.937
	[0.065]***
Grandmother's yrs. of schooling	0.158
	[0.030]***
Grandfather's yrs. of schooling	0.149
	[0.024]***
'Broken home' status	-0.249
	[0.144]*
Local unemployment	-0.134
	[0.071]*
Local wages	-4.883
	[2.120]**
Local tuition/1000	0.376
	[0.365]
Observations	2492
F-statistic	2.01
p-value	0.000***

Note: This table shows the result for a regression of maternal schooling on her characteristics and schooling cost variables, where schooling cost variables are also interacted with AFQT, grandparents' schooling, broken home indicator, and mother's birth cohort dummies. County fixed effects included. The table reports estimated marginal effects of a change in the variable indicated, evaluated at the mean. F-statistic and corresponding p-value refer to the joint test that all of these 47 schooling cost variables are zero. The sample is selected to be identical to the PIAT math regression in our main results, see Table 2.6. Standard errors, clustered by birth cohort and county are reported in brackets. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level. See text for details.

sults hold for other ages. We do not yet interact with the four group indicators as we do in the main results below. The table reports marginal effects of each regressor.[12] Mother's ability level and grandparents' schooling are important determinants of maternal education. The instruments are jointly significant at the 1% level although they are not all individually significant.

We have allowed the instruments to interact with a number of covariates reflecting maternal background to help improve the overall predictive ability of the instruments.

[12]The main effect of living near a college is not identified because it does not vary with time and we include county fixed effects. However we do interact it with a number of maternal background characteristics as described above.

In our sensitivity analysis we show that our results are robust to very flexible specification of the outcome equations by including polynomials in maternal covariates as well as interactions between them; thus the interactions in the instrument set are not picking up non-linearities left out of the outcome equations, but allow better predictions by modeling the heterogeneity in the schooling choice.

The second requirement for our instruments is that they should not have an independent effect on the outcome, conditional on other covariates. Thus the differential changes in the costs of schooling should not predict child outcomes, conditional on covariates. By controlling for county fixed effects we avoid biases due to geographical sorting. The latter relates to individuals moving to certain counties in a way which creates a correlation between the characteristics of the region (e.g. local labor market conditions, tuition fees, etc), and outcome relevant variables such as the unobserved human capital of the person moving - the mother in our case. The fact that such sorting takes place is well established (e.g., Solon (1999), Dahl (2002)).

The second concern relates to college quality as well as local labor market conditions. If higher tuition fees are associated with higher college quality, and if higher college quality makes mothers better at child rearing, then this could bias our results. First, we use tuition from public colleges only; any link between cost and quality can be expected to be weaker in comparison to private colleges. Second, a main determinant of college quality is the quality of the students; this aspect is captured by including an ability measure of the mother, and by including family background variables. But perhaps most importantly we *do not* rely on comparing mothers who faced different tuition levels. We exploit changing tuition, which relies on the trends being common across regions, as in the diff-in-diff context. Therefore, it does not seem likely that, after controlling for mother's ability, mother's family background, and county fixed effects, endogeneity of tuition due to college quality will pose a problem. A similar argument can be made for the local labor market conditions.

Our instruments are designed to relate mainly to late schooling or college choice.

They should be unrelated to early background characteristics of the mother. In our data there is a measure of mother's sociability at age 6, and a measure of maternal health limitations before age 5, which can be used to check the validity of our instruments.[13]

We next examine whether these instruments predict early sociability and health conditional on our controls. We regress these two measures on maternal schooling and the controls, instrumenting schooling with the variables described above. As in the rest of the chapter, the unit of observation in each regression is the child at age 7 or 8, even though the regression relates to the mother only. Therefore there may be more than one observation per mother, since some mothers have several children.

Table 2.4: Instrument validity

	Falsification exercise			
	Sociability at age 6		Early health limitations	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
Mother's schooling: All	0.019	0.007	-0.014	-0.010
	[0.009]**	[0.022]	[0.007]*	[0.020]
Mother's schooling: Male child	0.014	0.020	-0.017	0.016
	[0.010]	[0.026]	[0.009]*	[0.024]
Mother's schooling: Female child	0.028	-0.006	-0.012	-0.035
	[0.011]**	[0.026]	[0.008]	[0.023]
Mother's schooling: High AFQT	0.019	0.023	-0.009	0.000
	[0.012]	[0.033]	[0.008]	[0.027]
Mother's schooling: Low AFQT	0.020	-0.008	-0.026	-0.023
	[0.013]	[0.032]	[0.013]**	[0.030]
Mother's AFQT (corrected): All	-0.029	-0.020	-0.033	-0.045
	[0.031]	[0.036]	[0.027]	[0.031]
Observations	4322	4322	4395	4395
Mean	0.390	0.390	0.197	0.197
Standard deviation	0.488	0.488	0.398	0.398

Note: This table reports Minimum Distance estimates for the groups indicated based on equation (2.1), see text for details. A description of the outcome variables is found in Table 2.1 on page 25. Standard errors reported in brackets, clustered by county-cohort. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

Table 2.4 presents OLS and IV results for each early measure. Notice that final maternal schooling is strongly associated with both early sociability and early health

[13]Maternal sociability is an indicator for whether the mother indicates that at age 6 she was somewhat outgoing or extremely outgoing rather than somewhat shy or extremely shy. Early health limitations is an indicator for whether the mother reported any health limitations that she had either all her life or that began before age 5.

limitations of the mother in the OLS regressions, but not in the IV regressions. In the latter the coefficient on schooling is smaller and statistically not different from zero. This is what we would expect if our identification strategy is valid.

2.4 Results

2.4.1 Effects on Child Outcomes

Our main outcome variables are the PIAT mathematics and reading test, the BPI, and binary indicators for grade repetition and child obesity. The PIAT tests and the BPI are standardized to have mean zero and variance 1 in a nationally representative sample. We measure these variables at both ages 7-8 and 12-14.

(a) White Children

Tables 2.5 and 2.6 present our main results for white children. The first line shows the estimates for the whole sample, while the following four lines show effects for different subgroups of interest. The last line of the table corresponds to the overall effect of the mother's AFQT score on child outcomes. This variable is a very strong predictor of children's test scores and it is useful to compare the role of maternal schooling and ability in our results. Each estimate is computed as Minimum Distance estimates based on equation (2.1). Standard errors are clustered at the county-cohort level.

OLS results indicate that one year of additional mother's education increases mathematics standardized scores by 5% of a standard deviation at ages 7 and 8, while the IV coefficient is 10% (the difference between OLS and IV is significant at the 8% level). The results for the reading score at ages 7 and 8 are similar to those for the math score, but somewhat smaller. However, at ages 12 to 14 the effect of mother's schooling on both math and reading become small and insignificant in the IV results.

Mother's education also has strong effects on child behavioral problems (BPI) at both ages. There is an interesting pattern in these results: the effects on math and

Table 2.5: Child outcomes – OLS results: White children

	OLS estimates: White children									
	PIAT math		PIAT read.		BPI		Grade repetition		Overweight	
	7-8 yrs	12-14 yrs	7-8 yrs	12-14 yrs	7-8 yrs	12-14 yrs	7-8 yrs	12-14 yrs	7-8 yrs	12-14 yrs
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Mother's schooling: All	0.050	0.034	0.029	0.035	-0.087	-0.102	-0.005	-0.023	-0.009	-0.007
	[0.012]***	[0.017]**	[0.012]**	[0.014]**	[0.015]***	[0.018]***	[0.003]*	[0.005]***	[0.004]**	[0.005]
Mother's schooling: Male child	0.040	0.045	0.029	0.044	-0.078	-0.119	-0.007	-0.026	-0.014	-0.007
	[0.015]***	[0.021]**	[0.016]*	[0.019]**	[0.018]***	[0.022]***	[0.004]*	[0.007]***	[0.006]**	[0.008]
Mother's schooling: Female child	0.058	0.023	0.029	0.028	-0.099	-0.082	-0.004	-0.021	-0.005	-0.008
	[0.015]***	[0.021]	[0.015]*	[0.018]	[0.021]***	[0.023]***	[0.004]	[0.007]***	[0.006]	[0.006]
Mother's schooling: High AFQT	0.055	0.044	0.048	0.051	-0.093	-0.110	-0.006	-0.013	-0.009	-0.004
	[0.015]***	[0.022]**	[0.016]***	[0.018]***	[0.019]***	[0.023]***	[0.003]*	[0.006]**	[0.005]*	[0.006]
Mother's schooling: Low AFQT	0.040	0.022	-0.000	0.012	-0.076	-0.091	-0.005	-0.045	-0.010	-0.013
	[0.019]**	[0.024]	[0.020]	[0.021]	[0.026]***	[0.029]***	[0.009]	[0.009]***	[0.008]	[0.008]*
Mother's AFQT (corrected): All	0.145	0.185	0.132	0.250	-0.077	-0.010	-0.019	0.001	-0.014	-0.055
	[0.040]***	[0.050]***	[0.046]***	[0.054]***	[0.060]	[0.063]	[0.012]	[0.018]	[0.018]	[0.020]***
Observations	2492	2113	2353	2095	2565	2264	1191	1958	2533	2271
Mean	0.314	0.254	0.491	0.047	0.293	0.464	0.026	0.111	0.114	0.125
Standard deviation	0.782	0.874	0.805	0.854	0.990	0.986	0.159	0.314	0.318	0.331

Note: This table reports Minimum Distance estimates for the groups indicated based on equation (2.1), see text for details. A description of the outcome variables is found in Table 2.1 on page 25. Standard errors reported in brackets, clustered by county-cohort. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

Table 2.6: Child outcomes – IV results: White children

	IV estimates: White children									
	PIAT math		PIAT read.		BPI		Grade repetition		Overweight	
	7-8 yrs	12-14 yrs	7-8 yrs	12-14 yrs	7-8 yrs	12-14 yrs	7-8 yrs	12-14 yrs	7-8 yrs	12-14 yrs
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Mother's schooling: All	0.097	0.024	0.075	0.018	-0.092	-0.116	-0.028	-0.028	-0.015	-0.012
	[0.031]***	[0.033]	[0.033]**	[0.033]	[0.043]**	[0.041]***	[0.008]***	[0.011]**	[0.013]	[0.013]
Mother's schooling: Male child	0.060	0.037	0.053	0.043	-0.052	-0.091	-0.029	-0.016	-0.009	0.004
	[0.042]	[0.044]	[0.047]	[0.053]	[0.054]	[0.056]	[0.009]***	[0.015]	[0.018]	[0.020]
Mother's schooling: Female child	0.125	0.013	0.088	0.007	-0.131	-0.134	-0.026	-0.035	-0.020	-0.019
	[0.038]***	[0.041]	[0.040]**	[0.038]	[0.053]**	[0.049]***	[0.011]**	[0.013]***	[0.016]	[0.015]
Mother's schooling: High AFQT	0.146	0.032	0.107	0.024	-0.100	-0.114	-0.032	-0.020	-0.016	-0.020
	[0.045]***	[0.041]	[0.042]**	[0.043]	[0.057]*	[0.052]**	[0.010]***	[0.014]	[0.017]	[0.018]
Mother's schooling: Low AFQT	0.046	0.013	0.028	0.011	-0.081	-0.118	-0.018	-0.041	-0.014	-0.004
	[0.046]	[0.047]	[0.051]	[0.048]	[0.063]	[0.059]**	[0.013]	[0.017]**	[0.020]	[0.018]
Mother's AFQT (corrected): All	0.086	0.204	0.105	0.266	-0.087	-0.002	0.010	0.002	-0.016	-0.052
	[0.047]*	[0.054]***	[0.052]**	[0.057]***	[0.067]	[0.070]	[0.016]	[0.019]	[0.020]	[0.024]**
Observations	2492	2113	2353	2095	2565	2264	1191	1958	2533	2271
Mean	0.314	0.254	0.491	0.047	0.293	0.464	0.026	0.111	0.114	0.125
Standard deviation	0.782	0.874	0.805	0.854	0.990	0.986	0.159	0.314	0.318	0.331

Note: This table reports Minimum Distance estimates for the groups indicated based on equation (2.1), see text for details. A description of the outcome variables is found in Table 2.1 on page 25. Standard errors reported in brackets, clustered by county-cohort. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

reading decline with the age of the child, while the effect on behavior is increasing. At face value it seems that a better educated mother may be able to help accelerate academic achievement, an effect that is not sustained in the long run. However, the impact on behavior is sustained and possibly reinforced with time. The difference across ages for the effect on the math test is significant at the 11% level.

The results in columns (7) and (8) of Tables 2.5 and 2.6 examine grade repetition. A one year increase in mother's education reduces the probability of grade repetition by 2.8 percentage points for both age groups (IV). Child obesity is not influenced by maternal schooling at either age in the IV results. This is surprising, given the consensus that child obesity is largely affected by eating habits and physical exercise.

At the bottom of each table we report the impact of the maternal AFQT score on child outcomes. As expected and shown in other papers, the cognitive ability of the mother is a strong predictor of the cognitive ability of the child. The IV results show that the effect of mother's AFQT on child's performance in math and reading is larger at 12-14 than at 7 to 8. At ages 7 to 8, each year of maternal education produces a slightly larger increase in the math score of the child than a one standard deviation in maternal AFQT, so that (very roughly) a 4 year college degree produces the same increase in math at 7 and 8 as a 4 standard deviation increase in mother's cognition (a large effect). Equally striking is the result that mother's AFQT does not predict either child's behavior or child's grade repetition, although mother's schooling is a strong determinant of both.

These results resemble the findings of Cunha and Heckman (2006), who estimate that parental background has a strong effect on the child's cognitive skill at early ages which disappears later on, and a weaker initial effect on her non-cognitive skill which becomes stronger as the child ages. In their model, cognitive and non-cognitive skills are not equally plastic across ages and they estimate that cognitive skills are less malleable than non-cognitive skills. This result has been argued to be true in other papers (e.g., Knudsen, Heckman, Cameron, and Shonkoff (2006)). Our estimates would be consistent with such a model if we interpret maternal schooling as reflecting mostly environmental effects, and maternal cognition as being at least partly related with the heritability of cognitive ability. We would expect the environment to strongly affect child behavior at all ages, but to decrease its influence on cognition as the child grows, while the role of AFQT becomes stronger with child's age. Unless there is a strong environmental component to AFQT after controlling for maternal schooling, maternal AFQT may not be strongly related with the behavior of the child (unless cognitive and non-cognitive innate traits are positively correlated in the population[14]).

We also present estimates for four different subsamples, defined according to the gender of the child and the AFQT of the mother. We divide white mothers into two groups: white high AFQT mothers have a score above or equal to 0.4, while white low AFQT mothers have a score below 0.4. For blacks, we set the cutoff point at -0.25.[15]

When we break down the results by gender and (separately) by AFQT we find that our estimates are highest for female children and for high AFQT mothers (except for grade repetition at ages 12-14). The decline in the effect of mother's schooling on the math score can be attributed to the impact on girls, which is very strong at age 7-8 but virtually vanishes later. A similar decline can be observed for high AFQT mothers: they achieve a large improvement in the performance of their kids, but the impact vanishes by ages 12-14. In contrast, the effect on the behavioral problems index does not decline with age and the impact is substantial and significant. The lowest impact is on male children (not significant in the IV regression). The impact

[14]Heckman, Stixrud, and Urzua (2006) as well as Duckworth and Seligman (2005) argue that there is little correlation between cognitive and non-cognitive traits of children and adolescents. That is not the case in the data analyzed in Carneiro, Crawford, and Goodman (2007).

[15]This is done to account for the different distributions of AFQT between whites and blacks. There are two reasons why the effect of maternal education on child outcomes can vary across these two groups of mothers. First, this parameter can be a function of AFQT. Second, even within AFQT cells, this parameter can vary across observationally similar mothers. In that case the instrumental variables estimate will be an average of the effects of maternal education for the set of mothers affected by the instrument, and this set can be very different in the high and low AFQT groups, since AFQT and unobservable ability both determine the schooling decision of mothers. Unfortunately, our procedure confounds the two phenomena, but it is still of great interest especially if we can interpret it as (within each AFQT group) the effect of schooling for those mothers most likely to change schooling in response to a decrease in the costs of attending university (measured by our set of instrumental variables).

of mother's education on grade repetition is also persistent across ages. Overall, at ages 7-8, results are almost always stronger for mother's with high AFQT. At 12-14, however, for BPI and grade repetition the results are stronger for low AFQT mothers.

Generally, the IV results for white children are higher than the OLS ones. This may seem surprising because an ability bias intuition would tell us otherwise. However, this result is common in the returns to schooling literature (Card, 1999), and also emerges in the papers by Currie and Moretti (2003) and Oreopoulos, Page, and Stevens (2003). Part of the difference can be explained by measurement error in maternal education (Card, 1999), which could bias downwards the OLS results. Beyond these common arguments the standard intuition that is valid in the fixed coefficient model no longer applies when the impacts are heterogeneous. In this case IV estimates may well exceed OLS estimates of the effect of maternal schooling on child outcomes. On the one hand, with heterogeneous effects the OLS estimates do not have a clear direction of bias; on the other hand the IV estimates, valid only under a suitable monotonicity assumption (see Imbens and Angrist (1994)), pick up the effect on the marginal individual, which can be larger than the average effect.

A natural concern is that our instruments may be weak; we discuss this in our sensitivity analysis (section 2.4.4).

(b) Black children

It is well documented that there are large differences in the processes of human capital accumulation of blacks and whites.[16] Furthermore, ethnic differences in skill formation are an important source of concern for education policies in many countries. Therefore we compare the role of maternal education for white and black children.

Tables 2.7 and 2.8 present estimates of the effect of maternal education on outcomes for black children. Results are similar to the ones for white children, with the impacts

[16]See, e.g., Currie and Thomas (1995), Jencks and Phillips (1998), Fryer and Levitt (2004), Carneiro, Heckman, and Masterov (2005), Neal (2005), Todd and Wolpin (2006).

Table 2.7: Child outcomes – OLS results: Black children

	OLS estimates: Black children									
	PIAT math		PIAT read.		BPI		Grade repetition		Overweight	
	7-8 yrs	12-14 yrs	7-8 yrs	12-14 yrs	7-8 yrs	12-14 yrs	7-8 yrs	12-14 yrs	7-8 yrs	12-14 yrs
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Mother's schooling: All	0.074	0.075	0.062	0.079	-0.064	-0.064	-0.003	-0.032	0.009	0.015
	[0.020]***	[0.020]***	[0.019]***	[0.018]***	[0.027]**	[0.027]**	[0.009]	[0.009]***	[0.008]	[0.009]
Mother's schooling: Male child	0.068	0.082	0.065	0.095	-0.068	-0.077	0.003	-0.030	0.010	0.020
	[0.027]**	[0.027]***	[0.024]***	[0.023]***	[0.033]**	[0.031]**	[0.011]	[0.012]**	[0.010]	[0.010]*
Mother's schooling: Female child	0.078	0.070	0.060	0.060	-0.060	-0.047	-0.009	-0.034	0.009	0.004
	[0.022]***	[0.024]***	[0.023]***	[0.024]**	[0.032]*	[0.034]	[0.011]	[0.012]***	[0.011]	[0.014]
Mother's schooling: High AFQT	0.111	0.077	0.074	0.081	-0.063	-0.075	-0.004	-0.019	0.014	0.022
	[0.031]***	[0.032]**	[0.028]***	[0.032]**	[0.036]*	[0.038]**	[0.011]	[0.014]	[0.015]	[0.015]
Mother's schooling: Low AFQT	0.056	0.074	0.054	0.078	-0.064	-0.054	-0.001	-0.042	0.008	0.010
	[0.023]**	[0.025]***	[0.025]**	[0.023]***	[0.048]	[0.036]	[0.012]	[0.013]***	[0.009]	[0.012]
Mother's AFQT (corrected): All	0.298	0.326	0.345	0.254	-0.083	0.070	-0.007	-0.045	-0.025	0.036
	[0.070]***	[0.070]***	[0.069]***	[0.071]***	[0.099]	[0.086]	[0.027]	[0.029]	[0.028]	[0.033]
Observations	1271	1391	1181	1381	1233	1399	396	1168	1248	1446
Mean	-0.257	-0.402	0.066	-0.570	0.486	0.496	0.056	0.229	0.183	0.195
Standard deviation	0.833	0.857	0.806	0.845	0.985	0.993	0.229	0.420	0.387	0.396

Note: This table reports Minimum Distance estimates for the groups indicated based on equation (2.1), see text for details. A description of the outcome variables is found in Table 2.1 on page 25. Standard errors reported in brackets, clustered by county-cohort. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

Table 2.8: Child outcomes – IV results: Black children

	IV estimates: Black children									
	PIAT math		PIAT read.		BPI		Grade repetition		Overweight	
	7-8 yrs	12-14 yrs	7-8 yrs	12-14 yrs	7-8 yrs	12-14 yrs	7-8 yrs	12-14 yrs	7-8 yrs	12-14 yrs
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Mother's schooling: All	0.066	0.080	0.100	0.119	-0.067	-0.099	-0.000	-0.065	0.011	0.008
	[0.034]*	[0.028]***	[0.031]***	[0.030]***	[0.044]	[0.039]**	[0.009]	[0.016]***	[0.015]	[0.015]
Mother's schooling: Male child	0.081	0.083	0.121	0.126	-0.054	-0.106	0.004	-0.065	0.032	0.026
	[0.041]**	[0.039]**	[0.039]***	[0.043]***	[0.054]	[0.051]**	[0.011]	[0.020]***	[0.021]	[0.021]
Mother's schooling: Female child	0.051	0.077	0.075	0.113	-0.083	-0.092	-0.007	-0.064	-0.004	-0.009
	[0.042]	[0.040]*	[0.042]*	[0.041]***	[0.057]	[0.056]	[0.012]	[0.023]***	[0.018]	[0.021]
Mother's schooling: High AFQT	0.068	0.053	0.077	0.076	-0.118	-0.142	-0.004	-0.032	0.031	0.015
	[0.047]	[0.050]	[0.047]	[0.049]	[0.059]**	[0.053]***	[0.011]	[0.023]	[0.023]	[0.021]
Mother's schooling: Low AFQT	0.064	0.099	0.119	0.153	-0.003	-0.031	0.005	-0.099	-0.005	0.000
	[0.049]	[0.041]**	[0.042]***	[0.042]***	[0.066]	[0.069]	[0.013]	[0.023]***	[0.020]	[0.023]
Mother's AFQT (corrected): All	0.325	0.325	0.322	0.232	-0.091	0.095	-0.009	-0.030	-0.038	0.042
	[0.069]***	[0.070]***	[0.075]***	[0.073]***	[0.105]	[0.089]	[0.027]	[0.031]	[0.028]	[0.034]
Observations	1271	1391	1181	1381	1233	1399	396	1168	1248	1446
Mean	-0.257	-0.402	0.066	-0.570	0.486	0.496	0.056	0.229	0.183	0.195
Standard deviation	0.833	0.857	0.806	0.845	0.985	0.993	0.229	0.420	0.387	0.396

Note: This table reports Minimum Distance estimates for the groups indicated based on equation (2.1), see text for details. A description of the outcome variables is found in Table 2.1 on page 25. Standard errors reported in brackets, clustered by county-cohort. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

on math and reading, BPI, and grade repetition being large and significant, and the impact on obesity being imprecisely determined. There are, however, some differences. First, estimated impacts are stronger at 12-14 than at 7-8, and we do not observe the tendency of the math (and reading) impact to decline. Second, in the IV estimates the impact on grade repetition for 12-14 year olds is twice as large for black children than for whites, and the p-value for the difference is 5.7%. For children of low AFQT mothers, a year of education reduces the probability of grade repetition by almost 10 percentage points (which partly mirrors differences in prevalence of grade repetition). Third, maternal AFQT is a stronger predictor of child outcomes for blacks than for whites. Fourth, the role of maternal schooling is larger for males than for females.

2.4.2 Home Environments

The impact of mothers education on child outcomes is strong in a number of dimensions. Since we do not have an explicit model of child development, we cannot firmly establish the role of these channels. However, our results in this section paint a picture of how they may operate, and their detail makes them especially useful. The results for whites are reported in Table 2.9. We comment on the IV results, while in the Appendix 2.A we also report the OLS results for completeness. The maternal characteristics examined are maternal age at birth, educational aspirations for the child (does the mother believe whether the child will go to college), marital status, spouse's years of schooling (for those with a spouse), number of children, hours worked, and log family income (which includes spouse's income). All variables are measured when the child is 7 or 8.

An increase in mother's schooling by one year leads to increases in: maternal age at birth by one year, family income by 18%, the probability of being married of 4%, spouse's years of schooling by 0.5. The effect on fertility is surprisingly small.[17]

Several economists have argued that it is important to account for the effects of

[17] Note that we only have incomplete fertility and that more educated mothers delay childbirth.

Table 2.9: Family environment – IV results: White children

	IV estimates: White children (7-8 years)						
	Maternal age	Number of children	Marital status	Spouse schooling	Hours worked	Lg family income	Maternal aspirations
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mother's schooling: All	1.024	-0.017	0.041	0.549	55.633	0.177	0.048
	[0.139]***	[0.057]	[0.018]**	[0.092]***	[38.528]	[0.034]***	[0.018]***
Mother's schooling: Male child	1.074	-0.029	0.053	0.512	55.724	0.196	0.066
	[0.192]***	[0.072]	[0.021]**	[0.121]***	[45.630]	[0.046]***	[0.025]***
Mother's schooling: Female child	0.983	-0.008	0.029	0.572	55.524	0.157	0.039
	[0.176]***	[0.065]	[0.021]	[0.104]***	[48.314]	[0.047]***	[0.020]*
Mother's schooling: High AFQT	0.846	-0.107	0.045	0.486	24.112	0.177	0.057
	[0.200]***	[0.088]	[0.023]**	[0.137]***	[53.715]	[0.047]***	[0.020]***
Mother's schooling: Low AFQT	1.205	0.059	0.034	0.608	86.592	0.176	0.028
	[0.202]***	[0.080]	[0.029]	[0.132]***	[53.253]	[0.050]***	[0.030]
Mother's AFQT (corrected): All	-0.247	0.079	0.015	0.061	148.174	0.191	0.011
	[0.218]	[0.099]	[0.029]	[0.160]	[59.570]**	[0.056]***	[0.039]
Observations	4395	4395	4391	3335	4307	3796	1235
Mean	24.282	2.752	0.770	13.231	1152.305	10.361	0.764
Standard deviation	4.632	1.195	0.421	2.490	950.919	0.970	0.425

Note: This table reports Minimum Distance estimates for the groups indicated based on equation (2.1), see text for details. A description of the outcome variables is found in Table 2.1 on page 25. Standard errors reported in brackets, clustered by county-cohort. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

assortative mating because the causal effect of maternal education on child performance may come through her ability to find an educated father for the child. They also argue that maternal education can have ambiguous effects because if on one hand the child benefits from better home environments and perhaps richer investments, she will benefit of less maternal time because more educated mothers spend more time in the labor market. Two examples are Behrman and Rosenzweig (2002) and Plug (2004), who estimate small or no effects of maternal education on child's schooling, while father's education has large and strong effects on this outcome. Unfortunately we do not have good instruments for either of these variables and cannot directly assess the validity of these arguments. However, we can examine the effect of maternal schooling on spouse's schooling and on maternal labor supply.

As pointed out above, column (4) shows that an increase of one year in maternal education leads to an increase of 0.5 years of spouse's education. If we attributed all the effects of maternal education to assortative mating we would need father's schooling to have almost twice as large effects as the ones we estimate for mothers. Therefore, assortative mating effects are unlikely to fully drive our results. Column (5) looks at the effects of maternal education on maternal employment measured in terms of annual hours worked. Annual hours worked increase by 56 hours per additional year of maternal schooling (5% of the mean of 1,152 hours worked per year), or roughly 1.5 weeks of full-time work per year, although the effect is imprecisely estimated. If we compared a mother with a college degree and another without, our estimates suggest that the former would work 6 more weeks per year than the latter. Cumulating over several years of childhood, these will translate into much more family resources for the mother with a college degree, but less time at home. The latter can have an offsetting effect on the former, although it depends on what kind of substitutes educated mothers can find for their time with their child.

Column (7) shows that more educated mothers are 5 percentage points more likely to believe that their offspring will complete college. These expectations may translate into different behavior on the side of the mother and the child.

The estimates presented in Table 2.9 are fairly similar for boys and girls, and for children of mothers with high and low levels of AFQT. There are only a few cases of interesting differences across groups. In particular, the effect of maternal education on maternal aspirations and marital status are small for low AFQT mothers, which may be the reason why we found weak effects on child outcomes for this group of mothers.

One feature of the data set we use is the wealth of information on direct measures of home environments and parental investments, as reported in Table 2.10. For white children, an increase in mother's schooling by one year leads to increases in the probabilities that: there is a musical instrument in the home by 5.4%; there is a computer in the home by 5.7%; a child takes special lessons by 6.2%. Each extra year of schooling also means that mothers are 4.5% more likely to read to their child at least three times a week. There is no evidence that maternal education affects the amount of newspapers in the home, adult supervision out of school, and time spent with the child in a museum or sharing meals. Notice that more educated mothers do not seem to spend less time in activities with their children, even though they spend more time working.

Table 2.10: Investments – IV results: White children

	IV estimates: White children							
	Museum 7-8 yrs	Musical Instr. 7-8 yrs	Special lesson 7-8 yrs	Mother reads 7-8 yrs	Newspaper 7-8 yrs	Computer 12-14 yrs	Adult home 12-14 yrs	Joint meals 12-14 yrs
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mother's schooling: All	0.023	0.054	0.062	0.045	-0.006	0.057	0.018	-0.008
	[0.019]	[0.020]***	[0.017]***	[0.018]**	[0.021]	[0.016]***	[0.020]	[0.021]
Mother's schooling: Male child	0.045	0.075	0.100	0.064	-0.003	0.048	0.034	-0.007
	[0.028]	[0.026]***	[0.024]***	[0.025]**	[0.028]	[0.022]**	[0.026]	[0.028]
Mother's schooling: Female child	0.007	0.037	0.032	0.030	-0.007	0.064	0.004	-0.009
	[0.024]	[0.025]	[0.021]	[0.022]	[0.024]	[0.020]***	[0.025]	[0.026]
Mother's schooling: High AFQT	0.017	0.067	0.054	0.047	-0.008	0.047	0.008	-0.028
	[0.027]	[0.028]**	[0.021]**	[0.025]*	[0.027]	[0.020]**	[0.026]	[0.027]
Mother's schooling: Low AFQT	0.029	0.040	0.079	0.042	-0.002	0.074	0.030	0.015
	[0.028]	[0.029]	[0.030]***	[0.027]	[0.032]	[0.027]***	[0.027]	[0.029]
Mother's AFQT (corrected): All	-0.015	0.021	-0.002	-0.022	0.045	0.025	-0.067	-0.015
	[0.030]	[0.036]	[0.030]	[0.032]	[0.033]	[0.034]	[0.036]*	[0.037]
Observations	2646	2644	2643	2649	2646	1681	2036	2292
Mean	0.424	0.513	0.682	0.492	0.526	0.681	0.671	0.565
Standard deviation	0.494	0.500	0.466	0.500	0.499	0.466	0.470	0.496

Note: This table reports Minimum Distance estimates for the groups indicated based on equation (2.1), see text for details. A description of the outcome variables is found in Table 2.1 on page 25. Standard errors reported in brackets, clustered by county-cohort. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

This pattern emerges throughout this chapter, even much more strongly than here, and we will comment on it with detail when we examine the child's early years.

The results for black mothers are slightly different, and they are shown in Tables 2.11 and 2.12. Relatively to white mothers, education not only affects maternal age at

Table 2.11: Family environment – IV results: Black children

	IV estimates: Black children (7-8 years)						
	Maternal age	Number of children	Marital status	Spouse schooling	Hours worked	Lg family income	Maternal aspirations
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mother's schooling: All	0.896	-0.304	0.061	0.529	182.163	0.190	0.047
	[0.147]***	[0.063]***	[0.020]***	[0.079]***	[33.790]***	[0.033]***	[0.019]**
Mother's schooling: Male child	0.929	-0.326	0.073	0.484	220.602	0.238	0.046
	[0.200]***	[0.079]***	[0.024]***	[0.096]***	[52.013]***	[0.041]***	[0.025]*
Mother's schooling: Female child	0.867	-0.287	0.049	0.564	161.719	0.133	0.048
	[0.187]***	[0.073]***	[0.024]**	[0.089]***	[39.800]***	[0.043]***	[0.025]*
Mother's schooling: High AFQT	0.841	-0.257	0.059	0.484	138.268	0.257	0.036
	[0.225]***	[0.089]***	[0.031]*	[0.130]***	[46.324]***	[0.051]***	[0.028]
Mother's schooling: Low AFQT	0.937	-0.347	0.062	0.559	233.002	0.144	0.054
	[0.195]***	[0.085]***	[0.024]**	[0.105]***	[49.888]***	[0.042]***	[0.023]**
Mother's AFQT (corrected): All	-0.096	0.089	0.077	0.032	131.007	0.197	0.107
	[0.286]	[0.112]	[0.042]*	[0.227]	[79.503]*	[0.077]**	[0.063]*
Observations	2647	2647	2646	943	2624	2129	422
Mean	22.070	3.097	0.375	12.688	1139.074	9.638	0.656
Standard deviation	4.489	1.413	0.484	2.095	991.853	0.930	0.475

Note: This table reports Minimum Distance estimates for the groups indicated based on equation (2.1), see text for details. A description of the outcome variables is found in Table 2.1 on page 25. Standard errors reported in brackets, clustered by county-cohort. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

birth, aspirations, marital status, spouse's schooling and income, but it also has large effects on fertility and employment. Each additional four years in school (a four year university degree) decreases the number of children born to each woman by 1.2, and increase maternal employment by over 730 hours (or roughly 18 weeks) per year. The effects of education on income are especially large for high AFQT mothers, while the effects of education on employment and fertility are stronger for low AFQT mothers.

It is remarkable that each year of maternal schooling among blacks increases the proportion of children going to a museum at least several times per year by 3.2%, and the proportion of children who are read to at least three times a week by 5.4% (these are time intensive activities). Part of this may be due to the fact that more educated black mothers have less children to spend their time with. However, an extra year of maternal education also makes it 5.1% less likely that black children have adult

Table 2.12: Investments – IV results: Black children

	IV estimates: Black children							
	Museum 7-8 yrs	Musical Instr. 7-8 yrs	Special lesson 7-8 yrs	Mother reads 7-8 yrs	Newspaper 7-8 yrs	Computer 12-14 yrs	Adult home 12-14 yrs	Joint meals 12-14 yrs
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mother's schooling: All	0.032	0.017	0.101	0.054	-0.013	0.065	-0.051	0.014
	[0.019]*	[0.020]	[0.020]***	[0.018]***	[0.019]	[0.019]***	[0.017]***	[0.018]
Mother's schooling: Male child	0.021	-0.005	0.087	0.046	-0.014	0.053	-0.057	0.020
	[0.025]	[0.029]	[0.027]***	[0.023]**	[0.026]	[0.026]**	[0.024]**	[0.023]
Mother's schooling: Female child	0.044	0.034	0.112	0.064	-0.012	0.076	-0.047	0.006
	[0.027]	[0.025]	[0.024]***	[0.024]***	[0.027]	[0.026]***	[0.022]**	[0.025]
Mother's schooling: High AFQT	0.016	0.012	0.136	0.057	-0.014	0.088	-0.042	0.033
	[0.030]	[0.034]	[0.027]***	[0.025]**	[0.027]	[0.029]***	[0.026]*	[0.027]
Mother's schooling: Low AFQT	0.042	0.020	0.062	0.052	-0.012	0.047	-0.059	-0.006
	[0.024]*	[0.026]	[0.028]**	[0.024]**	[0.028]	[0.026]*	[0.023]**	[0.027]
Mother's AFQT (corrected): All	0.002	0.026	0.004	-0.070	0.101	0.082	0.024	-0.075
	[0.043]	[0.046]	[0.043]	[0.045]	[0.048]**	[0.049]*	[0.043]	[0.048]
Observations	1306	1305	1304	1308	1306	906	1306	1431
Mean	0.405	0.336	0.447	0.320	0.419	0.352	0.694	0.316
Standard deviation	0.491	0.472	0.497	0.467	0.494	0.478	0.461	0.465

Note: This table reports Minimum Distance estimates for the groups indicated based on equation (2.1), see text for details. A description of the outcome variables is found in Table 2.1 on page 25. Standard errors reported in brackets, clustered by county-cohort. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

supervision when they arrive home after school, which can have detrimental effects on their behavior (Aizer, 2004). This problem is worse for males than for females. The fact that the effects of maternal education on child outcomes are not only strong, but they are especially strong for black males, shows that mothers are able to overcome the problem of low adult supervision through other means. When we examine the remaining home environment variables, we only find statistically significant effects of the presence of a computer in the home and enrolment in special lessons.

In summary, there exists strong evidence that maternal education affects home environments and child outcomes. The size of several of our estimates in this section is large, and suggests that we should seriously look at education policy as a way of improving the home environments of future generations of children. Educated mothers provide better surroundings for their children by postponing and decreasing childbearing, by increasing family resources, and by assortative mating. There is also strong evidence that educated mothers invest more in their children. However, educated mothers also spend longer periods outside the home working and earning. Still, whatever the negative consequences of spending time away from the children may be, they are outweighed by the positive effects. With the exception of adult supervision for black children, more educated mothers do not spend less time with their children, either because they have less children, or less leisure time. If anything, our results indicate that the opposite is true.

At this point it is useful to compare our estimates of the effect of maternal education to those of other childhood interventions. The large class size reduction of the STAR experiment (a reduction from 22 to 15 pupils per class, studied by Krueger (1999)) yielded test score gains of 0.2 standard deviations, an equivalent of two years of maternal schooling. Dahl and Lochner (2006) estimate that a \$1,000 increase in family income improves performance on the math test score by 2.1% of a standard deviation (3.6% for reading). Using mother fixed effects, Currie and Thomas (1995) estimate that participation in Head Start increases performance in the PPVT vocabulary test by almost 6 percentile points (which is about 20 to 25% of a standard deviation). Bernal and Keane (2006) find that additional formal child care does not improve the average child test score performance, but may be beneficial for children of poorly educated mothers. Aizer (2004) estimates that adult supervision after school reduces the probability of a child engaging in risky behavior by about 7 percentage points. Dustmann and Schönberg (2007) find that increasing paid maternity leave does not significantly improve long-term child outcomes. Our claim is that, although the nature of the different interventions differs quite a lot, the effects of maternal education are large when compared to those of other interventions. If the objective is to increase children's outcomes, additional maternal education may be a serious competitor to the other types of interventions. Of course, in doing this kind of comparison, it is important to keep in mind that each of the interventions have different costs and may affect children along a variety of dimensions, and comparisons become difficult when trade-offs between different objectives are involved.

2.4.3 Early Childhood and Young Adulthood

In this section we investigate two issues. First, which of these effects are visible at earlier ages of the child? This question is particularly interesting given the recent academic and policy emphasis on the importance of the early years. Second, is there any evidence of effects of maternal schooling on environments and behavior during adolescence and young adulthood, when behavioral anomalies such as engagement in criminal activities, early dropping out of school, or early child bearing, may be the source of long run problems? Ideally, we would like to follow individuals well into their adult lives, but unfortunately this is not yet possible with this sample.

(a) Early Childhood

Here we present estimates of the effect of maternal schooling on the probability of the child having low birth-weight (weighing less than 5.5 pounds at birth), and the score on the MSD scale, which assesses the motor and social skills development, both for children up to 24 months. Results are shown for whites and blacks in Table 2.13.

Table 2.13: Early outcomes – IV results

	IV estimates: Children 0-1 years			
	Whites		Blacks	
	Low birthweight	MSD	Low birthweight	MSD
	(1)	(2)	(3)	(4)
Mother's schooling: All	-0.004	-0.076	-0.012	0.084
	[0.007]	[0.035]**	[0.013]	[0.049]*
Mother's schooling: Male child	-0.006	-0.080	-0.010	0.060
	[0.010]	[0.045]*	[0.016]	[0.056]
Mother's schooling: Female child	-0.003	-0.072	-0.016	0.138
	[0.011]	[0.047]	[0.020]	[0.079]*
Mother's schooling: High AFQT	-0.010	-0.054	0.008	0.013
	[0.010]	[0.043]	[0.017]	[0.065]
Mother's schooling: Low AFQT	0.002	-0.120	-0.036	0.157
	[0.011]	[0.061]**	[0.018]**	[0.066]**
Mother's AFQT (corrected): All	-0.008	0.025	-0.000	-0.242
	[0.013]	[0.071]	[0.025]	[0.137]*
Observations	5580	2136	2806	781
Mean	0.065	-0.039	0.130	0.184
Standard deviation	0.246	0.994	0.337	1.216

Note: This table reports Minimum Distance estimates for the groups indicated based on equation (2.1), see text for details. A description of the outcome variables is found in Table 2.1 on page 25. Standard errors reported in brackets, clustered by county-cohort. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

Currie and Moretti (2003) find that one extra year of maternal education reduces the probability that a child is born with low birth-weight by 1 percentage point. Our estimates for whites are lower and insignificant, whether we use OLS or IV, although we have a much smaller sample than Currie and Moretti (2003). Results are only statistically strong for black mothers with low AFQT scores, for whom the coefficient is -0.036 (the incidence of low birth-weight is of 14.9% for this group).

Looking at the relationship between maternal education and early motor and social skills of the child a new picture emerges. For whites, our estimates are small but negative, especially for low ability mothers. This is the first and only instance where increases in maternal schooling may not be good for their children, perhaps because of increased maternal employment and less time with the child.

Table 2.14 presents the results for early home environments of whites, where the following outcomes are considered: smoking in the year prior to the birth of the child,

Table 2.14: Early channels – IV results: white children

	IV estimates: White children 0-1 years							
	Smoking d. pregnancy	Weeks breastfeeding	Formal child care	Hours worked	Mother reads	Book	Soft toys	Outings
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mother's schooling: All	-0.069	2.307	0.013	102.498	0.006	0.071	-0.198	-0.005
	[0.016]***	[0.710]***	[0.007]*	[29.598]***	[0.014]	[0.030]**	[0.421]	[0.016]
Mother's schooling: Male child	-0.064	1.976	0.004	121.941	0.001	0.063	-0.374	-0.011
	[0.021]***	[0.941]**	[0.010]	[40.272]***	[0.020]	[0.045]	[0.525]	[0.022]
Mother's schooling: Female child	-0.074	2.717	0.022	86.187	0.011	0.077	0.059	0.000
	[0.022]***	[1.043]***	[0.010]**	[37.428]**	[0.021]	[0.042]*	[0.624]	[0.022]
Mother's schooling: High AFQT	-0.062	1.059	0.016	109.035	-0.008	0.026	-0.284	0.001
	[0.020]***	[0.968]	[0.011]	[39.213]***	[0.019]	[0.040]	[0.559]	[0.020]
Mother's schooling: Low AFQT	-0.081	3.801	0.011	93.057	0.027	0.139	-0.075	-0.017
	[0.028]***	[1.061]***	[0.009]	[47.500]*	[0.024]	[0.051]***	[0.677]	[0.029]
Mother's AFQT (corrected): All	-0.065	0.763	0.020	81.880	0.053	0.136	2.507	0.021
	[0.031]**	[1.364]	[0.010]*	[42.710]*	[0.030]*	[0.062]**	[0.775]***	[0.030]
Observations	2293	2220	4850	5942	2358	2382	2343	2380
Mean	0.287	15.370	0.066	926.749	0.607	3.240	16.654	0.691
Standard deviation	0.452	22.126	0.248	880.676	0.489	1.062	12.456	0.462

Note: This table reports Minimum Distance estimates for the groups indicated based on equation (2.1), see text for details. A description of the outcome variables is found in Table 2.1 on page 25. Standard errors reported in brackets, clustered by county-cohort. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

weeks of breastfeeding, use of formal child care arrangements, annual hours worked by the mother, whether the child is read to, how many books and soft toys the child has, and whether the child is taken on outings regularly.

The two health inputs, (not) smoking and breastfeeding, are strongly affected by maternal schooling. Notice also that the effect on maternal hours worked is much larger when measured during the child's early years than later on (as we saw in Table 2.9). At the same time, the increase in formal child care is modest and only statistically strong for girls. The strong increase in hours worked that results from additional education is not accompanied by a strong increase in formal child-care, raising the question of how these children are cared for. This could be seen as support to the argument that more educated mothers spend more time working, with detrimental effects on child development. Still, even if this is true, children seem to recover, so that BPI and grade repetition at 12 and 14 are lower when maternal education is higher. Finally, there is no evidence that, even though they work more, more educated mothers spend less time breastfeeding, reading to their children, or taking them on outings. This is consistent with recent findings from time diary studies summarized in Blau and Currie (2003):

Table 2.15: Early channels – IV results: Black children

	IV estimates: Black children 0-1 years							
	Smoking d. pregnancy	Weeks breastfeeding	Formal child care	Hours worked	Mother reads	Book	Soft toys	Outings
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mother's schooling: All	-0.026	1.422	0.019	194.011	0.050	0.130	-0.115	0.002
	[0.021]	[0.626]**	[0.008]**	[28.539]***	[0.022]**	[0.052]**	[0.456]	[0.019]
Mother's schooling: Male child	-0.005	1.223	0.017	183.880	0.063	0.150	0.395	-0.002
	[0.026]	[0.749]	[0.010]*	[36.948]***	[0.026]**	[0.060]**	[0.584]	[0.026]
Mother's schooling: Female child	-0.048	1.717	0.023	205.266	0.030	0.099	-0.415	0.005
	[0.026]*	[0.871]**	[0.012]*	[38.655]***	[0.031]	[0.070]	[0.504]	[0.026]
Mother's schooling: High AFQT	-0.034	-0.148	0.035	180.036	0.057	0.092	-0.125	-0.025
	[0.027]	[1.014]	[0.015]**	[39.661]***	[0.032]*	[0.067]	[0.527]	[0.023]
Mother's schooling: Low AFQT	-0.017	1.966	0.014	210.150	0.044	0.166	-0.097	0.044
	[0.029]	[0.684]***	[0.009]	[42.733]***	[0.029]	[0.065]**	[0.642]	[0.028]
Mother's AFQT (corrected): All	0.024	0.680	0.009	140.934	0.013	0.208	-1.184	0.024
	[0.048]	[1.249]	[0.015]	[61.143]**	[0.049]	[0.100]**	[0.950]	[0.042]
Observations	861	855	2257	2965	894	897	889	897
Mean	0.278	5.513	0.070	767.310	0.371	2.337	11.227	0.661
Standard deviation	0.448	13.905	0.254	885.509	0.483	1.190	10.086	0.474

Note: This table reports Minimum Distance estimates for the groups indicated based on equation (2.1), see text for details. A description of the outcome variables is found in Table 2.1 on page 25. Standard errors reported in brackets, clustered by county-cohort. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

mothers who work more do not spend less time with their children; instead, they have less leisure. It is also consistent with the analysis of (large) changes in maternity leave laws in Germany by Dustmann and Schönberg (2007) who find no positive effect on child outcomes. Notice also that young children of educated mothers have more books than other children, especially if their mothers have low cognitive ability.

In summary, it is difficult to make the case that the large increase in employment of white mothers that results from additional education has detrimental effects on children. There may be some delays in their motor and social development, especially for low AFQT mothers, but they do not appear to have any long term undesirable consequences. In fact, it is for low AFQT mothers that maternal education has the largest positive effects on home environments.

For black families this picture is even more evident. The main results are shown in column (3) and (4) of Table 2.13. The impacts of maternal education on birth-weight and motor and social development are positive and large, especially for low ability mothers. An additional year of education leads to about 200 extra hours of work, but also more regular use of formal child care arrangements, prolonged breastfeeding, more time reading to the child, and more children's books in the home (Table 2.15).

The estimates displayed in Tables 2.14 and 2.15 tell a clear and important story: improvements in maternal schooling promote much better home environments during the early years of the child; although more educated mothers work more, they do not spend less quality time with their children, and if anything the opposite is true; it is striking that for many outcomes, for both black and white mothers, it is for low ability mother that education has the largest impact on early home environments.

(b) Young Adulthood

Finally, we examine engagement in some risky behaviors in late adolescence: early dropping out of school, early childbearing, and criminal activity. It is important to keep in mind that many children of the NLSY79 cohort members have not yet reached adulthood. Thus, the children we observe in this age range are mainly from the early cohorts and from mothers with very low birth ages, and the sample size is smaller than for the younger cohorts. Still, at the very least, the following demonstrates that the effect of maternal education follows the children into adulthood.

Table 2.16: Young adults – IV results

	IV estimates: Young adults (18-19 years)					
	White			Black		
	Enrollment	Conviction	Own children	Enrollment	Conviction	Own children
	(1)	(2)	(3)	(4)	(5)	(6)
Mother's schooling: All	0.031	-0.002	-0.045	0.010	-0.018	-0.036
	[0.021]	[0.014]	[0.014]***	[0.021]	[0.013]	[0.014]**
Mother's schooling: Male young adult	0.033	-0.004	-0.047	0.005	-0.039	-0.017
	[0.032]	[0.023]	[0.020]**	[0.031]	[0.020]*	[0.018]
Mother's schooling: Female young adult	0.030	-0.001	-0.043	0.016	-0.005	-0.070
	[0.026]	[0.017]	[0.021]**	[0.030]	[0.016]	[0.024]***
Mother's schooling: High AFQT	0.033	-0.012	-0.052	-0.017	-0.040	-0.036
	[0.030]	[0.020]	[0.020]***	[0.033]	[0.018]**	[0.017]**
Mother's schooling: Low AFQT	0.030	0.007	-0.037	0.034	0.007	-0.037
	[0.029]	[0.019]	[0.020]*	[0.030]	[0.020]	[0.023]
Mother's AFQT (corrected): All	0.042	-0.046	0.010	-0.068	-0.049	0.000
	[0.052]	[0.031]	[0.026]	[0.058]	[0.025]*	[0.043]
Observations	935	1047	816	742	889	612
Mean	0.624	0.154	0.091	0.627	0.124	0.157
Standard deviation	0.485	0.361	0.296	0.484	0.329	0.398

Note: This table reports Minimum Distance estimates for the groups indicated based on equation (2.1), see text for details. A description of the outcome variables is found in Table 2.1 on page 25. Standard errors reported in brackets, clustered by county-cohort. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

Table 2.16 present estimates of the effect of maternal schooling on several outcomes: a dummy for school enrollment, a dummy for convictions, and the number of own children, all at ages 18 and 19. Among whites, we only observe strong effects on fertility. For blacks, the decrease in the conviction rate is notable for boys and children of high ability mothers, and so is the decrease in fertility, especially for girls.

2.4.4 Sensitivity Analysis

In this section we examine the sensitivity of our main results, presented in section 2.4.1 above. An important concern in this work is with the potential weakness of the instruments (although the p-values of the instruments in the first stage equations are very low). Most of the literature on weak instruments deals with models of fixed coefficients. In such cases, one standard recommendation is to estimate the model using LIML instead of two stage least squares, as we have done so far (e.g., Staiger and Stock (1997)). Therefore, we proceed by estimating the model by LIML. Here we present results for the main outcomes for the sample of white children. Panel B in Table 2.17 shows that, at ages 7-8, the LIML estimates are of the same sign than the original two stage least squares (TSLS) estimates in the chapter, but they have larger absolute magnitudes and they are more imprecise (which would be a prediction of most of the literature).[18] This means that the TSLS estimates are closer to OLS than LIML, which is what we would expect if the instruments were weak. Notice also that, even with the imprecise LIML estimates, the effect of maternal schooling on white children cognitive development drops substantially from ages 7-8 to ages 10-12, while that is not the case for grade repetition and BPI.

These results suggest that, although we may suffer from a weak instruments problem, if anything our estimates understate the true impact of maternal education on child outcomes since TSLS is biased towards OLS (and the latter are generally smaller than the former in absolute value). However, we need to be cautious about conclusions

[18]Panel A reproduces our base case result for easy reference.

Table 2.17: Sensitivity (white children)

	Sensitivity analysis (white children)									
	PIAT math		PIAT read.		BPI		Grade repetition		Overweight	
	7-8 yrs	12-14 yrs	7-8 yrs	12-14 yrs	7-8 yrs	12-14 yrs	7-8 yrs	12-14 yrs	7-8 yrs	12-14 yrs
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Base case										
IV – Base case	0.097	0.024	0.075	0.018	-0.092	-0.116	-0.028	-0.028	-0.015	-0.012
	[0.031]***	[0.033]	[0.033]**	[0.033]	[0.043]**	[0.041]***	[0.008]***	[0.011]**	[0.013]	[0.013]
	2492	2113	2353	2095	2565	2264	1191	1958	2533	2271
Panel B: LIML										
IV – LIML	0.140	-0.009	0.103	-0.009	-0.117	-0.247	-0.046	-0.031	-0.031	-0.017
	[0.069]**	[0.079]	[0.065]	[0.085]	[0.105]	[0.158]	[0.026]*	[0.040]	[0.032]	[0.024]
	2492	2113	2353	2095	2565	2264	1191	1958	2533	2271
Panel C: Including additional controls										
including polynomials	0.102	0.047	0.080	0.046	-0.093	-0.128	-0.025	-0.030	-0.018	-0.016
and interactions	[0.031]***	[0.034]	[0.033]**	[0.034]	[0.043]**	[0.042]***	[0.009]***	[0.011]***	[0.013]	[0.013]
	2492	2113	2353	2095	2565	2264	1191	1958	2533	2271
including group-specific	0.086	0.018	0.040	0.014	-0.094	-0.112	-0.022	-0.018	-0.012	-0.013
cohort dummies	[0.031]***	[0.034]	[0.033]	[0.032]	[0.044]**	[0.038]***	[0.008]***	[0.012]	[0.013]	[0.012]
	2492	2113	2353	2095	2565	2264	1191	1958	2533	2271
Panel D: Varying the set of instruments										
IV – Excluding distance variable	0.107	0.033	0.109	0.023	-0.104	-0.126	-0.032	-0.021	-0.015	-0.008
	[0.035]***	[0.038]	[0.038]***	[0.035]	[0.045]**	[0.043]***	[0.010]***	[0.012]*	[0.013]	[0.015]
	2492	2113	2353	2095	2565	2264	1191	1958	2533	2271
IV – Excluding distance variable	0.116	-0.006	0.098	0.038	-0.049	-0.081	-0.021	-0.023	-0.022	-0.012
and tuition variable	[0.041]***	[0.044]	[0.045]**	[0.042]	[0.062]	[0.056]	[0.011]**	[0.014]*	[0.017]	[0.018]
	2492	2113	2353	2095	2565	2264	1191	1958	2533	2271

Note: This table reports IV estimates, showing the estimated average effect across all groups using the MD procedure as before. Panel A reproduces the main results for easy reference. Panel B shows LIML estimates. Panel C adds additional controls. 'Polynomials and interactions' includes polynomials of AFQT and grandparents' education, and two-way interactions between AFQT, grandparents' education, and broken home status. All of these additional regressors are also interacted with the four group indicators. 'Group specific cohort dummies' adds interactions of cohort indicators with group indicators to the base specification. Panel D presents IV estimates based on a subset of the instruments, where distance (and corresponding interactions) and distance and tuition (and corresponding interactions) are, respectively, excluded from the analysis. See text for details. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

from these results, since the literature on weak instruments we draw on refers to a fixed coefficient model.

Another possible criticism of our procedure is that, since we are relying on interactions between controls and instruments, if the outcome equation is misspecified then some of our results might be driven by nonlinearities instead of genuine variation in the instruments. Therefore we re-estimate our model with a more flexible specification of the outcome equations, where we add the following variables to the set of controls: AFQT squared, grandmother's education squared, grandfather's education squared, and all two-way interactions between AFQT, grandmother's education, grandfather's education and whether the mother lived in a broken home at age 14. These additional controls are also interacted with the four group indicators. The IV estimates of the coefficient on maternal schooling are presented in the first row of Panel C of Table 2.17. The results are virtually unchanged by this additional set of controls.

All of our results presented included cohort fixed effects. Another specification check is reported in the second row of Panel C, in which we address the possible concern that the four subgroups of interest may follow group-specific trends, by including *group-specific* cohort indicators. Results are essentially unchanged except for PIAT reading at 7-8 and grade repetition at 12-14. Panel D shows results where we vary the set of instruments we use. We show results where we exclude the distance variable and the corresponding interactions, and then both distance and tuition (and corresponding interactions), so that the results rely only on opportunity cost variables. This kind of experiment is interesting as different instruments may affect different subgroups, and this approach has been used to compare returns for different groups (Cameron and Taber, 2004). There is of course a loss of efficiency connected to excluding some of the instruments, so the precision of these estimates is somewhat lower. The return in terms of PIAT scores for ages 7-8 goes up. When we exclude tuition as well, the BPI coefficient goes down and becomes insignificant. But overall, the results are very similar to the base case.[19]

2.5 Summary and Conclusion

In this chapter we study the effect of maternal education on their children's outcomes, including cognitive development as measured by test score performance, behavioral problems, grade repetition, and health outcomes. We also examine home environments and parental investments. We instrument maternal schooling with local tuition fees, distance to college, and local labor market variables. In the outcome equations we condition on county and time effects, thus removing the impact of permanent differences and aggregate trends. We obtain additional variation in the instruments by allowing the effect to vary with family background of the mother.

Our results show that mother's education increases the child's performance in both math and reading at ages 7-8, but these effects are not seen at ages 12-14. Maternal education also reduces the incidence of behavioral problems and reduces grade repetition, but we find no effect on obesity. More educated mothers delay childbearing, are more likely to be married, have substantially better educated spouses and higher family income. They are more likely to invest in their children through books, providing musical instruments, special lessons, or availability of a computer. Even though they work more, more educated mothers do not spend less time breastfeeding, reading to their children or taking them on outings. Finally, the effect of maternal education persists into adolescence, reducing the number of children born to the young adults at ages 18-19, and the number of criminal convictions for blacks.

A policy implication is that intergenerational transmission is important for understanding long term policy effectiveness. This is important because many programmes are struggling to improve outcomes for poor children. Programmes which manage to increase mothers schooling are likely to be important not only for mothers now but

[19]We should also mention that we have estimated more parsimonious models where we include state fixed effects instead of county fixed effects, which resulted in similar estimates to the ones we present.

also for their future children, and should be designed and judged with this in mind.

Our interest in understanding the effect of parental education on children's human capital is closely related to the study of intergenerational mobility. Solon (1999) points out that the high correlation between parental income and their offspring's income is well-documented, but that the underlying causes are not very well understood. Our findings suggest that parental educational choices may be an important transmission channel of intergenerational inequality. They imply that an additional year of parental education increases a child's test score performance by about 0.1 of a standard deviation. If a one standard deviation difference in age 7 test scores translates into wage increases of around 4% (Carneiro, Crawford, and Goodman (2007)), then the change in child's earnings due to the additional year of parental education is about 0.4%. If an additional year of parental education increases parental earnings by say 10% (Card, 1999), this mechanism implies that a one percent change in parental income is associated with about a 0.04 percent change in children's earnings. Comparing this to an empirical long-run elasticity between parental and children's earnings of around 0.4 (Solon (1999)), it becomes clear that parental education plays an important role in transmitting inequality. Of course, this is only a rough calculation. Still, it implies that parental education accounts for a substantive part of the intergenerational correlation in earnings, and it supports the view that educational policy can influence intergenerational mobility.

2.A Appendix

Table 2.18: Family environment – OLS results: White children

	OLS estimates: White children (7-8 years)						
	Maternal age	Number of children	Marital status	Spouse schooling	Hours worked	Lg family income	Maternal aspirations
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mother's schooling: All	0.984	-0.037	0.016	0.533	71.027	0.152	0.047
	[0.053]***	[0.024]	[0.006]***	[0.043]***	[13.050]***	[0.013]***	[0.008]***
Mother's schooling: Male child	0.936	-0.040	0.019	0.557	74.430	0.156	0.046
	[0.064]***	[0.026]	[0.007]***	[0.049]***	[16.365]***	[0.016]***	[0.011]***
Mother's schooling: Female child	1.041	-0.033	0.012	0.509	67.913	0.149	0.048
	[0.069]***	[0.027]	[0.007]	[0.050]***	[15.874]***	[0.015]***	[0.011]***
Mother's schooling: High AFQT	0.959	-0.025	0.015	0.548	53.733	0.155	0.045
	[0.070]***	[0.030]	[0.007]**	[0.059]***	[18.347]***	[0.017]***	[0.010]***
Mother's schooling: Low AFQT	1.016	-0.057	0.017	0.517	91.878	0.148	0.050
	[0.080]***	[0.039]	[0.011]	[0.063]***	[20.300]***	[0.019]***	[0.014]***
Mother's AFQT (corrected): All	-0.231	0.080	0.032	0.043	130.226	0.202	0.003
	[0.183]	[0.089]	[0.023]	[0.146]	[50.983]**	[0.049]***	[0.035]
Observations	4395	4395	4391	3335	4307	3796	1235
Mean	24.282	2.752	0.770	13.231	1152.305	10.361	0.764
Standard deviation	4.632	1.195	0.421	2.490	950.919	0.970	0.425

Note: This table reports Minimum Distance estimates for the groups indicated based on equation (2.1), see text for details. A description of the outcome variables is found in Table 2.1 on page 25. Standard errors reported in brackets, clustered by county-cohort. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

Table 2.19: Family environment – OLS results: Black children

	OLS estimates: Black children (7-8 years)						
	Maternal age	Number of children	Marital status	Spouse schooling	Hours worked	Lg family income	Maternal aspirations
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mother's schooling: All	1.025	-0.190	0.034	0.460	175.560	0.156	0.071
	[0.078]***	[0.032]***	[0.011]***	[0.073]***	[17.140]***	[0.016]***	[0.016]***
Mother's schooling: Male child	1.106	-0.210	0.048	0.446	178.186	0.169	0.081
	[0.098]***	[0.040]***	[0.013]***	[0.084]***	[19.795]***	[0.017]***	[0.018]***
Mother's schooling: Female child	0.941	-0.181	0.023	0.468	171.745	0.122	0.053
	[0.100]***	[0.034]***	[0.012]*	[0.076]***	[22.375]***	[0.022]***	[0.023]**
Mother's schooling: High AFQT	1.154	-0.158	0.021	0.386	93.849	0.163	0.039
	[0.129]***	[0.051]***	[0.019]	[0.115]***	[29.230]***	[0.032]***	[0.027]
Mother's schooling: Low AFQT	0.942	-0.212	0.040	0.510	215.690	0.154	0.083
	[0.102]***	[0.042]***	[0.013]***	[0.094]***	[20.712]***	[0.018]***	[0.018]***
Mother's AFQT (corrected): All	-0.148	0.017	0.095	0.086	154.825	0.208	0.100
	[0.272]	[0.108]	[0.041]**	[0.230]	[77.030]**	[0.073]***	[0.062]
Observations	2647	2647	2646	943	2624	2129	422
Mean	22.070	3.097	0.375	12.688	1139.074	9.638	0.656
Standard deviation	4.489	1.413	0.484	2.095	991.853	0.930	0.475

Note: This table reports Minimum Distance estimates for the groups indicated based on equation (2.1), see text for details. A description of the outcome variables is found in Table 2.1 on page 25. Standard errors reported in brackets, clustered by county-cohort. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

Table 2.20: Investments – OLS results: White children

	OLS estimates: White children							
	Museum 7-8 yrs	Musical Instr. 7-8 yrs	Special lesson 7-8 yrs	Mother reads 7-8 yrs	Newspaper 7-8 yrs	Computer 12-14 yrs	Adult home 12-14 yrs	Joint meals 12-14 yrs
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mother's schooling: All	0.039	0.041	0.050	0.042	0.013	0.056	-0.010	0.004
	[0.008]***	[0.007]***	[0.006]***	[0.007]***	[0.009]	[0.009]***	[0.009]	[0.009]
Mother's schooling: Male child	0.041	0.043	0.047	0.047	0.012	0.058	0.005	0.011
	[0.010]***	[0.009]***	[0.008]***	[0.009]***	[0.010]	[0.012]***	[0.011]	[0.012]
Mother's schooling: Female child	0.037	0.039	0.053	0.036	0.014	0.054	-0.028	-0.003
	[0.010]***	[0.009]***	[0.009]***	[0.010]***	[0.010]	[0.011]***	[0.012]**	[0.012]
Mother's schooling: High AFQT	0.038	0.035	0.048	0.040	0.016	0.040	-0.016	-0.007
	[0.010]***	[0.010]***	[0.008]***	[0.010]***	[0.011]	[0.012]***	[0.012]	[0.013]
Mother's schooling: Low AFQT	0.040	0.049	0.054	0.044	0.009	0.077	-0.004	0.015
	[0.012]***	[0.011]***	[0.011]***	[0.011]***	[0.012]	[0.013]***	[0.012]	[0.013]
Mother's AFQT (corrected): All	-0.023	0.040	0.009	-0.015	0.030	0.025	-0.047	-0.021
	[0.028]	[0.031]	[0.027]	[0.029]	[0.029]	[0.032]	[0.030]	[0.033]
Observations	2646	2644	2643	2649	2646	1681	2036	2292
Mean	0.424	0.513	0.682	0.492	0.526	0.681	0.671	0.565
Standard deviation	0.494	0.500	0.466	0.500	0.499	0.466	0.470	0.496

Note: This table reports Minimum Distance estimates for the groups indicated based on equation (2.1), see text for details. A description of the outcome variables is found in Table 2.1 on page 25. Standard errors reported in brackets, clustered by county-cohort. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

Table 2.21: Investments – OLS results: Black children

	OLS estimates: Black children							
	Museum 7-8 yrs	Musical Instr. 7-8 yrs	Special lesson 7-8 yrs	Mother reads 7-8 yrs	Newspaper 7-8 yrs	Computer 12-14 yrs	Adult home 12-14 yrs	Joint meals 12-14 yrs
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mother's schooling: All	0.029	0.035	0.064	0.060	-0.000	0.049	-0.030	0.004
	[0.010]***	[0.011]***	[0.011]***	[0.011]***	[0.013]	[0.014]***	[0.011]***	[0.011]
Mother's schooling: Male child	0.022	0.029	0.056	0.053	0.009	0.056	-0.023	0.006
	[0.013]*	[0.012]**	[0.015]***	[0.015]***	[0.014]	[0.017]***	[0.014]	[0.015]
Mother's schooling: Female child	0.038	0.047	0.072	0.065	-0.021	0.041	-0.035	0.003
	[0.015]**	[0.017]***	[0.014]***	[0.013]***	[0.018]	[0.019]**	[0.013]***	[0.013]
Mother's schooling: High AFQT	0.027	0.032	0.093	0.061	-0.010	0.056	-0.021	0.011
	[0.019]	[0.022]	[0.019]***	[0.017]***	[0.021]	[0.024]**	[0.018]	[0.016]
Mother's schooling: Low AFQT	0.029	0.036	0.050	0.059	0.006	0.046	-0.033	-0.002
	[0.012]**	[0.012]***	[0.013]***	[0.013]***	[0.017]	[0.017]***	[0.012]***	[0.016]
Mother's AFQT (corrected): All	-0.002	0.007	0.018	-0.070	0.091	0.090	0.009	-0.068
	[0.042]	[0.044]	[0.041]	[0.043]	[0.049]*	[0.048]*	[0.042]	[0.048]
Observations	1306	1305	1304	1308	1306	906	1306	1431
Mean	0.405	0.336	0.447	0.320	0.419	0.352	0.694	0.316
Standard deviation	0.491	0.472	0.497	0.467	0.494	0.478	0.461	0.465

Note: This table reports Minimum Distance estimates for the groups indicated based on equation (2.1), see text for details. A description of the outcome variables is found in Table 2.1 on page 25. Standard errors reported in brackets, clustered by county-cohort. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

Table 2.22: Early outcomes – OLS results

	OLS estimates: Children 0-1 years			
	Whites		Blacks	
	Low birthweight	MSD	Low birthweight	MSD
	(1)	(2)	(3)	(4)
Mother's schooling: All	-0.001	-0.044	-0.005	0.011
	[0.003]	[0.016]***	[0.007]	[0.035]
Mother's schooling: Male child	-0.004	-0.041	-0.006	0.019
	[0.004]	[0.018]**	[0.008]	[0.039]
Mother's schooling: Female child	0.004	-0.049	-0.003	-0.007
	[0.005]	[0.023]**	[0.009]	[0.053]
Mother's schooling: High AFQT	-0.003	-0.046	0.004	-0.022
	[0.004]	[0.018]**	[0.009]	[0.058]
Mother's schooling: Low AFQT	0.003	-0.037	-0.014	0.028
	[0.005]	[0.029]	[0.009]	[0.042]
Mother's AFQT (corrected): All	-0.007	-0.032	-0.003	-0.193
	[0.012]	[0.063]	[0.023]	[0.132]
Observations	5580	2136	2806	781
Mean	0.065	-0.039	0.130	0.184
Standard deviation	0.246	0.994	0.337	1.216

Note: This table reports Minimum Distance estimates for the groups indicated based on equation (2.1), see text for details. A description of the outcome variables is found in Table 2.1 on page 25. Standard errors reported in brackets, clustered by county-cohort. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

Table 2.23: Early channels – OLS results: white children

	OLS estimates: White children 0-1 years							
	Smoking d. pregnancy	Weeks breastfeeding	Formal child care	Hours worked	Mother reads	Book	Soft toys	Outings
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mother's schooling: All	-0.062	1.361	0.010	113.796	0.031	0.088	-0.185	-0.002
	[0.006]***	[0.337]***	[0.003]***	[10.507]***	[0.006]***	[0.013]***	[0.173]	[0.007]
Mother's schooling: Male child	-0.065	0.925	0.008	121.788	0.037	0.104	-0.167	-0.008
	[0.008]***	[0.393]**	[0.004]**	[11.946]***	[0.008]***	[0.018]***	[0.205]	[0.009]
Mother's schooling: Female child	-0.056	2.017	0.013	102.060	0.026	0.075	-0.218	0.004
	[0.009]***	[0.454]***	[0.004]***	[13.418]***	[0.008]***	[0.016]***	[0.266]	[0.009]
Mother's schooling: High AFQT	-0.053	1.516	0.013	96.074	0.029	0.084	-0.211	-0.012
	[0.008]***	[0.416]***	[0.004]***	[13.952]***	[0.007]***	[0.016]***	[0.226]	[0.008]
Mother's schooling: Low AFQT	-0.079	1.081	0.008	135.521	0.038	0.098	-0.140	0.015
	[0.011]***	[0.558]*	[0.004]*	[15.396]***	[0.012]***	[0.024]***	[0.301]	[0.011]
Mother's AFQT (corrected): All	-0.072	1.697	0.022	72.402	0.023	0.093	2.333	0.016
	[0.028]**	[1.185]	[0.009]**	[39.045]*	[0.028]	[0.053]*	[0.668]***	[0.026]
Observations	2293	2220	4850	5942	2358	2382	2343	2380
Mean	0.287	15.370	0.066	926.749	0.607	3.240	16.654	0.691
Standard deviation	0.452	22.126	0.248	880.676	0.489	1.062	12.456	0.462

Note: This table reports Minimum Distance estimates for the groups indicated based on equation (2.1), see text for details. A description of the outcome variables is found in Table 2.1 on page 25. Standard errors reported in brackets, clustered by county-cohort. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

Table 2.24: Early channels – OLS results: Black children

	OLS estimates: Black children 0-1 years							
	Smoking d. pregnancy	Weeks breastfeeding	Formal child care	Hours worked	Mother reads	Book	Soft toys	Outings
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mother's schooling: All	-0.044	1.528	0.021	175.101	0.059	0.183	0.324	0.007
	[0.013]***	[0.386]***	[0.005]***	[14.367]***	[0.013]***	[0.030]***	[0.233]	[0.011]
Mother's schooling: Male child	-0.027	1.544	0.017	172.874	0.069	0.189	0.536	0.009
	[0.015]*	[0.447]***	[0.006]***	[17.046]***	[0.015]***	[0.037]***	[0.301]*	[0.016]
Mother's schooling: Female child	-0.066	1.505	0.025	177.183	0.043	0.177	0.162	0.005
	[0.016]***	[0.493]***	[0.006]***	[16.732]***	[0.019]**	[0.038]***	[0.275]	[0.016]
Mother's schooling: High AFQT	-0.054	0.824	0.025	135.635	0.072	0.203	0.521	-0.015
	[0.016]***	[0.936]	[0.008]***	[27.401]***	[0.022]***	[0.042]***	[0.431]	[0.016]
Mother's schooling: Low AFQT	-0.031	1.598	0.019	190.480	0.052	0.170	0.233	0.028
	[0.020]	[0.395]***	[0.005]***	[17.002]***	[0.017]***	[0.036]***	[0.287]	[0.016]*
Mother's AFQT (corrected): All	0.040	0.396	0.007	152.300	0.002	0.133	-1.587	0.020
	[0.045]	[1.209]	[0.015]	[59.588]**	[0.047]	[0.095]	[0.921]*	[0.039]
Observations	861	855	2257	2965	894	897	889	897
Mean	0.278	5.513	0.070	767.310	0.371	2.337	11.227	0.661
Standard deviation	0.448	13.905	0.254	885.509	0.483	1.190	10.086	0.474

Note: This table reports Minimum Distance estimates for the groups indicated based on equation (2.1), see text for details. A description of the outcome variables is found in Table 2.1 on page 25. Standard errors reported in brackets, clustered by county-cohort. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

Table 2.25: Young adults – OLS results

	OLS estimates: Young adults (18-19 years)					
		White			Black	
	Enrollment	Conviction	Own children	Enrollment	Conviction	Own children
	(1)	(2)	(3)	(4)	(5)	(6)
Mother's schooling: All	0.025	-0.011	-0.022	0.031	-0.011	-0.042
	[0.013]*	[0.008]	[0.009]***	[0.017]*	[0.009]	[0.011]***
Mother's schooling: Male young adult	0.019	-0.025	-0.008	0.027	-0.043	-0.028
	[0.017]	[0.013]**	[0.013]	[0.023]	[0.014]***	[0.013]**
Mother's schooling: Female young adult	0.031	-0.001	-0.037	0.035	0.002	-0.067
	[0.017]*	[0.010]	[0.013]***	[0.023]	[0.010]	[0.018]***
Mother's schooling: High AFQT	0.020	-0.014	-0.019	0.015	-0.018	-0.037
	[0.017]	[0.012]	[0.014]	[0.029]	[0.015]	[0.016]**
Mother's schooling: Low AFQT	0.031	-0.008	-0.025	0.038	-0.007	-0.045
	[0.018]*	[0.012]	[0.012]**	[0.020]*	[0.011]	[0.014]***
Mother's AFQT (corrected): All	0.047	-0.041	-0.005	-0.073	-0.047	0.007
	[0.046]	[0.031]	[0.023]	[0.059]	[0.026]*	[0.042]
Observations	935	1047	816	742	889	612
Mean	0.624	0.154	0.091	0.627	0.124	0.157
Standard deviation	0.485	0.361	0.296	0.484	0.329	0.398

Note: This table reports Minimum Distance estimates for the groups indicated based on equation (2.1), see text for details. A description of the outcome variables is found in Table 2.1 on page 25. Standard errors reported in brackets, clustered by county-cohort. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level.

Chapter 3

Studying Abroad and the Effect on International Labor Market Mobility: Evidence from the Introduction of ERASMUS

3.1 Introduction

International labor market migration has risen dramatically in the recent past, especially among university graduates. Lowell (2007), for example, shows an increase in the emigration rate of university graduates from about 4 percent in 1980 to about 7 percent in 2000 for developed countries. The increased demand for skilled labor and the importance of highly skilled individuals for innovation has induced many countries to implement policies geared to attracting skilled migrants from abroad (OECD, 2002). Understanding the determinants of migration is key to formulating such policies. While attention has traditionally focused on wage differentials, going back to

Hicks (1932)[1], it is clear that other factors are important determinants of international mobility. One possible determinant which has received particular attention of policymakers over the past years is student mobility during tertiary education. In particular, it has been hypothesized that student mobility may act as a 'stepping stone' for later labor migration (Guellec & Cervantes, 2001). Numerous countries, including the United States, Japan, and the United Kingdom, attempt to attract highly skilled mobile workers through policies relating to student mobility programs (Guellec & Cervantes, 2001). These are based on the assumption that student mobility has a genuine effect on later labor market mobility. Despite the widespread belief in the link between studying abroad and international labor market mobility, empirical evidence is very limited. Establishing a causal link between studying abroad and labor market mobility later in life is a challenging task because students who decide to study abroad are in many ways different from students who undertake all of their education in their home country. The unobserved heterogeneity may also affect the decision of working abroad later in life. This may introduce a bias in OLS estimates of the effect of studying abroad on subsequent international labor migration decision.

In this chapter, we provide evidence on the *causal* effect of studying abroad on later labor market mobility by exploiting an exogenous change in student mobility: the introduction of the ERASMUS student exchange program. This program has been devised by the European Union to foster student exchange in Europe. Introduced in 1987 it offers the possibility of studying in another European country for up to 12 months at very low cost. Different universities and different departments introduced the program at very different times. We exploit the variation in scholarship availability as a source of exogenous variation in a student's probability to study abroad. In order to ascertain a student's exposure to the ERASMUS program we construct a unique data set, containing annual information on the number of exchange places for each subject at every German university. In order to assess the effect of studying

[1] For surveys on determinants of migration, see Greenwood (1975, 1985, 1997).

abroad on international mobility later in life we merge this data to a survey of German university graduates. We first show that the ERASMUS program has a strong impact on a student's probability of studying abroad. We then use the department level variation in international student exchange programs to identify the causal effect of studying abroad on the decision of working in a foreign country later in life. We find that studying abroad increases a person's probability of working abroad by about 15 percentage points. This result suggests that studying abroad has a strong causal effect on labor market mobility later in life. Qualitative evidence suggests that besides career concerns soft factors such as interest in foreign cultures or living with a foreign partner are important determinants for the decision to work abroad, and we suggest that the effect of studying abroad may work through these channels.

There are some papers analyzing the link between labor market mobility and previous mobility. Kodrzycki (2001) provides descriptive evidence on inter-state mobility in the US and links it to the preceding decision of attending college out of state.[2] Using individual-level data from the U.S., Groen (2004) documents that studying in a given state increases the probability of later working in that state, accounting for selection by exploiting information on the set of states individuals applied for. Bound, Groen, Kezdi & Turner (2004) estimate that increasing production of college graduates at the state level leads to moderate increases in the stock of college-educated workers in that state.

The link between *international* student mobility and the decision to work abroad after graduation has rarely been studied to date. One reason is data availability: Most surveys do not contain information on study abroad spells during a student's undergraduate career, and graduates who work abroad are generally not sampled in national surveys of the sending countries. Jahr and Teichler (2001) use data from a survey of European university graduates who have been internationally mobile. They

[2]She finds that individuals who attended college out of state are 54 percent more likely to live out-of-state five years after graduation. These results, however, cannot be interpreted as causal effects as she does not address the selection issues affecting mobility decisions.

investigate the effect of studying abroad on later international labor market mobility without controlling for possible selection of formerly mobile students. They find that formerly mobile students are between 15 and 18 percentage points more likely to work in a foreign country after graduation. Dreher and Poutvaara (2005) investigate the role of student mobility in explaining aggregate migration flows in a cross-country panel study, focusing on migration to the United States. They find strong effects of previous period's number of foreign students on current period's number of migrants, indicating that a ten percent increase in the number of foreign students increases subsequent migration by around 0.5 percent.

The paper which is most closely related to this work is a study by Oosterbeek and Webbink (2009). They employ a regression discontinuity design to control for unobserved heterogeneity between internationally mobile and non-mobile students. Using data on talented Dutch university students they find that studying abroad increases the probability of living in a foreign country by about 50 percentage points. A key difference to our work is that they look at a small sample of particularly talented students, while we use a nationally representative survey of German university graduates. Another important difference is that Oosterbeek and Webbink investigate the effect of *postgraduate* studies abroad. Students pursuing a postgraduate degree abroad may remain in the receiving country while looking for work. Part of the effect they find may also be driven by the fact that some of the respondents abroad are still enrolled in higher education at the time of the survey. In contrast, in our work, the intervention is international mobility during the undergraduate career, after which students return to complete their degree in Germany. Thus, our research design allows us – and in fact forces us – to separate the two mobility investments (studying abroad and working abroad). The effect we find is therefore informative about the dynamic effects of earlier mobility investments.

This chapter presents evidence that previous educational mobility is a very important determinant of mobility later in life. We thus establish a causal link of previous mobility decision to mobility later in life. This highlights the importance of taking earlier mobility into account in economic modeling but also for policy decisions. The European Union, for example, tries to foster labor market mobility in the EU (see 'Commission's Action Plan for skills and mobility' (2002)). Our research suggests that supporting international student mobility is a very successful policy instrument to foster labor market mobility later in life. Our results on the effect of the ERASMUS program on the probability of studying abroad also show that exchange programs are indeed effective in promoting student mobility. This will be important to policy makers as they spend large public funds on these programs.

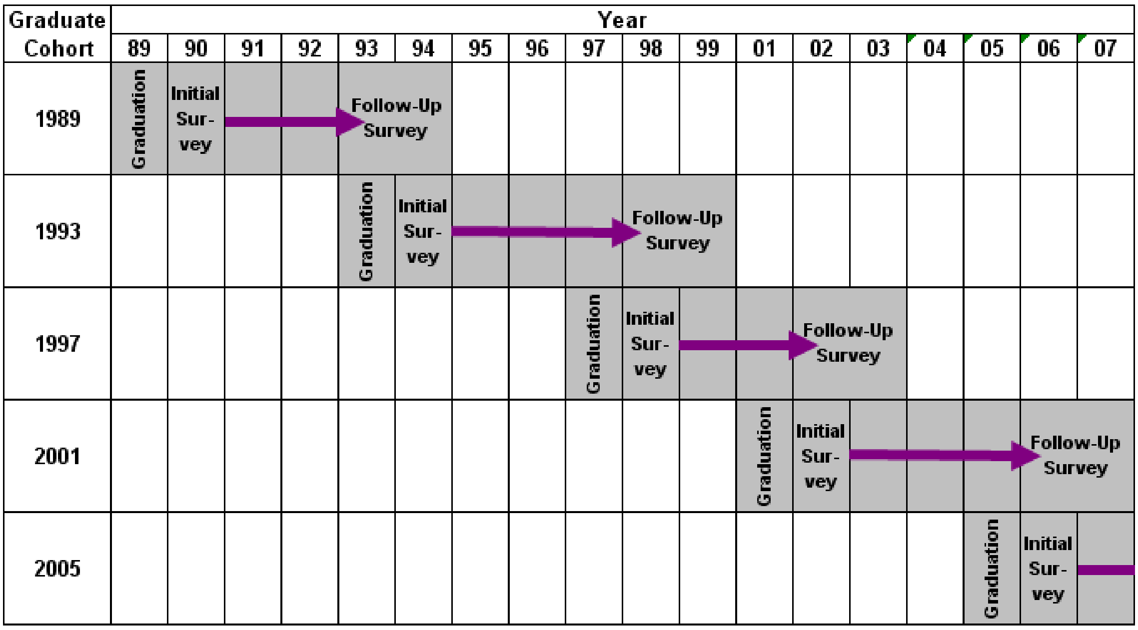
We emphasize that our primary interest lies in understanding the role of studying abroad as a determinant of individual international labor migration decisions, and the use of the ERASMUS program is motivated by the variation it induces in students' decision to study abroad. Our data does not allow to investigate the role of the ERASMUS program on immigration of skilled graduates from other countries to Germany, or the overall effect of studying abroad on the international distribution of human capital, although these are potentially interesting and important questions.

The chapter proceeds as follows: The next sections briefly describe the data we are using and provides some institutional detail on the ERASMUS program. Section 3.4 outlines our identification strategy. In the following section we report our first stage results and provide evidence that our instruments are both powerful and operate very precisely in the way we claim they do. Section 3.6 presents the main results and a number of sensitivity checks. We present descriptive evidence into the channels which lead students who studied abroad to work abroad later on. The last section concludes.

3.2 Data

We use data on German university graduates, which has been collected by the Higher Education Information System (HIS) institute. This survey is conducted to provide a nationally representative longitudinal sample of individuals who complete their undergraduate education in Germany. A sample of university graduates has been drawn from cohorts graduating in the academic years 1988-89, 1992-93, 1996-97, 2000-01, and 2004-05. In the following, we will refer to these five cross-sections as graduate cohorts 1989, 1993, 1997, 2001, and 2005. Graduates in each cohort are surveyed twice. The first survey takes place about 12 months after graduation (the *Initial Survey*). The same individuals participate in a follow-up survey about 5 years after entering the labor market (*Follow-Up Survey*).[3] The following Figure 3.1 illustrates the timing of the different surveys.

Figure 3.1: HIS Data



The data contains detailed information on the students' background, study history, and labor market characteristics. This allows us to relate study decisions, in particular international educational mobility, to later labor market outcomes. A large advantage

[3]For the 2005 cohort, only the initial survey is available so far.

of this dataset lies in the fact that individuals graduating from a university in Germany are followed even if they move to a foreign country. This feature makes this dataset particularly valuable to investigate questions concerning international mobility.

The data and the sampling process is described in detail in Briedis & Minks (2004). The sample was drawn as follows: For each cohort, university-subject-degree combinations where sampled randomly, and the respective universities mailed the questionnaire to each student who had graduated within the corresponding academic year. This procedure ensures that the sample contains individuals from a large number of different institutions and subjects. One key advantage of the data is that the population of interest includes all university graduates who completed their undergraduate studies during a given academic year at any institution of higher education in Germany.[4] The data contains no information on nationality of respondents. It contains, however, some information on where the students obtain their highschool degree. We limit our sample to all those individuals who obtain their highschool degree in Germany. The response rate to the survey is around 25%. While of course a higher response rate would be desirable, an analysis conducted by the HIS has come to the conclusion that the characteristics of the survey respondents are close to those of the target population. The total number of respondents corresponding to the five cohorts is 12,457 (1989), 11,314 (1993), 9,586 (1997), 8,124 (2001), and 11,784 (2005).

The key information for our purposes is whether the student has studied abroad during her undergraduate studies, and whether the graduate works abroad at the time

[4]The higher education system in Germany consists of a number of different university types catering to different types of students. We include five main types of higher education institutions in our estimation. This includes not only the traditional universities *(Universitäten)* but also the so-called Universities of Applied Sciences *(Fachhochschulen)*, the Comprehensive Universities *(Gesamthochschulen)*, the Colleges of Art and Music *(Kunst- und Musikakademien)*, and the Theological Universities *(Theologische Hochschulen)*. All institutions in our sample would be called universities in most countries outside Germany. Admission requirements differ by subject. For a subset of traditionally oversubscribed subjects, admissions are awarded centrally based on nationwide quotas. For all other subjects, higher education institutions may place local restrictions on admissions. These are likely to vary strongly between subjects. Criteria applied for admission decisions vary, but include overall highschool grade, waiting periods since graduating from highschool, and interviews. Further details and references on the higher education system in Germany can be found in KMK (2008).

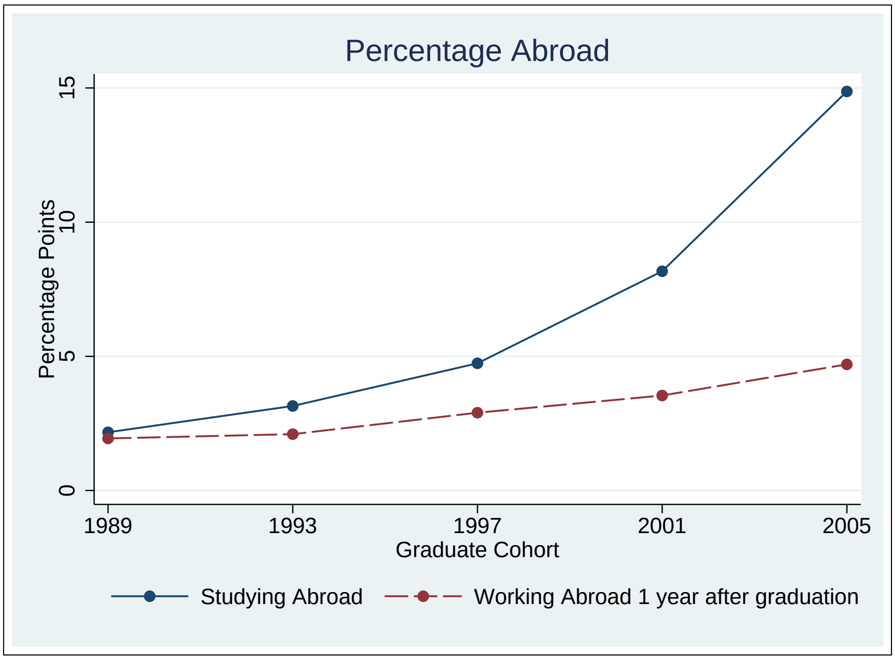
of the survey. We infer undergraduate mobility from the first question of the questionnaire, which asks the student to report her complete enrollment history. Respondents are instructed to report each change of degree program or university. The questionnaire makes explicit reference to study abroad as one form of change in status in the 2001 survey. For the 2005 cohort it contains an explicit question regarding study abroad spells. We use this information to construct an indicator of whether the student studied abroad during her undergraduate career. In order to exclude university mobility after finishing the first degree (e.g. to obtain a Master abroad), we only look at international mobility before the graduation date of the first degree. It is important to note that only students who obtain their degree in Germany are surveyed. We are, therefore, not able to observe students who first enrol in Germany and subsequently move to a foreign university and obtain their degree abroad. Also Germans who complete all of their higher education abroad are not included in our sample. These individuals may be different to students who study abroad as part of their degree in Germany. It is quite likely that those who complete their higher education abroad are even more likely to work in a foreign country after graduation than students who obtain their degree in Germany. If this was true we would underestimate the effect of studying abroad. Unfortunately, our data is not suitable to test this hypothesis.

For all students who have ever participated in the labor market, both the initial and the follow-up surveys contain questions about the current (or the last) employment, including the location of work. We infer from this question whether a former student now works in Germany or abroad, and create an indicator accordingly.

The following figure shows the percentages of studying abroad and working abroad (from the initial survey, one year after graduation) for the five graduation cohorts. It can be seen that both studying abroad and working abroad occurs more frequently among students of later graduation cohorts. It is important to note that we include dummies for the five graduation cohorts in all our regressions. Therefore, we do not identify the effect of studying abroad from the overall time-trend in the two variables.

In fact, in our sensitivity analysis, we show that our results are robust to allowing for not only a general time trend, but also for subject-specific time trends.

Figure 3.2: International Mobility in HIS Data



These percentages can be compared to information on international mobility from other data sources. Isserstedt & Schnitzer (2002) point out that different data sources use different ways to collect data and different definitions of a stay abroad. These differences may result in different estimates of student mobility. With this caveat in mind, we compare the incidence of international educational mobility in our data to data from the 16th Social Survey (*Sozialerhebung*), a large-scale survey of German students in 2000. Of all students surveyed in the Social Survey, about 13 percent of advanced students indicate that they spent part of their studies at a foreign university. The students surveyed in 2000 will mostly graduate before 2005. In the 2005 graduate cohort data about 15 percent have studied abroad. This is very similar to the fraction in the Social Survey. The figures from the Social Survey also replicate the strong over-time increase in the fraction of students who study abroad.

With similar caution we use data from the OECD Factbook 2006 to investigate the reliability of our data with respect to international labor market mobility. The OECD estimates that about 7.1 percent of Germans holding a university degree worked as expatriates in a foreign country in the year 2005. This number is higher than the percentage of people working abroad for the 2005 cohort in our dataset. This is due to the fact that the OECD figure measures stocks of expatriates while we consider the flow of university graduates to foreign countries.

We conclude that both the percentage of people studying abroad and the percentage of people working abroad in our data are comparable to estimates from other data sources. This is reassuring as there may be a worry that response rates to the HIS survey may differ for people living abroad. Unfortunately, there is no direct way of testing for differential response rates as we do not have any information on the individuals who do not respond to the HIS survey. One way of addressing this concern is to show that other data sources with different sampling frames exhibit similar numbers to our data.

In addition to the international mobility variables we also use a number of other control variables measured at the individual level. All sampled graduates received their *first* university degree. In the earlier cohorts students received a traditional German degree (*Diplom* or *Staatsexamen)*. A small proportion from the 2005 cohort was awarded a bachelor degree.[5] We therefore include an indicator for obtaining a bachelor degree in our regressions.

Furthermore, we create a measure of potential experience since graduation, defined as the number of months from graduation to the time of answering the questionnaire.[6]

[5]This reflects the recent introduction of bachelor and master degrees at most universities. Traditionally there was no distinction between bachelor and master degrees in the German higher education system. Students would enroll at a university after high school and study for about four to seven years obtaining one degree at the very end of their studies.

[6]There is some variation in experience because students were sampled according to whether their graduation fell in a particular academic year. Students graduating at the beginning of the academic year therefore have more potential experience than those graduating towards the end of the year. In addition, there is some variation with respect to when the questionnaires were sent out and how quickly graduates responded. We take this measure of potential experience rather than actual labor

Other controls include a female indicator, age at beginning of university studies, and an indicator for whether the student completed an apprenticeship before beginning her university studies. We also use variables which control for a student's earlier mobility decisions. In particular we include a variable which controls for whether the student's first university enrollment occurs in the state *(Bundesland)* where she obtained her final high school degree. Furthermore, we include the distance between the state of her university enrollment and the state where she obtained her high school degree.

We use a number of variables to control for a student's parental background. To control for parental education we use a variable that indicates the highest grade completed by either parent, where we split parental education into three categories to account for the characteristics of the education system in Germany.[7] We also construct indicator variables in five categories for each parent to control for parental occupation. As a proxy for credit constraints we use a variable measuring the proportion of expenses which the student covers by federal financial aid (BAFOEG). Students are eligible to this assistance if parental income is below a certain threshold. This threshold varies according to the number of children who are enrolled in a formal education program.

Our data also contains information on industry and occupational status of the surveyed graduates. Although our main analysis does not make use of this information, it may still be of interest to compare the respective distributions between internationally mobile individuals and individuals who remain in Germany.

In order to implement our Instrumental Variables strategy we combine the HIS

---

market experience, because actual labor market experience could be affected by a study period abroad and might then be endogenous to our outcome.

[7]The omitted category contains students with parents who obtained up to 13 years of education. This group consists of students with parents who did not receive a school degree (very few), parents with lower types of secondary schooling *(Hauptschule or Realschule)* usually followed by an apprenticeship, and parents who obtained a high school degree but no further education (very few). The second group is comprised of students where the better educated parent either obtained an advanced craftsmanship degree *(Meister)* or some higher education, such as a degree from a university of applied science *(Fachhochschule)* but not a degree from a university. The third group includes students who have at least one parent holding a university degree. Using a linear years of parental education variable or controlling for mother's and father's education separately does not affect our results.

graduate survey data with a unique dataset of ERASMUS participation. There is no readily available data on the ERASMUS exchange program for our time period of interest. We obtained data on the number of ERASMUS scholarship holders for each year and each participating institution on a subject-by-subject basis from 1993/94 to 2004/2005 from the German Academic Exchange Service (DAAD). To obtain the data for the earlier years we proceeded as follows: The DAAD provided us with the number of scholarships allocated to each ERASMUS inter-university agreement (Inter-university Cooperation Program, ICP). We combined this information with published listings of all ICPs, which give details about the participating universities and the subjects covered for each inter-university agreement (see, for example, DAAD (1992)). This allows us to construct a panel data set at the university-subject-year level that covers the entire history of the ERASMUS program in Germany. The typical (median) student goes abroad three years prior to his graduation, and we assign to each student the exposure to the ERASMUS program in that corresponding academic year.[8]

We restrict our sample to those observations for which all variables of interest are observed. As mentioned before, students from the graduate cohorts 1989, 1993, 1997, and 2001 have been surveyed twice, the first time one year after graduating from university and a second time five years after graduation. We thus have two observations for the location of work for most individuals from those cohorts. In the estimation below, we pool the observations from the initial and the follow-up survey for efficiency reasons.[9] This allows us to use the information provided in both questionnaires. Means and standard deviations of our estimation sample are reported in Table 3.1. It is evident from comparing columns (2) and (3) that individuals who studied abroad are also more likely to work abroad later in life. One can also see that

[8]This approach is preferable to simply assigning ERASMUS characteristics at a fixed point in the student's study period (say the second or third year): since our graduates are sampled when they exit university, and since there is substantial variation in length of studies, there might be a systematic relationship between individual study duration and other unobservable factors.

[9]By clustering the standard errors at the institution level, we fully account for the resulting dependence in the error terms.

Table 3.1: Summary Statistics

	(1) All	(2) Study Abroad = 0	(3) Study Abroad = 1	(4) Work Abroad = 0	(5) Work Abroad = 1
Working abroad	0.032	0.027	0.102	0.000	1.000
	(0.176)	(0,163)	(0.303)	(0.000)	(0.000)
Undergraduate study abroad	0.062	0.000	1.000	0.057	0.198
	(0.241)	(0.000)	(0.000)	(0.232)	(0.399)
ERASMUS indicator	0.490	0.472	0.767	0.485	0.637
	(0.499)	(0.499)	(0.423)	(0.500)	(0.481)
ERASMUS ratio	0.031	0.028	0.068	0.030	0.044
	(0.056)	(0.053)	(0.081)	(0.055)	(0.064)
Female	0.450	0.445	0.512	0,449	0.474
	(0.500)	(0.497)	(0.500)	(0.497)	(0.499)
Age when starting studies	21.637	21.682	20.959	21.655	21.082
	(2.559)	(2.603)	(1.595)	(2.577)	(1.831)
Experience	2.686	2.700	2.466	2.670	3.160
	(2.074)	(2.074)	(2.066)	(2.067)	(2.231)
Apprenticeship	0.301	0.313	0.194	0.309	0.206
	(0.461)	(0.464)	(0.396)	(0.462)	(0.405)
Mother's Education (years)	12.283	12.168	14.024	12.240	13.582
	(3.322)	(3.288)	(3.356)	(3.315)	(3.282)
Father's Education (years)	13.707	13.597	15.387	13.665	14.992
	(3.554)	(3.544)	(3.275)	(3.557)	(3.200)
Final University Grade[1]	2.041	2.057	1.812	2.048	1.848
	(0.681)	(0.681)	(0.633)	(0.682)	(0.604)
Credit Contrained[2]	0.119	0.120	0.098	0.120	0.099
(High Financial Assistance)	(0.324)	(0.325)	(0.297)	(0.325)	(0.298)
% in respective Industry:[3]					
Agriculture, Energy	2.6	2.6	1.7	2.6	2.9
Manufacturing	21.4	21.8	14.8	21.4	21.2
Services	40.9	40.9	40.9	41.0	37.7
Education, Culture	23.7	23.1	31.7	23.4	32.4
Administration, Organisations	10.7	10.7	10.5	10.8	4.9
Other	0.0	0.0	0.3	0.0	0.3
% in respective Occupation:[4]					
Manager	5.8	5.9	3.5	5.7	6.7
employee	69.6	69.4	71.4	69.2	81.7
self-employed	8.8	8.7	9.2	8.8	7.6
civil servant	11.8	11.9	9.6	12.1	1.7
other	4.1	4.0	6.3	4.2	2.3
Observations	54079	50741	3338	52355	1724

[1]The final university degree is only available for 52830 students in our sample. (The best grade is 1.0 the worst 4.0) [2]The question on financial assistance has only been administered between 1993 and 2001. In 1989 the students were directly asked about their financial situation. We therefore have the information on credit constraints for 45307 individuals. [3]The industry information is available for 53427 individuals. [4]The information on occupation is available for 53190 individuals. Note: This table contains sample means and (in brackets) standard deviations. For industries and occupations it contains percentages.

individuals with more exposure to ERASMUS (as measured by ERASMUS ratio or ERASMUS indicator, which are described in further detail below) are more likely to study abroad. In the following section we explain how we use the ERASMUS program to identify the causal link between studying abroad and international labor market mobility later in life.

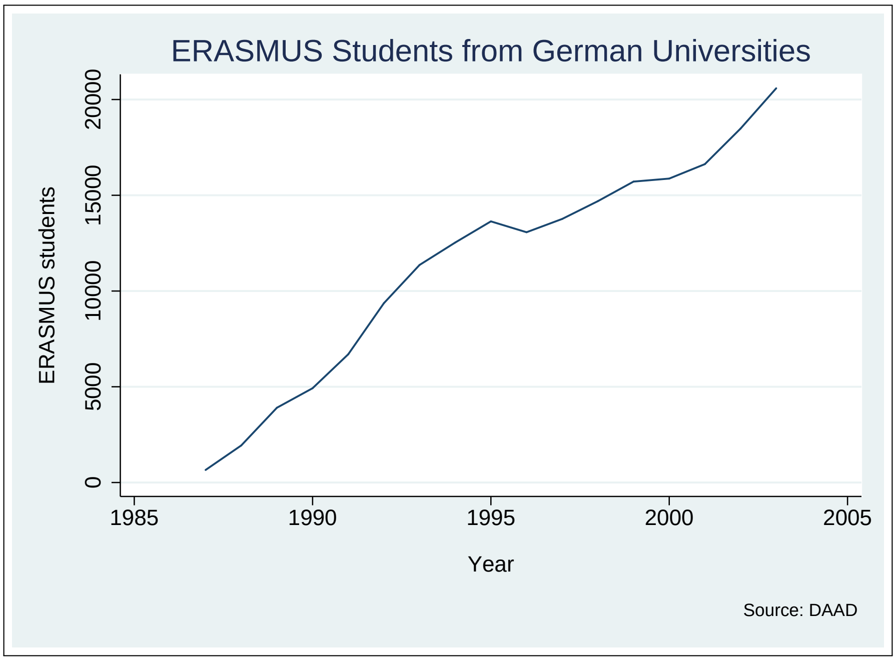
3.3 The ERASMUS Program

Our identification strategy relies on the large scale introduction and expansion of the ERASMUS program. In 1987, the Council of Ministers of the European Community passed the *European Community Action Scheme for the Mobility of University Students* (ERASMUS). The main objective of ERASMUS is 'to achieve a significant increase in the number of students [...] spending an integrated period of study in another Member State' Council of the European Communities (1987). Student mobility was to be increased through the creation of a European university network, individual scholarships, and mutual recognition of academic credits Smith (1988). Since then, ERASMUS has continually expanded. Looking across all participating countries, 1.37 million students have taken part in ERASMUS in the period of the academic years 1987/88 to 2004/05, with 15.7% of those outgoings coming from Germany. Figure 3.3 shows the number of German outgoing students for each year since the introduction of the program.

Due to this dramatic expansion, students in our five graduate cohorts are affected quite differently by the program. The expansion of ERASMUS has significantly contributed to the overall incidence of studying abroad. Our data shows that about 8 percent of the students in the 2001 graduate cohort have studied abroad as part of their undergraduate degree. It can be calculated that about 5 percent of the 2001 graduation cohort have studied abroad with an ERASMUS scholarship.[10] The ERAS-

[10]This number is obtained as follows: In the 2001 graduate cohort, the median student started her

Figure 3.3: ERASMUS in Germany



MUS program therefore accounts for more than half of international undergraduate mobility in Germany in the 2001 cohort.

Students participating in the ERASMUS program apply for an exchange scholarship at their home university usually one year before they intend to study abroad. The department then decides who is awarded an ERASMUS scholarship. The criteria for obtaining an award are mostly based on academic achievement and motivation (as demonstrated in a written statement of interest and/or an interview). In very rare cases the places are allocated on a first come first serve basis.[11] The award of the scholarship not only secures them a place at a certain partner university abroad but also provides them with a small mobility grant. In the academic year 2001/2002 (the year a typical student from the 2005 graduation cohort went abroad) an outgoing

tertiary studies in the academic year 1995/96. In that year, about 262,000 students entered university. The typical exchange student in that cohort studied abroad in the third year of her studies. In that year 13785 students from German universities participated in the ERASMUS program. This corresponds to about 5% of the entire cohort.

[11]For more information on the allocation process see Maiworm, Steube, and Teichler (1993).

student from Germany received about 146 Euros per month for her stay abroad. In addition to receiving the mobility grant the ERASMUS student receives a tuition fee waiver at the foreign university. Another important benefit of ERASMUS is that it significantly reduces the student's application costs and the time the student needs to apply in advance to be able to organize a stay at a foreign university.

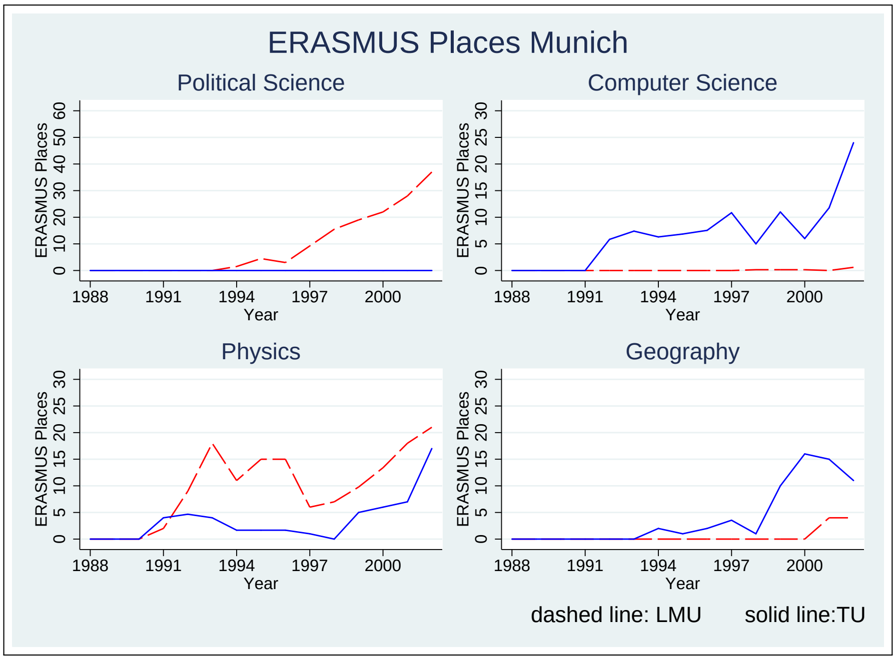
University participation in ERASMUS operated through Inter-University Cooperation Programs (ICP), in which groups of university departments from different countries formed a network covered by an ICP agreement, typically initiated through an active professor who happens to have contacts with professors at foreign universities. If new universities join the ICP additional places may become available. Many departments would at some point enter ERASMUS with a few links to departments at foreign universities. Over time other foreign departments would be taken into the network increasing the number of exchange places for German students. Similarly the German department itself would enter other (possibly new) cooperation networks. One way to interpret the evolution in ERASMUS scholarships is to think of the cooperations as an emerging network.

The professors involved in the organization of the ERASMUS student exchange program agree on the number of incoming and outgoing ERASMUS places for each participating university. These agreements are usually longer-term contracts covering a number of years. Thus, the number of exchange places with a certain foreign university stay constant for some years. Sometimes not all admitted students take up their place because they receive another scholarship or because they change their mind about wanting to study abroad. As the ICP agreements are long term agreements this does not affect the number of slots in the following year.

In order to give a insight into the variation, which is exploited in our identification strategy, we show the raw data on the number of ERASMUS students at four departments at the two large universities in Munich in the following figure.[12]

[12]We choose the Ludwig-Maximilians University and the Technical University Munich for this de-

Figure 3.4: ERASMUS in Munich



The introduction of the ERASMUS program at a certain department occurred at different points in time at the two universities even though the universities are of very similar quality. This indicates that a large degree of the variation in ERASMUS places is due to idiosyncratic shocks triggered by the contacts of some active professors.

3.4 Identification Strategy

To provide a simple conceptual framework, we start from the description of the individual migration decision from Borjas (1987). A university graduate deciding to work

scriptive analyis because they are located in the same city and are of similar quality and reputation. This is exemplified by the fact that these two unversities were among only three universities to be selected as winner of the 'Initiative for Excellence' in 2006. This initiative allocates federal funding to German universities which are considered to have the potential to become world-class research universities. This potential was evaluated based on the universities' past performance and on their strategic plans for the future.

abroad or at home, facing wages at home (w_0) and wages abroad (w_1) as follows

$$\log(w_1) = \alpha_1 + u_1 \tag{3.1}$$

$$\log(w_0) = \alpha_0 + u_0 \tag{3.2}$$

where (u_1, u_0) denote idiosyncratic error terms around means (α_1, α_0). The individual decides to work abroad if the return to migration exceeds the cost of migration (C). Thus, the resulting decision rule is

$$\text{Work abroad} = 1\left\{\log(w_1) - \log(w_0 + C) > 0\right\} \approx 1\left\{u_1 - u_0 > -(\alpha_1 - \alpha_0 - C/w_0)\right\}. \tag{3.3}$$

The key prediction of this Roy model in this context is that the probability of working abroad decreases with cost of migration C. Our focus lies in understanding the role of studying abroad as one important determinant lowering the cost for later labor market migration. There are a number of channels how studying abroad may reduce the cost for later migration decisions. Studying abroad allows the students to improve their foreign language skills. This would greatly reduce their costs of finding work in the foreign country. Furthermore, they will acquire a better knowledge of the foreign labor market and maybe get in contact with potential employers. Also personal contacts through friends in the foreign country may facilitate finding a job in a foreign country. We show below that individuals often return to work in very same country where they have studied abroad. This supports the hypothesis that these channels are indeed important. Other channels how studying abroad may lower the cost of migration are more subtle. The study abroad spell may act as a trial period of whether one likes to live in a foreign country and thus increase the interest in foreign cultures. Furthermore, studying abroad may foster private relationships abroad which draw the student to working abroad later on. Below we provide some suggestive evidence that these channels may indeed be affected by studying abroad.

In order to investigate the relationship between studying abroad and later labor market mobility we therefore estimate the following equation.

$$\begin{aligned}\text{Work Abroad} \;=\;& \beta_1 \;+\; \beta_2 \,\text{Study Abroad} \;+\; \beta_3\, X \;+\; \beta_4 \,\text{Cohort FE} \\ & +\, \beta_5 \,\text{Subject FE} \;+\; \beta_6 \,\text{University FE} + u \qquad (3.4)\end{aligned}$$

Where *Work Abroad* and *Study Abroad* are dummy variables indicating whether an individual worked abroad or studied abroad, respectively. X is a vector of personal characteristics, which may affect the decision to work abroad, such as gender, age, work experience or an individual's family background. We also include a full set of dummies for each graduate cohort, a student's subject, and university. Our main interest lies in obtaining consistent estimates of β_2.

The summary statistics presented above clearly indicate that students who study abroad differ systematically in their observable characteristics from those who remain in Germany throughout their undergraduate studies. Although our data set is rich in observed characteristics of the student, many dimensions which are likely to affect the students' mobility decision remain unobserved. A possible factor could be, for example, the students' unobserved motivation. If these unobserved factors are correlated with the outcome, estimating equation (3.4) using OLS would yield biased estimates, because we would mistakenly attribute the effect of the unobserved covariates to the stay abroad. While it is generally difficult to characterize these unobserved components in its entirety, there is some direct evidence of what factors may play a role. In their sociological analysis of determinants of studying abroad, Muessig-Trapp & Schnitzler (1997) identify as critical factors affecting the decision to study abroad the student's financial situation, whether she holds any part-time job, foreign language skills, the expected labor market benefit of going abroad, and her motivation and personality structure. Clearly, many of these dimensions will be unobserved to the econometrician. Thinking about our outcome of interest it is likely that the same unobserved factors which drive the decision to study abroad will also affect the decision of where to look for a job. It is therefore not clear what at all can be learned from a comparison of means of those who study abroad versus those who do not. This underlines that this context requires a credible identification strategy to learn about the causal impact of the study period abroad. We use the ERASMUS program as an instrumental variable to identify the causal effect of studying abroad. As our first stage we estimate the following equation:

$$\begin{aligned}\text{Study Abroad} \;=\;& \gamma_1 + \gamma_2\text{ERASMUS} + \gamma_3 X + \gamma_4\text{Cohort FE} \\ & + \gamma_5\text{Subject FE} + \gamma_6\text{University FE} + \epsilon \qquad (3.5)\end{aligned}$$

ERASMUS is a variable measuring a student's exposure to the ERASMUS program. In addition to the main variables of interest we include the same control variables as in equation (3.4).

It is important to be precise about the variation we exploit to identify the effect of studying abroad. We account for systematic differences between universities by including university fixed effects. Our empirical strategy thus relies on over-time changes in scholarship availability. At the same time, we include dummies for our five graduate cohorts, so that any difference that is common to all students in a cohort is taken out as well. This ensures that we are not relying on any long-term trends (which may possibly affect both the instrument and the outcome). In addition to that we include subject fixed effects in our estimation. This accounts for any systematic difference in international mobility of students in different subjects. We therefore rely on over-time changes in program intensity at a given subject and university combination. Probing the robustness of our findings we also include subject specific time trends in our specifications. These allow for a separate linear trend in the probability of studying abroad for each subject. The nature of our results is not affected by including those time trends. In another robustness check we further control for possible unobserved heterogeneity by including fixed effects for the interaction of a student's faculty (such as humanities or science faculty) and her university. We show below that our findings are robust to using these fixed effects.

We construct different measures of a student's exposure to the ERASMUS program. The first ERASMUS measure is an indicator, which takes the value 1 if the student's department offered an ERASMUS scholarship in the relevant year. In most cases this variable is 0 until a certain department joins the ERASMUS program and 1 thereafter, because very few departments leave the program after they have joined. We denote this variable *ERASMUS indicator,* which varies in the dimensions university, subject, and year. Using the ERASMUS indicator as an instrument amounts to a classical difference-in-differences estimator comparing students before and after the introduction of an exchange program for their subject at their university.

The second variable measures the exact number of ERASMUS scholarships, offered by each department at every university in a given year. In order to account for differences in size of different departments, we normalize the number of scholarships with the number of students enrolled in the respective department. We use the department level number of first year students in the fall semester of the academic year 1992/93 for this normalization. In the following we refer to this variable as *ERASMUS ratio.* This measure for a student's exposure to the scholarship program varies at the university, subject, year level as well.

The ERASMUS indicator is less powerful than the ratio because it does not capture changes in the number of ERASMUS scholarships, which certainly affect a student's probability of studying abroad. On the other hand, however, this disadvantage may be an advantage if student demand at a department affects the number of ERASMUS places. This would affect the credibility of any instrument using the actual number of ERASMUS scholarships. Even though we believe that this is not an important concern in practice the ERASMUS indicator is a way of dealing with this concern. The only way in which student demand may affect this instrument is through triggering the introduction of ERASMUS in the relevant department, which we believe is extremely unlikely. Administrative hurdles when setting up the program stand in the way of any short term responses to student demand. If a certain department wants to join the ERASMUS program, the university has to apply for a certification at the European Commission. Moreover, the department has to find partner universities, which are willing to exchange students with the given department. Clearing these administrative hurdles takes time. Another time lag is introduced by the fact that students have to apply for a certain ERASMUS slot almost one year before they actually study abroad. It is therefore very unlikely that departments are able to set up a new ERASMUS program in time for a certain cohort to be able to benefit from that introduction.

In the following, we address a number of possible concerns regarding the exclusion restriction. In particular, we consider the '*university quality*' argument, the '*big push*' argument, and the '*student selection*' argument.

One concern may be that university quality affects both scholarship availability and the outcome: If good universities offered more ERASMUS scholarships, and if at the same time good universities produced higher skilled graduates who are more likely to find a job in a different country, the exclusion restriction would be violated. We take care of this problem by including university fixed effects (FE) in all our regressions, which control for any permanent university attribute. A closely related criticism is that even within a given university some faculties, such as sciences, may be better than other faculties. We show that our results also hold if we include faculty times university fixed effects, which control for any permanent difference between faculties even within a given university.

A common concern in IV estimation is that using a particular policy may carry the risk of not accounting for other policies which were implemented at the same time. For example, the university could engage in more active exchange activities also outside Europe and possibly implement other measures which increase the employability abroad at the same time. We show below that ERASMUS had a very narrow effect and does not seem to be correlated with other policies. To check for the correlation with other programs we use information of where students went to study abroad, grouped into three categories (Europe, United States, and other areas). We show below that the ERASMUS program only affected the exchange to Europe but not to other areas.

Similarly, one may be worried that active professors who play an important role in expanding a department's exchange network may also be more involved in placing their students internationally once they graduate, having a direct effect on the outcome. We can assess this directly since our data contains information on whether students obtained their first position through intermediation of a professor. We find no evidence that there is any systematic relation between this job finding channel and ERASMUS scholarship availability, suggesting that ERASMUS exposure in a department is not correlated with a department's job placement activities.[13]

Another concern is that students may choose a particular university-subject combination because of scholarship availability. Particularly mobile students might choose universities and departments offering a large number of ERASMUS scholarships. This would again bias our IV results. We do not think that this is likely to occur, however. Since most of our sampled individuals started their university career long before the widespread availability of the internet, information about exchange programs was extremely difficult to obtain. Even nowadays it is hard to obtain information on the availability of ERASMUS scholarships on departmental websites of German universities. It is much more likely that enrollment decisions are based on factors such as reputation of the university or closeness to home. We also address the student selection argument by controlling for distance between the state of a student's highschool degree and her university. Controlling for earlier mobility does not affect our results.

Another way of addressing these concerns more directly is to define our measure

[13]In a simple Pearson's χ^2 test, we cannot reject the hypothesis that this job finding channel and the ERASMUS indicator are independent (p=0.62). When we regress an indicator for obtaining the first position through intermediation of a university professor on our ERASMUS measures in a full specification corresponding to our main model, we find no significant effects of the ERASMUS measures.

of ERASMUS exposure without exploiting the specific choice of university the student made.[14] For this purpose, we define a third measure (*ERASMUS subject ratio*) as the ratio of ERASMUS scholarships in the student's subject across all universities, relative to the overall number of students in that subject (again across all universities). This measure does not depend on the specific university a student chooses. As a variant we use the subject ratio measure but substract the ERASMUS slots in the student's own department.[15] In the tables this measure is denoted as *ERASMUS subject ratio, excluding own department*. As this measure does not include the student's own ERASMUS slots it will be completely unaffected by a possibly endogenous selection of a certain department with more ERASMUS places. We show below that our results are very similar when we use this alternative measure of ERASMUS.

A related worry is that students may change university or department after they figured out that their university and/or department offers little opportunity to study abroad.[16] Using the ERASMUS measures from a student's *first* enrollment enables us to avoid any problems of selective mobility after university entry of the student.

In summary, we believe that in our empirical framework ERASMUS scholarship availability provides us with exogenous variation in the student's decision to study abroad. In all regressions reported below we account for any dependence between observations by clustering all results at the university level. This leaves the error correlation within clusters completely unrestricted and allows for arbitrary with-in cluster dependence. The clustering, therefore, not only allows arbitrary correlations of errors for students from a graduate cohort at a certain university but also allows the

[14] This approach is based on our understanding that a school leaver's decision process can be thought of as first deciding on a subject, and then selecting between different universities given the subject. This is reflected, for example, in the subjects where university admissions are centrally administered: students can apply for one subject only, but in their application give a preference ranking for a number of different universities in this subject (ZVS 2009).

[15] This way of defining exposure is related to the instrument of Bartik (1991) for local labor demand conditions.

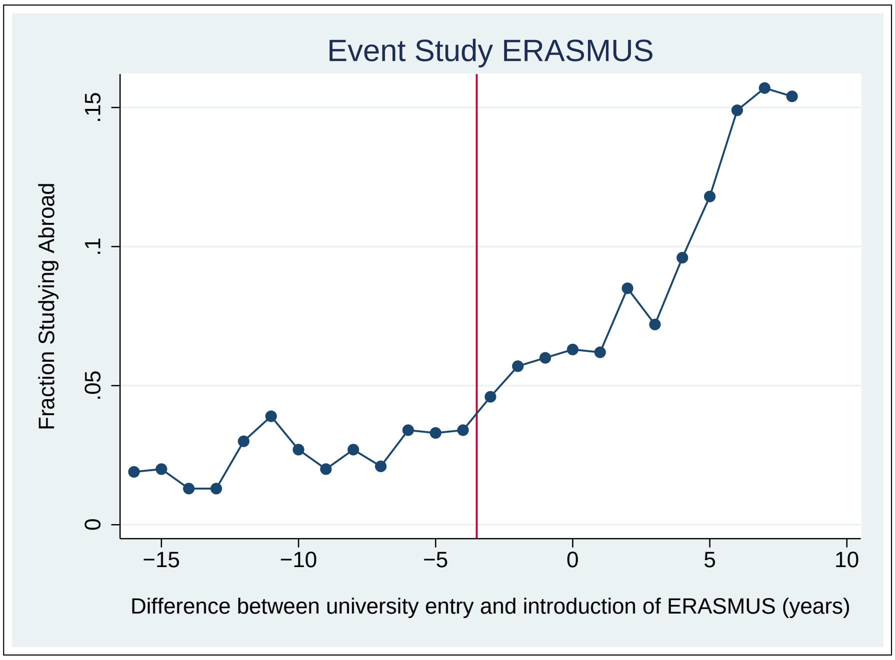
[16] Partly owing to the comparatively long duration of studies, it is more common for students to transfer between universities during the undergraduate studies than e.g. in the U.S. or the U.K.

errors to be serially correlated. An alternative way of addressing the possible serial correlation of error terms is to collapse the data into a pre and post period as suggested by Bertrand, Duflo, and Mullainathan (2004). We show in column (5) of Table 3.10 that this alternative way of obtaining standard errors yields very similar results to clustering at the university level.

In order to visualize how students are affected by these shocks of being faced with more or less exchange opportunity, we perform the following event study: For each student's initial university and subject choice, we observe whether there was at any point an ERASMUS cooperation in the time period we observe. We group students by whether they entered the university before or after the introduction of the ERASMUS scheme, and by how many years. In the following figure we plot the time difference between the introduction of ERASMUS and university entry against the probability of going abroad. Keeping in mind that students usually start two or three years before going abroad, we get the following prediction: According to our hypothesis, the probability of studying abroad should be flat for the cohorts starting more than three years before the introduction. The cohorts starting three or two years before the introduction of ERASMUS would then be the first ones to be affected, and we expect an increase in the proportion of students studying abroad from then on. The results can be seen in Figure 3.5.

This figure provides evidence that the ERASMUS scheme affects the different cohorts in a very precise way. Closely following our prediction, the probability of studying abroad is low and flat before the introduction of ERASMUS, and goes up steeply afterwards. Furthermore, our data provides evidence that institutions which have *not yet* introduced ERASMUS are similar to those which *never* introduce ERASMUS: Students at institutions which never introduce ERASMUS have a probability of studying abroad of 2.2%, which closely matches the average for the not-yet-affected students in the graph above. In the following section we show how the exposure to ERASMUS affects the probability of studying abroad.

Figure 3.5: Event Study ERASMUS



3.5 First Stage Results

Table 3.2 presents the results from our first stage estimates. In this context the first stage regressions are interesting in their own right as one can learn about the factors affecting an individual's decision to study abroad. We regress an indicator for studying abroad on our measure for exposure to the ERASMUS program and other control variables. In column (1) we use the ERASMUS indicator as our measure for a student's exposure to the program. The coefficient on ERASMUS is highly significant with an F-statistic of 40.5. The coefficient indicates that a student's probability of studying abroad increases by about 2.5 percentage points if her department participates in the ERASMUS program. Analyzing the effect of our control variables one can see that a student's gender does not seem to affect her probability of studying abroad. The quadratic in age indicates that students who begin their studies at a higher age are much less likely to study abroad (in the relevant age range).

Table 3.2: First Stage

	(1)	(2)	(3)	(4)
Instrument	Dummy	Ratio	Subject Ratio	Subject Ratio excluding own department
ERASMUS	0.0247	0.4490	0.9121	0.8382
	(0.0039)**	(0.0639)**	(0.1364)**	(0.1297)**
Female	-0.0022	-0.0026	-0.0029	-0.0029
	(0.0034)	(0.0033)	(0.0033)	(0.0033)
Apprenticeship	-0.0013	-0.0012	-0.0012	-0.0012
	(0.0037)	(0.0037)	(0.0037)	(0.0037)
Age (when starting Studies)	-0.0096	-0.0103	-0.0101	-0.0101
	(0.0027)**	(0.0027)**	(0.0027)**	(0.0027)**
Age Squared	0.0001	0.0001	0.0001	0.0001
	(0.0000)*	(0.0000)**	(0.0000)**	(0.0000)**
Experience	0.0001	0.0000	0.0003	0.0001
	(0.0018)	(0.0018)	(0.0018)	(0.0018)
Bachelor	0.0119	0.0123	0.0127	0.0130
	(0.0328)	(0.0318)	(0.0326)	(0.0327)
Follow-up Survey (Dummy)	✓	✓	✓	✓
Graduate Cohort FE	✓	✓	✓	✓
Subject FE	✓	✓	✓	✓
University FE	✓	✓	✓	✓
N	54079	54079	54079	54079
R-squared	0.087	0.092	0.090	0.089
F-stat of Instrument	40.536	49.394	44.688	41.746

Dependent Variable: Indicator for Study Abroad. ** denotes significance at the 1%, * denotes significance at the 5% level. All standard errors are clustered at the university level.

In column (2) we use the ratio of ERASMUS places to the number of students in the relevant cohort as our measure for exposure to the ERASMUS program. Once again the coefficient on the ERASMUS measure is highly significant with an F-statistic of 49.4. The coefficient indicates that an increase in the ratio of ERASMUS places from say 5 percent to 10 percent increases an individual's probability of studying abroad by about 2.2 percentage points. The coefficients for the control variables are very similar to the ones reported in column (1).

In columns (3) and (4) we report the first stage for the ERASMUS subject ratio and the subject ratio, excluding own department. As we would expect, the strength of the instrument is somewhat lower than for the ratio, but the F-statistic is still above

40.[17]

In the following we show that the ERASMUS program has a very specific effect on studying abroad, as it only affects the probability of studying abroad in a European country but not in countries outside Europe. This is a clear indication that the introduction of ERASMUS was not one of many policies to improve university quality, which in turn could affect the outcome as well. In order to demonstrate the precise effect of studying abroad we create three indicator variables, which take the value 1 if an individual studied abroad in Europe, the USA, or in any other foreign country respectively. We expect that our instrument only affects the probability of studying abroad in Europe as the ERASMUS program only offers scholarships for studying abroad in European partner universities. In columns (1) and (4) of Table 3.3 we replace the dependent variable of our usual first stage regression (studying abroad in any country) with an indicator for studying abroad *in Europe*.[18] ERASMUS is a strong and highly significant determinant of studying abroad in Europe. The magnitudes of the ERASMUS coefficients is similar to the one obtained when we use the general definition of studying abroad.

The regressions reported in columns (2) and (5) use an indicator for studying abroad in the US as the dependent variable. The coefficients on the ERASMUS measures is not significantly different from 0. Furthermore, the point estimates of the ERASMUS measures are very close to 0. In columns (3) and (6) we report specifications where we use an indicator for studying abroad in any country outside Europe or the US as the dependent variable. Again the coefficients on ERASMUS are small and not significantly different from 0. The evidence from Table 3.3 strongly suggests that the introduction of the ERASMUS program was not correlated with the introduction of

[17]One common concern in IV estimation is a potential bias due to weak instruments (see Bound, Jaeger & Baker (1995) and Stock, Wright and Yogo (2002)). The F-statistics from the first stage, reported at the bottom of Table 3.2, show that weak instruments are not likely to pose a problem in our analysis.

[18]We do not observe study abroad destinations in the 1989 cohort, so that our sample in this analysis is correspondingly smaller.

Table 3.3: Falsification Exercise: First Stage with Different Destinations

	(1)	(2)	(3)	(4)	(5)	(6)
Instrument	Dummy	Dummy	Dummy	Ratio	Ratio	Ratio
Study Abroad in	Europe	USA	Rest	Europe	USA	Rest
ERASMUS	0.0200	-0.0016	0.0013	0.3861	0.0102	0.0281
	(0.0036)**	(0.0018)	(0.0012)	(0.0597)**	(0.0156)	(0.0144)
Controls						
Follow-up Survey (Dummy)	✓	✓	✓	✓	✓	✓
Graduate Cohort FE	✓	✓	✓	✓	✓	✓
Subject FE	✓	✓	✓	✓	✓	✓
University FE	✓	✓	✓	✓	✓	✓
N	41065	41065	41065	41065	41065	41065
R-squared	0.075	0.023	0.039	0.080	0.023	0.039
F-stat of Instrument	30.80	0.77	1.18	41.83	0.43	3.79

Dependent Variable: Indicator for Study Abroad in a certain area. ** denotes significance at the 1%, * denotes significance at the 5% level. All standard errors are clustered at the university level.

a broader set of policies, which might themselves affect later labor market outcomes. These results increase our confidence for using the ERASMUS program as an instrumental variable for studying abroad. In the following section we use this IV to obtain estimates of the effect of studying abroad on the probability of working in a foreign country later in life.

3.6 Main Results and Sensitivity Analysis

The OLS results reported in column (1) of Table 3.4 confirm that graduates who spent some time at a foreign university are more likely to work abroad later in life. Our OLS result indicates that the effect of studying abroad is about 6.5 percentage points. As discussed before we do not want to attribute causality to the OLS results. This is because the factors affecting an individual's decision to study abroad are likely to affect her decision to work abroad later on as well. Therefore, we now turn to our IV results.[19]

Table 3.4: Main Results

	(1)	(2)	(3)	(4)	(5)
Estimation Method	OLS	IV	IV	IV	IV
Abroad	0.0646	0.2439	0.1224	0.1488	0.1346
	(0.0066)**	(0.1078)*	(0.0450)**	(0.0598)*	(0.0671)*
Female	-0.0002	0.0002	-0.0001	-0.0000	-0.0000
	(0.0020)	(0.0021)	(0.0020)	(0.0020)	(0.0020)
Apprenticeship	-0.0051	-0.0049	-0.0050	-0.0050	-0.0050
	(0.0023)*	(0.0024)*	(0.0023)*	(0.0023)*	(0.0023)*
Age (when starting Studies)	-0.0052	-0.0035	-0.0046	-0.0044	-0.0045
	(0.0018)**	(0.0022)	(0.0018)*	(0.0019)*	(0.0019)*
Age Squared	0.0001	0.0001	0.0001	0.0001	0.0001
	(0.0000)*	(0.0000)	(0.0000)*	(0.0000)*	(0.0000)*
Experience	0.0067	0.0067	0.0067	0.0067	0.0067
	(0.0012)**	(0.0012)**	(0.0012)**	(0.0012)**	(0.0012)**
Bachelor	-0.0013	-0.0033	-0.0020	-0.0023	-0.0021
	(0.0097)	(0.0096)	(0.0092)	(0.0092)	(0.0092)
Follow Up Survey (Dummy)	✓	✓	✓	✓	✓
Graduate Cohort FE	✓	✓	✓	✓	✓
Subject FE	✓	✓	✓	✓	✓
University FE	✓	✓	✓	✓	✓
Instruments:					
ERASMUS		Dummy	Ratio	Subject Ratio	Subject Ratio excluding own department
N	54079	54079	54079	54079	54079
R-squared	0.034				
F-stat First Stage		40.537	49.394	44.688	41.746

Dependent Variable: Working Abroad. ** denotes significance at the 1%, * denotes significance at the 5% level. All standard errors are clustered at the university level. Note: Dependent variable is an indicator for whether the respondent works abroad at the time of the survey. *Study abroad* is an indicator for whether the student spends part of her university career at a foreign university. See text for further details.

In column (2) of Table 3.4 we present the first set of IV results using the ERASMUS indicator as an instrument. We find that studying abroad increases an individual's probability to work in a foreign country by about 24 percentage points. The effect is significant at the five percent level. We find no significant difference in terms of gender.

[19]In Table 3.10 in the Appendix 3.A, we also present the reduced form estimates corresponding to the main results. Column (5) of Table 3.10 shows the results from collapsing the data into a pre and a post period as suggested by Bertrand, Duflo, and Mullainathan (2004) using the ERASMUS dummy measure. The corresponding uncollapsed reduced form is presented in column (1). The results are very similar.

Furthermore, we find that individuals who completed an apprenticeship before they enrolled at university are about 0.5 percentage points less likely to work abroad. People who complete an apprenticeship may be more likely to go back to work at the same firm where they completed their apprenticeship, which will usually be located in Germany. We also find that labor market experience has an effect on the probability of working abroad. The coefficient indicate that individuals with one more year of experience in the labor market are about 0.7 percentage points more likely to work abroad. Within a survey wave, there is relatively little variation in potential experience, and this estimate also captures the increased probability of working abroad from the initial to the follow-up survey. Over and above this annual measure of potential experience, the indicator variable for the follow-up survey does not show up significantly.

In column (3) we present the results from using the ERASMUS ratio as instrument. Making use of the additional variation in number of scholarships increases precision significantly. The point estimate goes down as well compared to column (2), but is still substantially higher than the OLS estimate. The effect is statistically significant at the one percent level. It is important to note that the point estimate is highest when we use the ERASMUS indicator. Given these results we are confident to say that our results reflect a supply-side increase in scholarship availability, rather than students' demand. If the number of ERASMUS places was driven by the demand of very motivated students we would expect higher coefficients on ERASMUS when using the ERASMUS ratio instrument.

We further probe our results by using the ERASMUS measures which exploit subject level variation rather than conditions at the actual department (columns (4) and (5)). It is reassuring to find that the estimates are similar to the ones reported in the previous columns. In the following, we show that our results are robust to a number of specification checks.

There may be a worry that students from different family backgrounds not only choose universities with different provision of ERASMUS scholarships but also exhibit different propensities to work in a foreign country. As long as this effect is constant over time we deal with this problem by estimating all equations including university fixed effects. It could be possible, however, that people from different backgrounds react differently to the introduction of an ERASMUS program or changes in the number of scholarships. In order to address this concern we add controls for parental education and occupation to our main specification. It is evident from looking at the second panel of Table 3.5 that including the measures for parental background hardly affects our estimates of the effect of studying abroad. The results indicate that students from better educated parents are between 0.5 and 1 percentage points more likely to work abroad.

Another concern is that students with a taste for mobility choose universities or departments with a lot of ERASMUS scholarships. Our IV estimates would be biased if these individuals were more likely to work abroad later in life. In the following we present a powerful test, which directly addresses this concern. We add two variables which control for a student's mobility at the start of her university career. The first variable indicates whether the student enrols in university in the state *(Bundesland)* where she obtained her highschool diploma *(Abitur)*. The second mobility variable measures the distance from the state where she obtained her highschool diploma to the state of her first university enrolment. The coefficients on the distance measures for early mobility are not found to be significant. Including those two mobility variables hardly affects the estimates for the effect of studying abroad as can be seen from the third panel in Table 3.6.

Individuals may be more likely to work abroad if they know more foreigners. There are at least two channels through which the number of contacts to foreigners may affect the likelihood of working abroad. One channel may be an increased number of contacts to future business partners. A further channel may be that contacts to foreigners increase an individual's taste for foreign cultures which may affect her probability of working abroad. As the ERASMUS program is at least partly reciprocal, universities

Table 3.5: Sensitivity Analysis 1 (Further Controls)

		(1)	(2)	(3)	(4)	(5)
Estimation Method		OLS	IV	IV	IV	IV
Main Specification	coefficient	0.0646	0.2439	0.1224	0.1488	0.1346
	(st. err.)	(0.0066)**	(0.1078)*	(0.0450)**	(0.0598)*	(0.0671)*
	F-stat 1st stage		*40.537*	*49.394*	*44.688*	*41.746*
Including Parental Background	coefficient	0.0628	0.2406	0.1209	0.1447	0.1300
	(st. err.)	(0.0066)**	(0.1101)*	(0.0452)**	(0.0602)*	(0.0673)
	F-stat 1st stage		*39.376*	*49.733*	*43.890*	*41.285*
Adding Controls for Early Mobility	coefficient	0.0618	0.2407	0.1178	0.1413	0.1255
	(st. err.)	(0.0066)**	(0.1101)*	(0.0456)**	(0.0608)*	(0.0683)
	F-stat 1st stage		*39.884*	*49.010*	*43.551*	*40.565*
Further adding foreign students at home university	coefficient	0.0618	0.2406	0.1176	0.1407	0.1248
	(st. err.)	(0.0066)**	(0.1103)*	(0.0457)**	(0.0609)*	(0.0684)
	F-stat 1st stage		*39.780*	*48.935*	*43.356*	*40.335*
Follow Up Survey (Dummy)		✓	✓	✓	✓	✓
Graduate Cohort FE		✓	✓	✓	✓	✓
Subject FE		✓	✓	✓	✓	✓
University FE		✓	✓	✓	✓	✓
Instruments: ERASMUS			Dummy	Ratio	Subject Ratio	Subject Ratio excluding own department
N		54079	54079	54079	54079	54079

Dependent Variable: Dummy for Working Abroad. ** denotes significance at the 1%, * denotes significance at the 5% level. All standard errors are clustered at the university level. Note: This table only shows results for the coefficient of interest, studying abroad. Regressors not listed include female indicator, apprenticeship, potential experience. Panel 1 reports the results from Table 3.4. Panel 2 adds controls for parental education and parental occupation. Panel 3 adds controls for previous mobility. The specifications reported in Panel 4 add further controls for the ratio of foreign students at the home university. See text for details.

Table 3.6: Sensitivity Analysis 2 (Time Trends and Additional FE)

		(1)	(2)	(3)	(4)	(5)
		OLS	IV	IV	IV	IV
Main specification	coefficient	0.0646	0.2439	0.1224	0.1488	0.1346
	(st. err.)	(0.0066)**	(0.1078)*	(0.0450)**	(0.0598)*	(0.0671)*
	F-stat 1st stage		*40.537*	*49.394*	*44.688*	*41.746*
Including Subject-Specific	coefficient	0.0639	0.2550	0.1023	0.1447	0.1261
Time Trends	(st. err.)	(0.0065)**	(0.1288)*	(0.0558)	(0.0865)	(0.1062)
	F-stat 1st stage		*29.915*	*35.184*	*20.813*	*15.619*
Including	coefficient	0.0638	0.2231	0.1408	0.1946	0.1805
University * Subject group FE	(st. err.)	(0.0067)**	(0.1243)	(0.0630)*	(0.0675)**	(0.0770)*
	F-stat 1st stage		*39.155*	*35.220*	*43.391*	*36.864*
Graduate Cohort FE		✓	✓	✓	✓	✓
Subject FE		✓	✓	✓	✓	✓
Instruments: ERASMUS			Dummy	Ratio	Subject Ratio	Subject Ratio excluding own department
N		54079	54079	54079	54079	54079

Dependent Variable: Dummy for Working Abroad. ** denotes significance at the 1%, * denotes significance at the 5% level. All standard errors are clustered at the university level. Note: This table only shows results for the coefficient of interest, studying abroad. Regressors not listed include female indicator, apprenticeship, potential experience. Panel 1 reports the results from Table 3.6. Panel 2 adds time trends for each subject to the main specification. In these regression we also include university FE as before. Panel 3 adds Fixed Effects at the university*subject group level to the main specification. The specifications reported in Panel 3 do not include university FE because we use the finer level of university*subject group FE. See text for details.

offering more ERASMUS scholarships may also enroll more foreign students. This could then increase the student's propensity to work abroad later on and therefore bias our IV results. In the fourth panel of Table 3.5 we present the results from adding the university wide ratio of foreign students over the total number of students in a student's cohort[20] to our specification. Adding this control does not change the coefficient on studying abroad at all. The coefficient on our measure for the exposure to foreign students is highly significant, but rather small in magnitude. The estimated coefficient indicates that increasing the percentage of foreign students at a student's home university from say 5 to 15 percent increases her probability of working abroad by about 0.03 percentage points. This exercise is interesting also because it adds university-specific covariates which vary over time, and it is reassuring that the results remain unchanged.

In the following we check whether our results are driven by time trends in our variables of interest. Including graduate cohort FE (as in all specifications) guarantees that we do not identify the effect of studying abroad on working abroad from overall time trends. There may be a worry, however, that students studying certain subjects exhibit time trends in both studying abroad and working abroad. To address this issue we include linear subject specific time trends. The results of this exercise are reported in the second panel of Table 3.7. Apart from the specification reported in column (3) the inclusion of the subject specific time trends hardly affects the coefficient of studying abroad.

It may be the case that groups of departments within a university differ in quality or in their ability to foster international exchange. We address this concern by including a full set of department group times university fixed effects. We thus use separate fixed effects for say sciences or languages at a certain university. Including this fine level of FEs hardly affects the estimates using the ERASMUS ratio instrument. Again, we

[20]We use the ratio at the middle of the average student's university career as the relevant measure for contacts to foreigners.

Table 3.7: Destinations of International Mobility

		Work Abroad Location														
		GB Ire-land	France	Spain	Bene-lux	Scan-dina-via	Austria Switz.	East. Euro.	Other Euro.	USA	Other Amer.	Asia	Aus-tralia NZ	Other	Total	% of total studying abroad
Not Studying Abroad		15	4	5	8	5	41	4	2	7	2	2	3	2	100	
	GB/Ireland	**46**	0	0	15	0	31	0	0	8	0	0	0	0	100	8
	France	3	**50**	0	12	0	29	3	0	0	3	0	0	0	100	21
	Spain	5	0	**48**	5	10	29	0	0	0	0	0	0	5	100	13
	Benelux	20	0	20	**60**	0	0	0	0	0	0	0	0	0	100	3
	Scandinavia	9	4	0	9	**52**	17	0	0	9	0	0	0	0	100	14
Study	Austria/Switzerland	0	0	0	9	0	**82**	0	0	9	0	0	0	0	100	7
Abroad	Eastern Europe	0	0	25	25	0	0	**50**	0	0	0	0	0	0	100	2
in:	Other Europe	0	11	11	11	0	33	0	**33**	0	0	0	0	0	100	6
	USA	10	5	0	5	5	38	0	0	**24**	5	10	0	0	100	13
	Other America	0	0	0	20	20	20	0	0	0	**40**	0	0	0	100	3
	Asia	14	0	0	0	0	29	0	0	0	0	**57**	0	0	100	4
	Australia/ NZ	14	0	0	0	0	43	14	0	0	14	0	**14**	0	100	4
	Other	0	0	0	0	0	0	0	0	0	0	0	100	**0**	100	1
% of total working abroad		13	7	6	9	6	38	3	2	6	3	2	3	2		

Note: For all graduates working abroad, this table shows conditional probabilities of working abroad in the different countries conditional on the study abroad treatment and the destination of the stay abroad. Based on 496 observations from graduate cohort 2005.

find that the magnitude of the estimates to be very similar to our main specification.

It is reassuring that the inclusion of time trends or a finer set of fixed effects does not have a huge impact on our estimates. This and the fact that our estimates are hardly affected by including controls for parental background, for early mobility, and for the number of foreign students at the home university makes us confident that using the ERASMUS program as a source of exogenous variation is a credible identification strategy to estimate the causal effect of studying abroad on later labor market mobility.

One defining feature of our results is that the IV results are substantially higher than the corresponding OLS result. We have tested whether the IV estimates are significantly different from the OLS estimates. The results indicate that the IV estimates are not statistically different at a 5% level, which reflects the lower precision of the IV estimates.[21] Although the difference in the estimates needs to interpreted in the light of this test, we are nonetheless interested in understanding why our point estimates are consistently higher than the OLS estimates, and we interpret this finding in terms of heterogeneity in returns: It is unlikely that all students will be affected in the same way by the intervention of studying abroad. It is much more plausible that the effect of studying abroad itself varies across the student population. We follow Imbens & Angrist (1994) and interpret our estimates as a Local Average Treatment Effect (LATE): The IV results show the average effect for the subgroup which has been affected by the instrument. In the context of our instrument, this group is well-defined: It is the group of students who would not have studied abroad without the ERASMUS program, but study abroad when the ERASMUS is implemented. Since they are the students who have been affected by the ERASMUS program, our estimates are of immediate interest to policy makers.

We therefore investigate heterogeneity in returns along two important dimensions: parental education and whether the student was credit constrained during her studies.

[21] In our main specification, the IV dummy specification is statistically different from the OLS estimate at 10%.

Parental education may be important because students with better educated parents may be better informed about the benefits from studying abroad. Furthermore, we investigate heterogeneity according to the financial situation of the student. In the absence of credit constraints all students for whom the cost of studying abroad is above the returns from studying abroad will not study in a foreign country. Some credit constrained students, however, will not be able to invest in studying abroad even though this investment offers a positive return. The introduction of ERASMUS can be understood as a price change which makes the investment into studying abroad worthwhile for these marginal students.

In order to investigate heterogeneity in returns we therefore split our sample into four different subgroups: students with high parental education who have not been credit constrained, students with high parental education who have been credit constrained, students with low parental education who have not been credit constrained, and lastly the most disadvantaged group: students with low parental education who have been credit constrained. We classify students to being from a high parental education background as those whose parents have at least 16 years of education, i.e. both parents have at least a university degree. Low parental education is defined as all those with parents who have less than 16 years of education. Credit constraints are proxied with an indicator variable which takes the value 1 if the student covers 50 percent or more of his expenses with federal financial assistance (BAFOEG).[22] We then follow Kling (2001) in interpreting the IV estimate as a weighted average of the causal effect of studying abroad, where the weight of each subgroup j is given by the following formula:

[22]Unfortunately, we do not have any information on the student's financial situation for the 2005 wave. In 1989 the question on BAFOEG was not administered but the students were asked to evaluate their financial situation on a 1 to 5 scale. We classify all those who answered 5 (unsatisfactory financial situation) as being credit constrained. This corresponds almost exactly to the sample proportion who indicate that they financed 50 percent of their expenses with BAFOEG in the later cohorts.

$$weight_j = \frac{w_j \lambda_j \Delta(StudyAbroad)_j}{\sum_j w_j \lambda_j \Delta(StudyAbroad)_j} \tag{3.6}$$

Here w_j is the sample fraction of each subgroup j, λ_j is the variance of the instrumental variable for subgroup j conditional on all other regressors x, and $\Delta(StudyAbroad)_j$ is the impact of the ERASMUS instrument on the probability of studying abroad for subgroup j. The last term is obtained from estimating the first stage regression separately for each subgroup.[23] We use this decomposition to compute the corresponding weight for our four subgroups.

In our sample about 39% of all students come from the most advantaged background (see column (1) in Table 3.8), and this group is found to respond strongest to the introduction of ERASMUS (see column (2)). Even though the conditional variance of ERASMUS is lowest for them (column (3)) they contribute about 46% to the final IV estimate which is more than their sample proportion. The other group that contributes more than proportionately to the IV estimates is the group of students with the most disadvantaged background. Column (5) reports the corresponding IV estimates if the regression is estimated separately for the four subgroups. The much smaller samples lead to a loss in precision; comparisons of the point estimates for the four subgroups should therefore be made with caution. With this caveat in mind it is evident that the least advantaged group of students seems to have the highest return from studying abroad. This suggests that credit constraints and information asymmetries may indeed prevent some students from realizing the return from studying abroad.

[23] See Kling (2001) for further details.

Table 3.8: Heterogeneity in Returns

	(1)	(2)	(3)	(4)	(5)	(6)
	Fraction in Sample w_j	Delta (First Stage) $\Delta(StudyAbroad)_j$	Lambda λ_j	Kling Weight	IV (Ratio)	se(IV)
Not Credit Contrained/ High Parental Edu	0.39	0.027	0.083	0.46	0.178	(0.117)
Not Credit Contrained/ Low Parental Edu	0.49	0.014	0.119	0.43	0.085	(0.114)
Credit Contrained/ High Parental Edu	0.05	0.006	0.118	0.02	0.154	(0.234)
Credit Contrained/ Low Parental Edu	0.07	0.019	0.129	0.09	0.341	(0.163)*

Credit constrained = 1 for students who defray 50 percent or more of their living costs with federal financial aid for the 1993 to 2001 waves. The question was not administered in the other waves. For the 1989 wave credit constrained students are all students who indicate that their financial situation during their studies was unsatisfactory. High parental education = 1 for students whose less educated parent has at least 16 years of schooling (has at least an undergraduate degree).

3.7 How Studying Abroad Affects International Labor Market Mobility

The results presented in the previous sections indicate that individuals who study abroad are more likely to work in a foreign country. It is interesting to understand how studying abroad affects an individual's decision to migrate to a foreign country later in life. We address this in two ways: First, we make use of observed location choices to study the type of skills acquired during the stay abroad. Second, the survey provides us with direct qualitative evidence on why graduates move abroad, and we show how this varies depending on whether the student studied abroad earlier.

We can think of the effect of studying abroad as affecting the set of skills the student acquires during her studies. One important question is whether these skills have a strong location-specific component. We can shed some light on this question by investigating whether individuals who have studied abroad return to work in the same country when they decide to work in a foreign country. There are a number of reasons why mobile graduates may be more likely to work abroad in the countries where they studied abroad before: While they were studying abroad they may have obtained skills that are of particular relevance in that specific labor market, e.g. language skills, knowledge about the local labor market, or personal contacts which facilitate a match. On the other hand, it is possible that studying abroad affects the probability of working abroad equally for different work destinations. This would be the case, for example, if studying abroad widens the horizon of the student generally and leads her to search for a job internationally, independent of where she studied before. Especially, studying abroad could operate as a stepping stone to increase the set of feasible destinations. This question is also highly relevant from a policy perspective: The ability of the ERASMUS scheme or other student mobility programs to achieve an integrated European labor market depends on the assumption that students who went abroad to study in Europe are internationally mobile after graduation, but remain in Europe.

Here we present descriptive evidence to address this question from the 2005 cohort.[24] For each study abroad treatment and study abroad location, Table 3.7 shows the conditional probability of being in each work location. The table provides evidence that choices about study abroad locations are sticky, that is that students tend to return to work to the country or region where they studied abroad. A χ^2-test of independence between the study abroad location and the work abroad location is rejected at the 0.01% level with a test statistic of 768.7.

We now turn to qualitative evidence from the survey on why graduates moved abroad. As these qualitative questions were only administered to the 1997 cohort we cannot apply our instrumental variable strategy here. We therefore provide a descriptive analysis, which – if only suggestive – may shed light on the way studying abroad affects later labor market mobility.

Graduates who had worked in a foreign country for at least one month in the five years since graduation were asked to identify the reasons for their decision to work abroad. In Table 3.9 we present the percentage of the people who indicated that a certain reason had been important in their decision to work abroad. The table shows that the main reasons for working abroad are interest in foreign cultures, interesting offers from abroad, and the initiative of the employer. We split the sample into those who complete all their university education in Germany and those who study abroad for some time during their undergraduate education. Interestingly, while the means are similar in some categories, there are a number of noteworthy differences. Those who have studied abroad are more likely to indicate that their interest in foreign cultures has led them to seek employment abroad. It may be the case that studying in a foreign country increased the individual's taste for living abroad, which may in turn increase her probability of migrating later in life. Students who have studied abroad are also significantly more likely to indicate that they chose to work abroad to be with their

[24]We only observe country by country locations for the studying abroad and the work abroad spell for the latest cohort.

Table 3.9: Reasons for working abroad

	All	Study Abroad = 0	Study Abroad = 1	Difference in means (p-value)
Interest in Foreign Cultures	52.95	50.93	67.21	0.000
	(1.59)	(1.71)	(4.27)	
Received Interesting Offer	35.85	35.35	39.34	0.389
	(1.53)	(1.63)	(4.44)	
At Employer's Instance	33.40	34.07	28.69	0.239
	(1.51)	(1.62)	(4.11)	
Better Career Prospects	25.36	25.81	22.13	0.382
in Germany after Return	(1.39)	(1.49)	(3.77)	
Obtain Qualifications Abroad	16.80	16.86	16.39	0.897
	(1.19)	(1.28)	(3.37)	
International Research Project	14.77	14.65	15.57	0.788
	(1.13)	(1.21)	(3.30)	
Partner	10.90	9.77	18.85	0.003
	(0.99)	(1.01)	(3.56)	
Employment Outlook Abroad	8.66	8.02	13.11	0.061
	(0.90)	(0.93)	(3.07)	
Career Prospects Abroad	6.52	5.70	12.30	0.006
	(0.79)	(0.79)	(2.99)	
Number of Observations	982	860	122	

Note: Based on all respondents from the 1997 follow-up survey who have work experience abroad. Table shows percentage of respondents who indicate that a particular reason led them to take up work abroad. Example: 50.93% of respondents indicate that interest in foreign cultures led them to take up work abroad.

partner. The answers to this question may suggest that people who studied abroad may have met their partner while studying abroad and therefore consider to work abroad later in life. Of course, this difference may also be driven by assortative mating with more mobile people having more mobile partners, and the way this question was asked makes it impossible to distinguish between these alternatives. Meeting a partner abroad may, nonetheless, be a possible channel of the effect of studying abroad. The summary statistics also indicate that those who have studied abroad are somewhat more likely to say that they work abroad because of better employment opportunities in the foreign labor market, where we obtain a p-value of 0.06 when we test for a significant difference in the means of the two groups for this response. It is possible that a stay at a foreign university makes it easier to realize opportunities in foreign labor markets, either because those who studied abroad have better information on the foreign labor market or because employers are more willing to offer employment to those individuals. Interestingly, rather than the employment outlook, it is the career prospects abroad where the means are significantly different at the 1% level, suggesting that those with international study experience seem to be more likely to consider a career abroad.

The statistics presented here provide some suggestive evidence of how studying abroad may alter later international labor market mobility. Further research is necessary to get a better insight into the channels of the effect of studying abroad on working abroad later on.

3.8 Conclusion

Using exogenous variation in scholarship availability, we are able to identify a causal effect of undergraduate student mobility on later international labor migration. Our strategy exploits the introduction and expansion of the ERASMUS scholarship program. The extent to which students were exposed to the scholarship scheme varied widely. We exploit cross-sectional and over time-changes in scholarship availability. Accounting for permanent differences between different institutions, different subjects, and different graduate cohorts, our identification relies only on differential over-time change, and can be interpreted as a Diff-in-Diff estimator. Our first-stage shows that the ERASMUS scheme has indeed a strong effect on the students' decision to go abroad, which is not surprising given its scale. We show that the instrument is precise in that it only affects the decision to study in Europe, but not in other locations. Our event study adds further credibility to our instrument, by showing that the probability of studying abroad is low and flat before ERASMUS is introduced, and increases strongly for those students affected by the scholarship.

Our OLS results indicate that the group of students who studied abroad are about 6 percentage points more likely to work abroad later on, controlling for a set of background characteristics, institution and time fixed effects. Our IV results are substantially higher than that, and indicate that the effect of study abroad is about 15 percentage points. We interpret the difference between OLS and IV as an indication of heterogeneity in effects: The population which is affected by our instruments reacts particularly strongly to the incentives of the mobility program. This Local Average Treatment Effect (LATE) interpretation is of particular interest to policy makers, since it evaluates the effect for the affected sub group. We show that the most disadvantages students have the highest returns from studying abroad suggesting that credit constraints and information asymmetries play a role in this setup.

Our results show that educational mobility programs may have a potentially large role in affecting students' behavior in their labor market mobility decision. These results imply that an opportunity to attract talented graduates is to provide student exchange opportunities. Attractive universities and scholarship programs may yield a return through attracting students, part of whom will remain as skilled workers later on. In the context of the policy change under consideration, ERASMUS is successful in that this student mobility scheme appears to contribute to the development of an integrated European labor market. This is especially so if we take into account the descriptive evidence from the previous section that location choices are sticky, i.e. that mobile students tend to return to the country where they studied before.

More generally, our work allows insights into the dynamic implications of educational mobility decisions. Our results indicate that the effects of educational mobility programs go far beyond affecting the decision to study abroad for some time period, but rather reach far into the labor market, and it will be interesting to follow the sample of graduates as their careers unfold. But already at this early stage our results indicate that even short-term mobility investments can lead to significant further mobility investments later on.

3.A Appendix

Table 3.10: Reduced Forms

	(1)	(2)	(3)	(4)	(5)
	OLS	OLS	OLS	OLS	BDM
ERASMUS	0.0060 (0.0027)*	0.0550 (0.0221)*	0.1357 (0.0573)*	0.1128 (0.0583)	0.0065 (0.0030)*
Instruments: ERASMUS	Dummy	Ratio	Subject Ratio	Subject Ratio excluding own department	Dummy
N	54079	54079	54079	54079	

Dependent Variable: Indicator for Working Abroad. ** denotes significance at the 1%, * denotes significance at the 5% level. Standard errors in columns (1) to (4) are clustered at the university level. In column (5) the data is first collapsed into pre-post subject times university cells as suggested by Bertrand, Duflo, and Mullainathan (2004).

Chapter 4

Vocational Schooling versus Apprenticeship Training: Evidence from Vacancy Data

4.1 Introduction

Recent research has emphasized the occupation-specific nature of human capital (Johnson and Keane, 2007; Kambourov and Manovskii, 2008), indicating that human capital is encapsulated in the ability to perform specific tasks. This suggests that the way young labor market entrants are taught the skills they need in the workplace is crucial for their labor market outcomes. At the same time, little is known about how these skills are best conferred. Alternative *templates* (Goldin, 2001) compete with respect to structured vocational training: full-time vocational schooling, largely firm-based apprenticeships[1], and on-the-job training. The co-existence of these alternatives, and the pronounced differences between countries in the approach to vocational training,

[1] Ryan (1998) defines *apprenticeship* as 'employer-sponsored programmes which integrate part-time schooling with part-time training and work experience on employers' premises [...] within an externally defined curriculum which contains mandatory part-time schooling, leads to a nationally recognised vocational qualification and takes at least two years to complete'.

as documented by Ryan (2001), underlines that no consensus has been reached on how to best equip young people with the skills they are likely to need in the workplace.

In this chapter, we compare labor market outcomes between apprenticeship training and full-time vocational schooling, focusing on wages, unemployment and measures of mobility. As identifying source of variation, we exploit the following idea: The apprenticeship system fundamentally links the educational opportunities of young people to the provision of apprenticeship places by firms. Conceptually, the same individual will make different educational choices, depending on where and when she grows up and the corresponding fluctuations in apprenticeship places. Using unique data on apprenticeship vacancies from Germany, together with detailed panel data on labor market outcomes, we document how apprenticeship choice is affected by the availability of apprenticeship places. We show that at the margin, young people substitute between apprenticeship-based training and full-time vocational schooling, rather than between apprenticeship and direct entry as unskilled worker. Thus, the variation we exploit is informative about the relative effect of apprenticeship versus schooling-based training. We then employ this variation in the opportunities of young people as instrumental variable to learn about the causal effect of the apprenticeship scheme. To motivate this instrumental variable, we provide a simple small open economy model with educational choice, in which aggregate price shocks affect the local number of apprentices, but have no differential effect on factor rewards.

Our main results indicate that vocational schools and apprenticeship training provide similar levels of productivity as measured by wages in the age range between 23 and 26. This suggests that these two alternatives are similar in the skills they confer. At the same time, the probability of unemployment is substantially lower for apprenticeship graduates. Investigating the pattern of unemployment over time, we find that the effect is transitory, and fades out over time. This suggests that apprenticeship training provides a benefit to participants in that it improves labor market attachment early in their career.

We perform our analysis on data for Germany, where full-time vocational schools exist as alternative next to the dual apprenticeship system[2]. This allows us to investigate the relative return in a within-country framework. The early tracking of pupils allows us to abstract from the college-going decision: As we describe below, individuals are tracked at ages 10–12 into either a university-bound upper track or a lower- or medium-schooling track, so that the decision to go to university is already pre-determined through the tracking decision earlier on. With respect to alternative entry as unskilled worker, we treat this as an empirical question, and document in several ways that our measure of apprenticeship availability moves individuals between apprenticeship training and full-time vocational school.

In a policy context, understanding the implications of these different templates is crucial for a number of reasons. Given the increased demand for skilled labor, a well-trained workforce is believed to be central to a productive and competitive economy. In many countries, governments and individuals invest heavily into vocational training schemes, and it is important to know if this money is well spent or could be better invested elsewhere. Vocational schooling plays a large role in many countries: On average across OECD countries, 48% of youth are enrolled in vocational or pre-vocational programs at upper secondary level, of which about a third is a combination of school- and work based programs (OECD, 2008). There is wide variation among countries: In some countries, formal vocational training is entirely or mostly school-based (e.g. Sweden, Belgium), while in others, firm-based programs play an important role (e.g. Denmark, Germany, Switzerland). In the U.S., 16% of high school graduates obtain more than a quarter of their credits in career/technical education (CTE) courses[3], and vocational training plays an important role in community colleges.

Young adults who are not college-bound benefit from knowledge about the effects of taking alternative paths, and from the provision of the most effective training scheme.

[2]The term *dual* refers to the shared provision of training through both the firm and a part-time vocational school, in which the student typically spends one to two days per week.

[3]Source: NCES (2008).

From a social policy perspective, a well-functioning school-to-work transition can avoid potentially damaging unemployment or inactivity and the social problems associated with that. It is not surprising that this transition is often found at the heart of policy proposals. In a number of cases, policy initiatives focus on fostering the role of firms in training young adults through apprenticeship–type programs, as for example in the School-to-Work Opportunity Act of 1994 in the US, or the 1995 Modern Apprenticeship program in the United Kingdom.

Investigating the effects of apprenticeship-type training is of broader importance. The apprenticeship system shares important features with other institutions in different countries and at different educational levels: As Becker (1962) points out, there are essential similarities between apprenticeships and the training of lawyers or physicians. College students around the world attempt to gain practical experience through internships. These activities come at substantial costs to individuals, who work at low or sometimes without pay, and often increase duration of their studies. In the United States, many colleges offer Cooperative Education programs.[4] In a 1996 representative survey of 500 U.S. colleges and universities, the American Council on Education (ACE) finds that 91% of institutions offer unpaid internships, 69% offer paid internships, and 57% offer cooperative education programs (NCCE, 2008b).

Empirically, establishing which of the different templates for vocational skill formation is most effective is difficult because in countries where alternatives coexist, individuals select into the different paths based on individual unobservable characteristics and preferences. Simple comparisons of means between the different paths are likely to be misleading because these characteristics affect labor market outcomes at the same time. These selection problems are well known. Ryan (2001, p.74) highlights these challenges with respect to vocational schooling and firm-based training and concludes that 'a large microeconometric evaluation literature is correspondingly

[4]NCCE (2008a) defines *Cooperative Education* as a 'structured educational strategy integrating classroom studies with learning through productive work experiences in a field related to a student's academic or career goals.'

uninformative'. As a result, in trying to understand the implications of the different templates, large emphasis has been placed on evidence from comparative studies. In the comparison of apprenticeship with vocational schooling, the within-country studies which do address selection have regularly relied on excluding family background variables to identify the model, which is difficult to reconcile with the evidence that parental characteristics have a direct effect on a range of parental investments and child outcomes.[5]

Bonnal, Mendes, and Sofer (2002) and Winkelmann (1996) study the transitions immediately after completion of the training period, and find that apprentices are less likely to transit into unemployment. Sollogoub and Ulrich (1999) find that 4.5 years after graduation, apprentices have lower wages (after correcting for selection), but have spent a larger fraction of this period in work. Plug and Groot (1998) find that earnings and earning growth are not statistically different. Blanchflower and Lynch (1994) use the NLSY to estimate the effect of different forms of training on wage growth in a first-difference framework between ages 20 and 25.

This chapter makes a number of contributions. The small open economy framework with educational choice provides an economic setting which generates the exclusion restriction that is required for an instrumental variable strategy. We discuss identification in a multinomial choice setting, and argue that a univariate instrument may still recover a well-defined alternative-specific treatment effect in a potentially important special case. We show that in this application, we cannot reject that this condition is satisfied. This allows us to account for selection in a transparent manner, and identify an effect along a clearly defined margin that is of interest to policy-makers. We trace out the differential effect of training form along a number of important labor market outcomes. The panel nature of our administrative data allows to follow individuals for longer than typically possible in school-to-work transition surveys, and provides us

[5]See, for example, Haveman and Wolfe (1995), Currie (2007), and the evidence presented in chapter 2 of this thesis.

with a large representative sample. This work also empirically investigates the differential responsiveness to negative shocks between apprenticeships and vocational schools, which has been an influential argument in the literature which compares different forms of vocational preparation.

A number of recent papers investigates the role of apprenticeship training along other margins. Comparing apprenticeship training as alternative to on-the-job training, Adda, Dustmann, Meghir, and Robin (2006) estimate a dynamic discrete choice model of apprenticeship choice. The length of vocational training is investigated in Oosterbeek and Webbink (2007) using a reform of compulsory schooling laws, and in Fersterer, Pischke, and Winter-Ebmer (2008), who study the variation of apprenticeship length induced by firm failures. A paper which compares vocational education and academic schooling is the work by Malamud and Pop-Eleches (2008); they do not distinguish between vocational schooling and apprenticeship training.

The chapter proceeds as follows: The next section reviews the arguments relating to the relative merits between apprenticeship training versus full-time vocational schooling, and then provides a brief background on the German educational system. Following on, we briefly describe the data. Section 4.4 addresses identification. Section 4.5 documents how individuals' educational choice responds to availability of apprenticeship vacancies. Section 4.6 contains the main results. Section 4.7 presents a number of sensitivity checks, and the last section concludes.

4.2 Background

4.2.1 Differences between apprenticeship training and full-time schooling

In this part we briefly review the main arguments in the comparison between vocational schooling and apprenticeship training.

Schools may be able to provide broader knowledge and more conceptually oriented instruction. This relates to the literature which investigates firm incentives to provide training, starting from the seminal work of Becker (1993). Firms will not invest in the workers' general human capital, since in a competitive labor market the firm will not be able to recover the revenue from this investment. Acemoglu and Pischke (1998) show firms may be willing to provide some general training, since the informational advantage of knowing the worker's quality results in a rent to the firm; but investment in general training is still inefficiently low. A number of authors highlight the role of the apprentice as a form of unskilled worker to the firm. Within the regulatory constraints and contractual commitments to the trainee, firms maximize profits through the use of the apprentice as unskilled worker at low wages (Heckman, 1993). Along similar lines, the employee-type status of the apprentice gives the firm discretion, even given a regulatory framework, and this in turn may lead to commitment problems on the side of the firm. The central drawback of apprenticeship training is that it is thought to be too firm-specific, and may not be sufficiently portable to other firms. The large number of moves young people make in their transition from school to work is well documented (Topel and Ward, 1992), and in the context of technological and structural change, transferability and the ability to acquire further skills are important criteria.

A number of advantages of apprenticeship training are of educational nature. Apprenticeship training is believed to be the more practical approach to learning, which contextualizes knowledge in the workplace. This may be especially relevant for less academically able young adults, and may increase motivation (Ryan (1998, 2001)). School-based instruction relies essentially on a simulated work environment, which may make it harder to link theory and practice. The combination of two learning places in dual systems may on the one hand lead to additional returns from the interaction of two forms of learning. On the other hand, it carries the risk of two unconnected approaches. In terms of skills, apprenticeship training may confer additional work-related skills, like team-work, discipline, the ability to integrate into a working environment and the corresponding working hours and conditions. Firms may know better what skills are required and are more likely to employ the latest technology and practices. Firm involvement in financing training may lead to efficiency gains (Plug and Groot, 1998). Furthermore, apprenticeships may serve a useful function in terms of job search and matching. Firms learn about the quality of the worker, and apprentices also benefit from a reduction of uncertainty about the employer. The on-the-job aspect of apprenticeship training is likely to provide not only more information than can be transferred through certificates, but also information about the specific value of the firm-worker match.

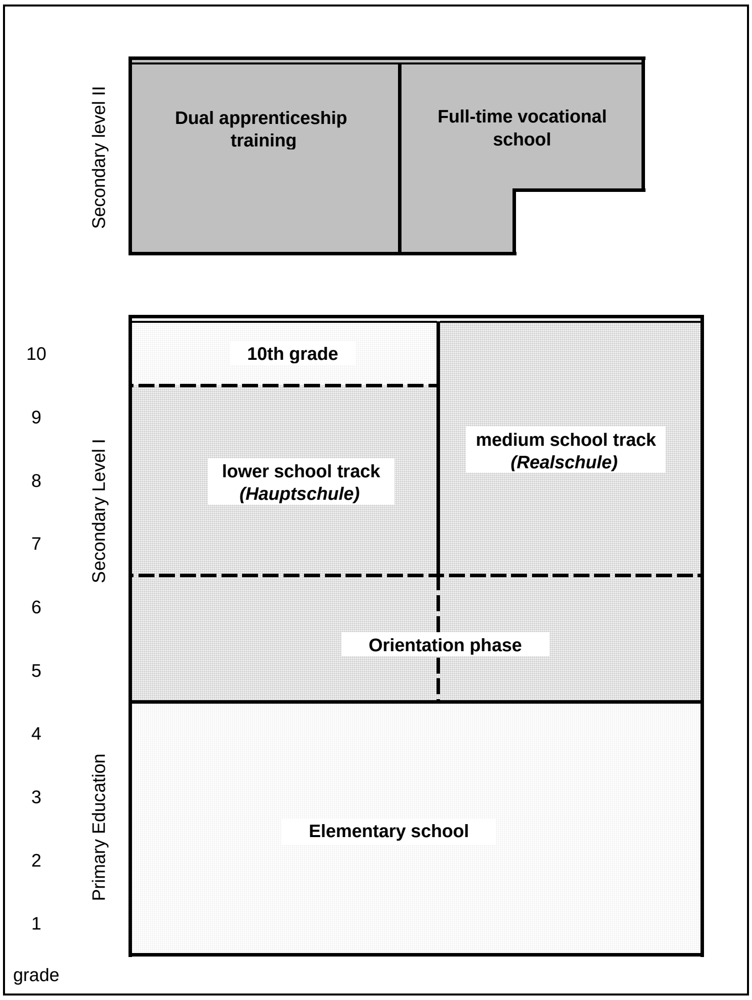
4.2.2 Institutional background on the German educational system

In this section, we provide a brief review of the relevant institutional background in which our study is conducted.[6] When aged between 10 and 12, students are typically tracked into three school streams: the *Gymnasium* as the track for later university students, and the lower and medium school (*Hauptschule* and *Realschule*) leading towards vocational education. Mobility between tracks is rare; since this chapter focuses on vocational education, we limit attention to the lower and medium schooling track. Figure 4.1 shows the structure of the educational and vocational system for these groups. Students complete general secondary school after grade nine or ten, and usually enter vocational education after that. The dual apprenticeship system is particularly well known. In this system, young adults can train and obtain a vocational degree in one of a large number of occupations. Apprenticeships have a full duration of at least two years, with most apprenticeships having a full duration of three, or three and a half years.[7] Apprentices and firms write a contract, which is registered and supervised by the Chambers of Industry and Commerce or the Chamber of Handicrafts. The contract

[6] For further reference, key features of the apprenticeship system in Germany are described in Winkelmann (1996), Soskice (1994), Witte and Kalleberg (1995) and Dustmann (2004). Summary descriptions of the German vocational system can also be found in KMK (2008).

[7] Of the 25 most popular apprenticeship degrees for males in Western Germany in 2004, all but one have a regular duration of at least three years, and only one lasts for two years (data for 2004 for males in Western Germany. Source: Federal Institute for Vocational Education and Training (2006, 2008)).

Figure 4.1: Diagram of the German education system up to vocational degree



Note: Diagram shows structure of education system up to vocational degree, for individuals streamed into lower and medium schooling track. Own representation based on KMK (2008, Figure 2.4.7). See text for details.

typically specifies an initial probationary period, after which firing from the firm's side is difficult. Apprentices spend about one third of the time in school-based instruction, which typically amounts to one or two days per week and is run by the regional government. Two thirds of the time are spend in the firm, where the apprentice works and the employer provides training. The three most frequent apprenticeship degrees are 'motor vehicle mechatronics technician', 'industrial mechanic', and 'management assistant for retail services'.

Firms can report their apprenticeship vacancies to the local employment office with a request for placement. This typically involves that the firm contacts the employment office and reports their apprenticeship vacancy. The employment office attempts to assist in the matching for this vacancy through advertising it to young adults, and possibly through suggesting candidates to the firm. The firm can then interview the candidate and can, but is under no obligation to do so, offer the apprenticeship to the young adult. The firm is not charged for using this service. Although the firm does not have to report apprenticeship vacancies, this service is regularly used by firms. As an indication, we can compare the number of vacancies reported to the employment office with official statistics on new apprenticeship contracts, on which reliable data is available since apprenticeship contracts need to be specially registered. For example, in the year 1985, which is in the middle of the period we consider here, firms reported 481,000 apprenticeship vacancies in the twelve months up to September 1985. At the end of September, 697,000 new apprenticeship contracts had been registered, and 31,000 apprenticeship vacancies were still unfilled (BA, 1991, Table 45). Hence, vacancies reported to the employment office make up a large share of all apprenticeships.

Full-time vocational schooling is an alternative form of vocational preparation.[8] These schools have a duration ranging from one to three years. One-year vocational

[8]Two main types are the full-time vocational schools (*Berufsfachschulen*) and the schools of health professions (*Schulen des Gesundheitswesens*).

courses (*Berufsgrundbildungsjahr* or *Berufsvorbereitungsjahr*) are preparatory courses and do not lead to vocational degrees on their own, but can typically be credited towards further vocational training, especially apprenticeship training. Two or three year courses lead to a recognized occupational certificate, and can lead to the same occupational degrees as apprenticeship training (KMK, 2006). There are a number of vocational degrees which can only be obtained in these full-time vocational schools. These programs lead to a degree as 'assistant' in a range of different occupations. For males, the most common ones are 'technical IT assistant', 'commercial assistant', and 'carer for the elderly'.[9]

In many of the frequently chosen occupational fields, occupation-specific qualifications exist in both the firm-based dual system as in the full-time vocational school. For example, in the field of information technology, young people could obtain an apprenticeship degree as IT specialist in the firm, or a school-based degree as mentioned above. Nonetheless, it is known that the distribution across occupational groups differs between the two training forms; therefore one important sensitivity check investigates how our main results change when we explicitly account for occupation fixed effects.

We conclude this section by briefly reviewing alternative available data sources for the fraction of individuals who obtain different forms of vocational qualification. Official statistics put the shares for highest vocational qualification in Germany at 19% unskilled, 61% dual apprenticeship, and 19% vocational schools.[10] Troltsch et al. (1999) survey the evidence on the share of unskilled youth without formal vocational qualification; estimates range between 10 and 20%. Witte and Kalleberg (1995) report that 16% of men have a school-based vocational education.[11]

---

[9]Figures for males from West Germany for 2006, see Federal Statistical Office (2007).

[10]Own calculation based on Federal Institute for Vocational Education and Training (2006), excluding all individuals with college education.

[11]Estimates based on GSOEP. Their sample excludes unskilled workers. For women, the corresponding number in their data is 23%, reflecting the higher proportion of females in schools of health professions.

4.3 Data

The analysis in this chapter is based on a large administrative panel data set of individual employment histories for German employees, the IABS, provided by the Institute for Employment Research (IAB). The sample contains 2% of all employees who have ever been subject to social security contributions over the period 1975 to 2001. We provide details on the data and the sample we use in the Data Appendix 4.A.2, and limit the discussion here to the key aspects. The data contains detailed records of both employment and unemployment spells. Crucially for this analysis, it contains not only regular employees, but also records firm-based apprentices. We limit our analysis to West German males (excluding West Berlin) from the cohorts 1964 to 1975.[12] To focus on non-college bound youth we eliminate all individuals who hold a schooling degree from the college-bound schooling track (*Abitur*), or who ever hold a degree from a university (or a university of applied sciences) in our sample. As a measure of where the young adult grows up, we record the first employment office district in which he is recorded in the data.

There are two key educational variables in the data: First, there is a variable which indicates whether the individual has obtained a vocational qualification. This is defined more broadly than firm-based apprenticeship training, and explicitly includes school-based degrees as long as they lead to a recognized vocational qualification.[13] This variable allows us to distinguish unskilled individuals (who might have completed lower or medium-level (general) schooling, but no vocational qualification) from skilled individuals, who have obtained a vocational qualification through completing either apprenticeship or a degree from a full-time vocational school.[14] Second, we compute

[12]We restrict our analysis to males because incorporating fertility decisions would complicate the analysis considerably, which are likely to be important during the age range we consider.

[13]The reporting instructions for firms explicitly clarify that the firm is to report educational qualification as 'completed vocational degree' for adults who have either completed a dual firm-based apprenticeship or have obtained a recognized degree from a full-time vocational school (BA, 2008).

[14]Fitzenberger, Osikominu, and Völter (2005) suggest to use an imputation rule to ensure consistency of this variable over time. We follow this approach by measuring educational status as the highest

for each individual the number of years spent in apprenticeship training (up to a given age). Apprenticeship degrees have a regular duration of up to 3.5 years; to reflect that we top-code the number of years of apprenticeship training at this value. This constitutes our main measure of exposure to apprenticeship training. It can be thought of as similar to a years of education measure, but, importantly, it only refers to years spent as apprentice in a firm.[15]

Our main outcomes of interest are unemployment and wages. For unemployment, we take indicator variables for whether the individual has been registered unemployed for at least a given number of days during the calendar year. In our main results, we focus on unemployment for at least 30 days. For wages, we take log average daily wages in regular full-time employment over the calendar year. We also study measures of annual mobility, for which we define indicators for changing industry and occupation, respectively.

To measure availability of apprenticeships, we make use of a unique data set, which annually records apprenticeship vacancies at a fine regional level, dividing Western Germany into 141 local labor markets.[16] This statistic contains the total number of apprenticeship vacancies that have been reported by firms to their local employment office with a request for assistance with placement of the vacancy. We normalize the vacancy data by an estimate of the number of young people who grow up in each district. We then assign each individual this measure of apprenticeship availability at age 16, in the relevant local labor market. — Means and standard deviations for our sample of the relevant variables are displayed in Table 4.1.

---

value reported up to the age of interest.

[15]We discuss possible measurement error in this variable in Section 4.7.3.

[16]In the following, we refer to these interchangeably as *employment office districts* or *regions*.

Table 4.1: Sample summary statistics

Variable	Mean	St. dev.
Apprenticeship training (years)	2.182	1.239
apprenticeship vacancies (at age 16)	0.622	0.254
unskilled vacancies (at age 16)	5.790	1.044
skilled vacancies (at age 16)	6.620	0.925
unskilled market wage (at age 16)	4.227	0.072
skilled market wage (at age 16)	4.372	0.069
unemployment rate (at age 16)	0.047	0.026
age (years)	24.507	1.114
German national (indicator)	0.879	0.326
Observations	242,014	

Note: Table reports means and standard deviations, reported for the sample of our outcome regression for probability of unemployment of at least 30 days, as reported in Table 4.5 below. Wages are log daily wages. For wages and general vacancies, the skilled group refers to those with either form of vocational degree (apprenticeship or full-time vocational schooling), and the unskilled group to those without vocational degree. See text for details.

4.4 Identification

This section has three parts. First, we describe a small open economy setting and incorporate educational choice between apprenticeship and vocational school. We investigate the effect of price shocks on educational choice and factor rewards to motivate an instrumental variable strategy. Second, we discuss how treatment effect identification is affected if one recognizes that individuals can choose between *three* different alternatives: apprenticeship training, vocational schools, or direct entry as unskilled worker. In this multinomial choice setting, we show that under specific circumstances, a univariate instrument continues to recover a well-defined treatment effect which is of interest. Third, we discuss the empirical implementation in this study.

4.4.1 A small open economy model with educational choice

We begin by describing each region as a small open economy, integrated through trade, and follow the standard assumptions of neoclassical trade models. Here, we summarize key properties of the model; a formal description is contained in the Appendix 4.A.1.

To simplify the discussion, consider an economy in which two products are produced using two labor inputs (apprenticeship graduates and vocational school graduates). Technology is identical across regions and characterized by constant returns to scale, and positive and diminishing marginal productivities. Assume that products are freely traded across regions, while factors are mobile across sectors but immobile across regions. Product prices $(1, p)$ are determined at the world market.

With regards to the school leaver's educational choice, young people choose between apprenticeship training and vocational schooling alternatives. This choice is made by comparing alternative-specific utilities. These are made up of the respective present discounted value of future wages earned in the main labor market, a random individual-level shock which affects the utility of apprenticeship training, and a region-specific parameter which affects apprenticeship utility. This latter parameter may be interpreted, following Findlay and Kierzkowski (1983), as the local availability of an education-specific fixed factor. This specification leads to a region-specific educational decision rule which determines the number of apprentices as a function of the wage difference.[17] We follow standard assumptions of neoclassical trade models by assuming that one sector (say sector 1) is always characterized by a higher intensity of apprenticeship input than sector 2, and that equilibrium with incomplete specialization is possible.

In this setting, consider an increase in the price of good 2. This leads to an increase in the factor reward for the type of labor used intensively in that sector (i.e., the vocational school graduates). Sector 2 expands and sector 1 declines, whereby both sectors increase their factor intensity in apprenticeship graduates. In terms of educational choice, young people respond to this change in wages by moving into the non-apprenticeship track. Since the function which maps wage differences into apprentices is location-specific as described above, the response in the number of ap-

[17] Since higher wages will make this option more attractive and increases the number of individuals who choose this alternative, this specification satisfies normality as in McKenzie (1955).

prenticeship graduates differs across regions. Importantly, this model is characterized by factor-price equalization: For given product prices, factor rewards are identical across regions. Thus, this model describes a setting where a price shock translates into a differential response in terms of number of apprentices, but does not lead to a differential response in terms of factor rewards. Thus, this economic framework generates an exclusion restriction, which motivates an instrumental variable approach for comparing productivity across the two types of labor. Since industry price shocks and the location-specific characteristic (interpreted as availability of an education-specific factor) are not directly observed in this data, we implement this procedure by instead using the number of apprenticeship vacancies in each region. We assume that the economy adjusts quickly to changes in prices, so that the observed responses reflect movements between equilibria. – Although not explicitly modeled here, a straightforward comparison of means would pick up selection effects arising from individual heterogeneity.

One important assumption of the model and the empirical strategy is that local variation is relevant for young people, in that they cannot adjust by moving to other regions. Regional mobility is generally thought to be low in Germany, and since apprenticeship wages are low, apprentices usually have to rely on living at home, so that factor immobility seems to be a sensible assumption for this group.

4.4.2 Identification with a univariate instrument in a multinomial choice setting

The interest of this chapter is in comparing two alternative forms of obtaining a vocational qualification. It is natural to ask whether individuals might want to adjust to apprenticeship availability by entering the labor market as unskilled worker. This transforms the decision problem into a multinomial choice problem. In this section, we discuss how treatment effect identification is affected by this more general choice problem, and under what conditions the IV estimator recovers an alternative-specific parameter of interest; the corresponding empirical analysis is found in section 4.5.2 below.

Heckman and Vytlacil (2007) consider identification of treatment effects in an unordered multinomial choice model with a binary instrument. Define an alternative-specific treatment effect as difference in outcomes between choices j and m, $\Delta_{j,m} \equiv Y_j - Y_m$. Assuming that an instrument Z_j affects only the utility of choice alternative j, and is excluded both from the vector of potential outcomes as from the other choice-specific utilities, they show that the Wald estimand that arises from changing Z_j is equal to a weighted average of $\Delta_{j,m}$ across the other alternatives $m = 1...M$. In general, the IV estimator will not recover a comparison between two specific alternatives, but rather a *weighted average* across all possible alternatives. This reflects that in response to a change of the value of option j, 'movers' respond by changing into different alternatives, and the weights correspond to the probability of choosing a particular option m as next-best alternative.[18]

In the context of this study, there is an important special case to this. If a change in the instrument Z_j induces all 'movers' to switch into the *same* second-best alternative, then the weighted average of alternative-specific causal effects collapses into a single alternative-specific causal effect. In this special case, the instrument Z_j (which modifies the value of taking option j) recovers $\Delta_{j,m}$ for a specific alternative m. Thus, there is a special case in which a univariate IV does recover one alternative-specific causal effect in a multinomial choice setting. Whether this special case applies in a particular application can be empirically verified: this reflects that we observe individuals' educational choices and can estimate how they respond to changes in the value of the instrument. We provide empirical evidence on this below in section 4.5.2. The evidence presented there clearly indicates that the IV estimator recovers the relative return between apprenticeship training and full-time vocational school alternatives.

[18] Conceptually, this is similar to the analysis of Angrist and Imbens (1995) for the variable treatment intensity case.

In our three-alternatives model, this is shown by documenting that the probability of working as unskilled does not respond to apprenticeship availability, which corresponds to a weight of 0 in the discussion above. This implies that the IV estimates presented in this chapter do recover a well-defined treatment coefficient of interest even in this generalized multinomial setting.

A related concern might be that individuals respond to shocks in apprenticeship availability by going to college instead. The early tracking of pupils in the German education system allows us to abstract from the college-going decision: As described above, individuals are tracked at ages 10–12 into either a university-bound upper track or a lower- or medium-schooling track, so that the decision to go to university is already pre-determined through the tracking decision earlier on. Correspondingly, we eliminate all individuals from the university-bound upper track from our sample, and limit attention the lower- and medium level track.

4.4.3 Empirical implementation

We implement this approach in a linear empirical framework, which we now describe. Denote $Y_{i,cj,ta}$ as a labor market outcome of interest for individual i, who grew up in cohort c in region j, measured at age a in time period t. Our model for the outcome equation is

$$Y_{i,cj,ta} = \alpha_1 S_i + \alpha_{2j} + \alpha_{3c} + \alpha_{4t} + \alpha_{5a} + \alpha_6 X_i + \alpha_7 X_{cj} + \alpha_{8j} \cdot c + \epsilon_i \tag{4.1}$$

where S_i indicates apprenticeship training, and where we think of α_1 as being heterogeneous in the population. α_{2j}, α_{3c} correspond to region and cohort fixed effects, respectively, α_{4t} and α_{5a} to year and age indicators. X_i are individual characteristics, and X_{cj} to labor market characteristics at the time and region where the individual grew up; $\alpha_{8j} \cdot c$ denote region-specific trends. We instrument for S_i with apprenticeship availability, so that the corresponding first stage is

$$S_i = \gamma_1 Z_{cj} + \gamma_{2j} + \gamma_{3c} + \gamma_{4t} + \gamma_{5a} + \gamma_6 X_i + \gamma_7 X_{cj} + \gamma_{8j} \cdot c + u_i \tag{4.2}$$

where Z_{cj} denotes the availability of apprenticeships, corresponding to the cohort c in region j. As before, γ_{2j}, γ_{3c} correspond to region and cohort fixed effects, respectively, γ_{4t} and γ_{5a} to year and age indicators.[19] In all regressions presented, we account for permanent differences with fixed effects for regions, cohorts, and region-specific trends. We also control for labor market characteristics at age 16, which we discuss below in more detail.

Since we include district and cohort fixed effects, our estimates have the interpretation of diff-in-diff estimates, where differential developments in apprenticeship availability are used to identify the effect. In addition to that, our specification allows for linear region-specific time trends. This more general specification allows each district to follow a separate time trend.[20]

We now explain how we control for conditions in the local main labor market: first, we use a similar general vacancy measure, and control for the (log) number of general vacancies in the relevant (main) labor market (at age 16 as before), by skill level.[21] Second, we include local wages and unemployment rate for males (aged 25 to 40), in the relevant labor market at age 16. These additional controls allow for a broader description of the individual's choice problem.

---

[19] These last two sets of regressors capture year and age effects specific to the time when the outcome is measured. The problem of separately identifying cohort, age and time effect is well known (Heckman and Robb, 1985). Year effects would not be separately identified if the cohort and age effects were fully interacted. Here we limit ourselves to the three additive sets of indicators, and follow the approach taken in Hall (1971) by excluding an additional dummy to avoid perfect collinearity. No attempt is made here to interpret the coefficients on these indicator variables, so that it is not of importance which of the indicator variables is eliminated (see Berndt and Griliches (1990)).

[20] We test for differential trends in our sample, by performing an F-test on the set of region-specific trends in the estimation equations. We find that the region-specific trends are significant at the 1% level both on the first stage and in the IV estimates. To perform this test, we cluster estimates by region-cohort cell.

[21] Here, skill level refers to those with either form of vocational degree (skilled), compared to school-leavers who enter directly as regular employees (unskilled).

4.5 The effect of apprenticeship vacancies on educational choice

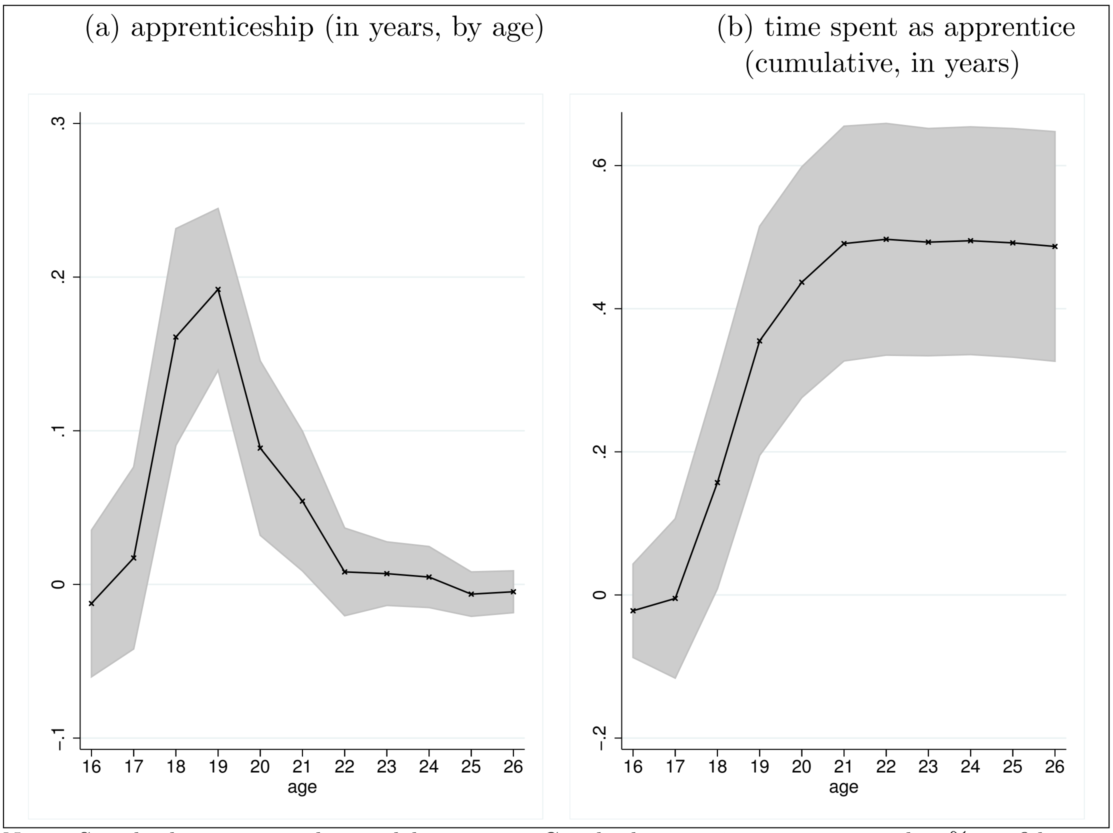
In this section, we proceed as follows: We first document the effect of apprenticeship vacancies on educational choice, where we show that an increase in apprenticeship vacancies significantly increases the apprenticeship training undertaken. We then turn to clarifying the margin of adjustment, and document that variation in apprenticeship availability leads individuals to substitute between apprenticeship training and full-time vocational schools. We validate this with a falsification exercise. Finally, we present the first-stage regression that corresponds to our main results.

4.5.1 Vacancies and apprenticeship training

To start with, we consider how the amount of apprenticeship training undertaken at each age is affected by vacancies. For this purpose, we estimate equation (4.2), but take as outcome only apprenticeship training undertaken at a specific age.[22] The resulting coefficient (γ_1) measures the effect of apprenticeship vacancies on apprenticeship training at a particular age, measured as a fraction of the year. We obtain a separate coefficient for each age, from 16 to 26, and plot the resulting coefficients, each of which is estimated in a separate regression. The allows to show at which ages individual training decisions are affected. The results are shown in Figure 2 (a). The figure shows that apprenticeship availability has a pronounced and significant effect on the time spent as apprentice at ages 18 and 19, and the effect then declines to zero after that. This documents that – as expected – apprenticeship availability affects the educational choice of young adults after they leave school. We then repeat this exercise, but now look at the total years of apprenticeship training obtained up to a particular age. The dependent variable here varies from 0 for someone who has not (or not yet) entered the apprenticeship system, to 3.5 for someone who has done a full-length apprenticeship

[22]Estimates are obtained using the same set of controls that we employ in the main results below.

Figure 4.2: Effect of apprenticeship vacancies on apprenticeship training



Note: Standard errors are clustered by region. Graph shows point estimates and 10% confidence intervals. See the data section for the definition of the variables.

training. The resulting coefficients are shown in Figure 2 (b). At age 21, the effect of apprenticeship vacancies is fully realized, and the effect is flat afterwards. To get a sense of the magnitudes, recall that the instrument has a standard deviation of about 0.25; thus, a one standard deviation increase in the instrument moves expected apprenticeship training by about 0.125 years. As can be seen from this graph, apprenticeship vacancies at school-leaving age have a lasting effect on the individual's educational choice.

4.5.2 Clarifying the margin of adjustment

We interpret the availability of apprenticeship places as varying the utility specific to the apprenticeship option. In order to interpret the main results below, it is important to clarify where these marginal individuals, who have been affected by the vacancy variable, come from in a multinomial setting; this is the purpose of this section. Recall that we limit attention to non-college bound youth; then there are three potential avenues for a young adult who leaves school: the apprenticeship system, a full-time vocational school, or a direct entry into the labor market as unskilled worker. We now document the substitution behavior that is associated with an increase in apprenticeship vacancies. One hypothesis we investigate is that the marginal apprentice enters as unskilled worker when apprenticeship availability is low, and as apprentice when availability is high. We call this the *substitution for unskilled work* hypothesis. Alternatively, young people at the margin who do an apprenticeship when availability is good might be drawn from the pool of individuals who would otherwise obtain a vocational degree in a full-time vocational school, and we term this the *substitution for full-time vocational schooling* hypothesis.

To investigate this, we select all individuals aged 24, and group them into these three categories as follows. The *unskilled*, i.e. those who have neither form of vocational qualification, make up 21.5% in our sample. Second, individuals who have a vocational qualification based on a a full-length firm-based apprenticeship training, which we define as having a vocational degree and at least 1.5 years of apprenticeship training.[23] This group (*apprentices*) account for 62.5% in our sample. The remaining individuals (*full-time vocational school*) make up 16% in our sample.[24] The proportions we obtain in our data fit well with estimates from other sources (see section 4.2.2 on page 122).

We now use this grouping to estimate a trivariate probit model, in which the main explanatory variable is apprenticeship vacancies. We include the same set of controls as outlined before.[25] The resulting marginal effects are reported in Table 4.2.

Table 4.2: Trivariate probit: Marginal effect of apprenticeship vacancies

category	variable	Marginal effects
group 1 unskilled	apprenticeship vacancies (at age 16)	-0.00164 [0.0342]
group 2 vocational school	apprenticeship vacancies (at age 16)	-0.124 [0.0386]***
group 3 apprenticeship	apprenticeship vacancies (at age 16)	0.126 [0.0396]***
	cohort fixed effects	Yes
	region fixed effects	Yes
	region trends	Yes
	labor market conditions at entry	Yes
	Observations	61358

Note: Standard errors reported in brackets, clustered by region. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level. See text for details.

The effect of apprenticeship availability on the probability of being in the unskilled group is small and not statistically different from zero. Instead, the marginal effects on full-time vocational schooling and apprenticeship training are of similar magnitude and opposite signs, they are both statistically significant at the 1% level. This indicates that the apprenticeship vacancy variation induces individuals to move between vocational

[23]This can be thought of as a minimum requirement. Most apprenticeship programs have a full length of either three or three and a half years.

[24]Most of this third group have obtained their vocational degree entirely outside the apprenticeship system, and some are recorded with very short apprenticeship spells only.

[25]Since we restrict the sample to those aged 24 in this exercise, we do not need to account separately for age or year effects.

schooling and apprenticeship training.

Survey evidence supports this result. Troltsch et al. (1999) report evidence from a representative telephone survey of unskilled youth without any vocational qualification. Of those interviewed, the majority did not search for a training position, had an offer but rejected it, or started a training program but then dropped out. The fraction of individuals who indicate that they searched unsuccessfully is low (14%), suggesting that there is little room for apprenticeship vacancies to have any effect. Similarly, a number of characteristics strongly increase the probability of being in the unskilled group (e.g. dropping out of secondary school), which make it very difficult to enter either apprenticeship (because of firm hiring decision) or full-time vocational schools (because of school admission criteria).

In summary, the above estimates provide evidence in favor of the *substitution for full-time schooling hypothesis* and against the *substitution for unskilled work hypothesis*. We replicate this result in a linear regression framework, where we regress an indicator for being unskilled on the vacancy measure (and the set of controls) and find no effect, as reported in Table 4.11 (in Appendix 4.A.3).

For further evidence, we now turn to a falsification exercise, which exploits that our two substitution hypotheses imply different predictions about the age at which individuals are first seen in this employment data. We present these as IV estimates, where the regressor of interest is years of apprenticeship training. First, we look at the age at which an individual is first seen in the data, excluding apprenticeship spells. If individuals substitute between apprenticeship and *unskilled work*, we would expect a coefficient of 1: an additional year of apprenticeship training delays the first non-apprenticeship spell accordingly. Under the *substitution for full-time vocational schooling* hypothesis, on the other hand, we would expect a coefficient of 0, if the full-time vocational schools have roughly the same length as apprenticeship training. Estimates are found in Table 4.3.[26]

[26]The corresponding first stage is discussed in detail below.

Table 4.3: Falsification exercise: Age at first regular job

	Age first seen in data (*excluding* apprenticeship spells)		**Age first seen in data (*including* apprenticeship spells)**	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
Apprenticeship training (years)	0.183	**-0.0703**	-0.787	**-0.969**
	[0.0153]***	**[0.313]**	[0.0159]***	**[0.278]*** **
German national (indicator)	-0.353	-0.0368	-0.375	-0.149
	[0.0404]***	[0.386]	[0.0387]***	[0.343]
cohort fixed effects	Yes	Yes	Yes	Yes
region fixed effects	Yes	Yes	Yes	Yes
region trends	Yes	Yes	Yes	Yes
age fixed effects	Yes	Yes	Yes	Yes
year fixed effects	Yes	Yes	Yes	Yes
Labor market controls at entry	Yes	Yes	Yes	Yes
Observations	241585	241585	242014	242014
First stage F-statistic		24.96		25.55
First stage p-value		0.00000172		0.00000133
Mean (dependent variable)	19.62	19.62	17.50	17.50
St. dev. (dependent variable)	1.796	1.796	1.928	1.928
Hypothesis test: Substitution for unskilled work				
Corresponding parameter value		1		0
F-statistic		11.68		12.16
p-value		0.000827		0.000651
Hypothesis test: Substitution for full-time vocational schooling				
Corresponding parameter value		0		-1
F-statistic		0.0504		0.0121
p-value		0.823		0.913

Note: Standard errors reported in brackets, clustered by region. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level. See text for details. Hypothesis test at the bottom of the table refers to the coefficient of the variable 'years of apprenticeship training'.

Column (1) reports the OLS results, which suggest that in a simple comparison of means, an additional year of apprenticeship delays the first non-apprenticeship spell. In column (2), we instrument using the vacancy data. The coefficient is now very close to zero and insignificant. This confirms that individuals at the margin switch between apprenticeship training and vocational schools.

In columns (3) and (4), we repeat this exercise for the age at which the individual is first ever seen in the data, *including* in apprenticeship training. Under the *unskilled* hypothesis, individuals are seen in the data at the same age as apprentices, so that the corresponding coefficient is 0. Under the *full-time vocational schooling* hypothesis, individuals are seen in the data a year earlier, because the firm registers the apprentice similar to its regular workers. The coefficient corresponding to this hypothesis is then -1. The resulting IV estimate is not statistically different from -1, as indicated by our test at the bottom of the table, while we reject a coefficient of 0 at the 1% level.

We conclude that the estimates which we present in the following should be interpreted as being the treatment effect of the individual who switches from vocational full-time schooling to firm-based apprenticeship, depending on the local availability of apprenticeships. Doing so we follow the work of Imbens and Angrist (1994) on the Local Average Treatment (LATE) parameter, an interpretation that requires a monotonicity assumption on how individuals react to changes in the instrument.

The estimates above do not only clarify the interpretation of our estimates provided below; they also convey an important substantive point relating to economic policy. They indicate that when full-time vocational schooling exists as alternative, measures which increase supply of apprenticeship vacancies are likely to draw individuals from these full-time vocational schools rather than individuals who would have entered the labor market directly as unskilled workers. In that sense, the results suggest that policies which expand availability of apprenticeships are effective in increasing the take-up of firm-based apprenticeship training, but they are not necessarily effective in reducing the number of unskilled workers, when vocational schooling exists as alternative.

4.5.3 First stage results

Table 4.4 presents the first stage results which correspond to our main outcome equation.[27] The dependent variable of interest is years of apprenticeship training obtained. As sample we select individuals aged 23 through 26. Each column in Table 4.4 corre-

[27]Between different outcomes, the available sample differs slightly. We present the first stage here for one of our main outcomes, unemployment for at least 30 calendar days, as reported in Table 4.5 on page 140 below. In the IV results below, we report the corresponding F-statistic along with the estimates.

Table 4.4: First stage

	Apprenticeship training (years)		
	OLS	OLS	OLS
	(1)	(2)	(3)
apprenticeship vacancies (at age 16)	0.505	0.490	0.492
	[0.0970]***	[0.0981]***	[0.0974]***
unskilled market wage (at age 16)		-0.174	-0.166
		[0.240]	[0.239]
skilled market wage (at age 16)		0.248	0.235
		[0.493]	[0.492]
unemployment rate (at age 16)		0.438	0.435
		[0.452]	[0.456]
unskilled vacancies (at age 16)			0.0301
			[0.0185]
skilled vacancies (at age 16)			0.000405
			[0.0305]
German national (indicator)	1.243	1.243	1.243
	[0.0283]***	[0.0283]***	[0.0283]***
cohort fixed effects	Yes	Yes	Yes
region fixed effects	Yes	Yes	Yes
region trends	Yes	Yes	Yes
age fixed effects	Yes	Yes	Yes
year fixed effects	Yes	Yes	Yes
Observations	242014	242014	242014
First stage F-statistic	27.09	24.94	25.55
First stage p-value	0.0000	0.0000	0.0000
Mean (dependent variable)	2.182	2.182	2.182
St. dev. (dependent variable)	1.239	1.239	1.239
Minimum (dependent variable)	0	0	0
Maximum (dependent variable)	3.500	3.500	3.500

Note: Standard errors reported in brackets, clustered by region. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level. See text for details.

sponds to a different set of control variables. Column (1) only includes an indicator of German nationality. The variable of interest measures apprenticeship vacancies. The corresponding coefficient for the apprenticeship degree indicator is 0.505, and it is significant at the one percent level. Keeping in mind that the standard deviation of this variable is about 0.25, a one standard deviation change in apprenticeship vacancies increases average apprenticeship training by about 0.125 years.

In columns (2) and (3), we add further controls for local labor market conditions at age 16, allowing us to investigate whether the first stage coefficient of interest is sensitive to a slightly extended specification of the educational choice stage. These variables vary at the same level as our instrument. Column (2) adds average local wages and unemployment rates, computed for males aged 25 to 40.[28] The coefficient on apprenticeship vacancies goes down somewhat, but the change is small and does not affect the significance of the coefficient at all. In column (3) we control for the number of all open vacancies (in logs) at age 16 in the main labor market in the relevant region, separately by required skill level.[29] Columns (1) to (3) demonstrate that shocks in apprenticeship availability translate into differences in educational attainment as measured by apprenticeship training, and further that this effect is statistically strong and robust to an extended specification of the educational decision problem.

An important concern in IV estimation is the problem of weak instruments (Bound, Jaeger, and Baker, 1995; Staiger and Stock, 1997; Stock, Wright, and Yogo, 2002). One way to assess this is to consider the F-statistic from the first stage. As indicated in the bottom of table, the F-statistics from the instrument here are above 25, respectively, and well above the rule of thumb of an F-statistic of 10. This suggests that weak instruments should not be a concern in this application.

4.6 Effect of training form on labor market outcomes

4.6.1 Unemployment and wages

As outcomes, we focus on unemployment and wages. We define an indicator for unemployment of at least 30 days during the calendar year. Wages refer to log average daily wage in full-time employment during the calendar year. We select all young adults aged 23 to 26, and pool annual observations for efficiency. All standard errors are clustered at the region level, allowing for arbitrary within-cluster dependence in the error term, including serial correlation. Estimates are reported in Table 4.5.

[28] Average wages are included separately for skilled versus unskilled workers, referring to whether the individual has completed either form of vocational training, or no vocational training.

[29] Here, skill level refers to whether a position requires some form of completed vocational qualification, or not.

Table 4.5: Main outcomes

	Indicator: Unemployed at least 30 days		**Log average daily wages**	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
Apprenticeship training (years)	-0.0355	-0.105	0.0309	0.0289
	[0.00133]***	[0.0457]**	[0.00133]***	[0.0501]
unskilled market wage (at age 16)	0.0183	0.00851	-0.0738	-0.0743
	[0.0584]	[0.0634]	[0.0521]	[0.0520]
skilled market wage (at age 16)	-0.0474	-0.0260	-0.0822	-0.0817
	[0.144]	[0.152]	[0.112]	[0.112]
unemployment rate (at age 16)	-0.384	-0.317	-0.111	-0.109
	[0.117]***	[0.115]***	[0.102]	[0.108]
unskilled vacancies (at age 16)	0.00679	0.00860	0.00373	0.00377
	[0.00446]	[0.00498]*	[0.00441]	[0.00454]
skilled vacancies (at age 16)	-0.00124	-0.000175	-0.00185	-0.00181
	[0.00638]	[0.00647]	[0.00777]	[0.00782]
German national (indicator)	-0.0127	0.0737	0.00570	0.00822
	[0.00470]***	[0.0557]	[0.00460]	[0.0635]
cohort fixed effects	Yes	Yes	Yes	Yes
region fixed effects	Yes	Yes	Yes	Yes
region trends	Yes	Yes	Yes	Yes
age fixed effects	Yes	Yes	Yes	Yes
year fixed effects	Yes	Yes	Yes	Yes
Observations	242014	242014	218438	218438
First stage F-statistic		25.55		16.43
First stage p-value		0.00000133		0.0000835
Mean (dependent variable)	0.149	0.149	4.203	4.203
St. dev. (dependent variable)	0.356	0.356	0.320	0.320

Note: Standard errors reported in brackets, clustered by region. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level. See text for details.

Column (1) presents the OLS estimate for the unemployment outcome. According to this estimate, an additional year of apprenticeship training reduces the probability of unemployment by 3.55 percentage points. Column (2) instruments for apprenticeship. The effect goes up substantially in magnitude, and the resulting effect is 10.5 percentage points. Although precision decreases, the effect remains significant at the 5% level. This evidence suggests that in this age range, former apprenticeship graduates have a lower probability of being unemployed.

In a constant coefficient framework, the difference between OLS and IV estimates is informative about the direction of selection bias, which here would suggest negative selection.[30] In a heterogeneous treatment effect framework, the IV estimate reflects the treatment effect of the marginally affected subgroup (who switch between apprenticeship training and full-time vocational schools), indicating that this group is characterized by a higher treatment effect.

We now turn to estimating the effect on productivity as measured by wages. The main IV estimate is reported in column (4) of Table 4.5. Interestingly, the OLS and IV coefficients are very similar at about 3 per cent. Once we instrument, the effect on wages is not statistically different from zero. This suggests that the two alternative forms of training lead to similar levels of productivity, but it is important to keep in mind that the standard error of this estimate is relatively large.

To the young trainee, one benefit of apprenticeship training may be the access to the firm's internal labor market, as argued for example by Soskice (1994). Our estimates may, at least in partly, reflect that a fraction of the apprenticeship graduates may stay on in their training firm, while the vocational school graduates are more likely to go through search unemployment at the end of their training, and then over time catch up with apprenticeship graduates. Our data allows us to investigate this directly by splitting up the sample by age, and estimate separately for each age. These results are reported in Table 4.6. Columns (1) through (4) show the OLS effects by age, which basically remain constant at 0.035. Columns (5) to (8) show the IV results. Here, we find a pronounced pattern over time: The effect declines rapidly with age. At age 26, it is no longer significant. The IV result suggests that the beneficial effect of lower unemployment probability is not permanent, but transitory.[31] This is consistent with the interpretation that the apprenticeship training smooths the initial transition into the main labor market, but that the vocational school graduates then catch up

[30] A priori, the direction of selection bias is unclear. On the one hand, firms select positively from the applicants. On the other hand, comparing vocational full-time schools and apprenticeship training, it is possible that more academically inclined individuals have a preference for schools, and that school are rigorous in enforcing admission standards. Bonnal, Mendes, and Sofer (2002) find evidence of negative selection into apprenticeship for France. Plug and Groot (1998) find no evidence of self-selection for a sample of Dutch young adults.

[31] Note that the mean of the dependent variable remains very similar across the age range.

Table 4.6: Effect on unemployment by age

	Unemployed for at least 30 days at age. . .				**Unemployed for at least 30 days at age. . .**			
	23	**24**	**25**	**26**	**23**	**24**	**25**	**26**
	OLS	OLS	OLS	OLS	IV	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
apprenticeship training (years)	-0.0354	-0.0344	-0.0363	-0.0359	**-0.148**	**-0.134**	**-0.0987**	**-0.0573**
	[0.00167]***	[0.00153]***	[0.00167]***	[0.00163]***	**[0.0716]****	**[0.0683]***	**[0.0519]***	**[0.0566]**
cohort fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
region trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
age fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Labor market controls at entry	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Personal characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	59315	61358	60785	60556	59315	61358	60785	60556
First stage F-statistic					15.49	20.43	26.67	35.46
First stage p-value					0.000	0.000	0.000	0.000
Mean (dependent variable)	0.158	0.151	0.144	0.142	0.158	0.151	0.144	0.142
St. dev. (dependent variable)	0.365	0.358	0.351	0.349	0.365	0.358	0.351	0.349

Note: Standard errors reported in brackets, clustered by region. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level. See text for details.

over time.

4.6.2 Effect on mobility

We also investigate year-on-year mobility as measured by changing occupation or industry. Mobility measures inform us about transferability of skills, especially given that one may be concerned that apprenticeship training is very specific and leads to lock-in effects. A number of studies document an unexpectedly high fraction of former apprentices who work in occupations different from the one they trained in (see e.g. Werwatz (2002)). This has been interpreted as indicating that the training provides more general skills, or acts as a signal for general worker quality (Heckman, 1993). Our set-up here allows us to compare mobility rates between different forms of vocational training, and we look at occupational and industry mobility. The rate of year-on-year mobility in both dimensions is high at around 15 percent. Estimates are reported in Table 4.7. The OLS result suggests that apprenticeship training has a negative effect on mobility. Once we instrument, the coefficient turns positive, but our results are imprecisely estimated; we find no significant evidence of differential mobility behavior between the two groups.

4.6.3 Responsiveness to negative shocks

One of the key questions in the debate on how apprenticeship training compares to other forms of training relates to the individual's ability to adjust to negative shocks. For example, Heckman, Roselius, and Smith (1993) suggest that narrow technical training may reduce options later in life by introducing rigidities; an overly tight link to a specific task or firm may result in constraints (Witte and Kalleberg, 1995).[32] One particular concern is that the benefit from apprenticeship training may be very specific

[32] Although long-run career constraints are clearly of interest, the individuals in our sample are still too young to be informative about that. At the same time, it is likely that the role of the employment experience in the main labor market increases relative to initial vocational training (Witte and Kalleberg, 1995), so that our age group should be of particular interest.

Table 4.7: Mobility

	change in occupation		**change in industry**	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
apprenticeship training (years)	-0.0178	0.0573	-0.0118	0.0152
	[0.000953]***	[0.0637]	[0.000980]***	[0.0573]
unskilled market wage (at age 16)	-0.0304	-0.00302	0.0553	0.0658
	[0.0442]	[0.0534]	[0.0470]	[0.0522]
skilled market wage (at age 16)	-0.0743	-0.0888	-0.0468	-0.0538
	[0.108]	[0.115]	[0.104]	[0.105]
unemployment rate (at age 16)	-0.140	-0.227	-0.0495	-0.0774
	[0.101]	[0.138]	[0.0941]	[0.113]
unskilled vacancies (at age 16)	0.00332	0.000984	0.00789	0.00703
	[0.00339]	[0.00406]	[0.00329]**	[0.00409]*
skilled vacancies (at age 16)	-0.00382	-0.00505	-0.00847	-0.00895
	[0.00583]	[0.00609]	[0.00606]	[0.00592]
German national (indicator)	-0.0262	-0.119	-0.0125	-0.0456
	[0.00387]***	[0.0786]	[0.00408]***	[0.0702]
cohort fixed effects	Yes	Yes	Yes	Yes
region fixed effects	Yes	Yes	Yes	Yes
region trends	Yes	Yes	Yes	Yes
age fixed effects	Yes	Yes	Yes	Yes
year fixed effects	Yes	Yes	Yes	Yes
Observations	183245	183245	183085	183085
First stage F-statistic		9.104		7.960
First stage p-value		0.00303		0.00548
Mean (dependent variable)	0.144	0.144	0.150	0.150
St. dev. (dependent variable)	0.351	0.351	0.357	0.357

Note: Standard errors reported in brackets, clustered by region. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level. See text for details.

to the firm, and not sufficiently transferable if the firm-worker match is hit by a negative shock. In this section, we assess this empirically by studying the responsiveness of the young person to a job destruction shock. To do this, we follow the literature on firm closures as a negative shock. This approach has been widely used as a source of exogenous job destruction (Jacobson, LaLonde, and Sullivan, 1993; Oreopoulos, Page, and Stevens, 2008; Sullivan and von Wachter, 2007). We implement this as follows: For each worker, our data contains information on firm closures by recording the last year in which employees were recorded under this firm identifier. This allows us to identify firm closures. We define workers to be *at risk* if their employing firm ceases to exist in the same or the following year. We then consider the probability of being unemployed in period t, and investigate the effect of having been at risk in period $t-1$. Since we are concerned with the differential effect between forms of training, our coefficient of particular interest is the interaction between apprenticeship training and the indicator for being at risk. We control for the at-risk indicator.[33] Our specification now has two endogenous variables (apprenticeship training, and the interaction between training and being at risk in the previous period), and we instrument for these variables with apprenticeship availability, and availability interacted with the at-risk indicator. Results are found in Table 4.8, where we present results for being unemployed for at least 30 days, at least 45 days, and for log wages.

Columns (1) and (2) report the first stages. As indicated in the F-statistics, both first stages are statistically significant with a low p-value. Column (3) reports the IV estimate for the indicator of being unemployed for at least 30 days. The first row shows the effect of apprenticeship training on unemployment. This essentially replicates the baseline result we reported above, that apprenticeship training reduces the probability of being unemployed. The interaction in the second row shows how the effect differs when the individual is hit by a negative shock. Although this coefficient is not individually significant, it indicates that the adverse effect of firm closures is stronger for (former) apprentices. We repeat this exercise for the indicator of being unemployed for at least 45 days, and find essentially the same pattern, except that here the interaction term is individually significant at the 10% level. Column (5) considers log wages; the coefficient on the interaction is negative but insignificant.

As reported at the bottom of the table, the coefficients are jointly significant for the unemployment outcome, indicating that the form of training matters for the pattern of unemployment in this context. Interestingly, the two coefficients are of similar magnitude, so that the coefficient on the at-risk interaction offsets the beneficial effect

[33]Thus, we follow the firm closures literature in maintaining the assumption that the firm closure is a random event. In this context, a natural way of examining this is by regressing the at-risk indicator on apprenticeship vacancies at age 16. We find that apprenticeship vacancies do not have a significant effect on the probability of being at-risk, consistent with the interpretation of firm closures as a random shock.

Table 4.8: Responsiveness to negative shock (firm closure)

			Unemployed at least 30 days	Unemployed at least 45 days	Log average daily wages
	First (i)	First (ii)	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)
apprenticeship training (years)			-0.132	-0.103	0.0326
			[0.0611]**	[0.0533]*	[0.0568]
apprenticeship * closure indicator			0.165	0.169	-0.114
			[0.105]	[0.0958]*	[0.0887]
apprenticeship vacancies (at age 16)	0.379	0.0161			
	[0.125]***	[0.0133]			
apprent. vacancies * closure indicator	-0.131	0.218			
	[0.0759]*	[0.0898]**			
closure indicator	-0.0774	1.895	-0.205	-0.228	0.119
	[0.0533]	[0.0595]***	[0.211]	[0.194]	[0.180]
cohort fixed effects	Yes	Yes	Yes	Yes	Yes
region fixed effects	Yes	Yes	Yes	Yes	Yes
region trends	Yes	Yes	Yes	Yes	Yes
age fixed effects	Yes	Yes	Yes	Yes	Yes
year fixed effects	Yes	Yes	Yes	Yes	Yes
Labor market controls at entry	Yes	Yes	Yes	Yes	Yes
Personal characteristics	Yes	Yes	Yes	Yes	Yes
Observations	176492	176492	176492	176492	170140
First stage F-statistic	6.049	4.988			
First stage p-value	0.00302	0.00808			
Mean (dependent variable)	2.196	0.0652	0.0977	0.0849	4.235
St. dev. (dependent variable)	1.213	0.428	0.297	0.279	0.307
Hypothesis test: Joint significance					
F-statistic			4.259	3.676	1.149
p-value			0.0160	0.0278	0.320
Hypothesis test: add to zero					
F-statistic			0.0650	0.339	0.518
p-value			0.799	0.561	0.473

Note: Standard errors reported in brackets, clustered by region. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level. See text for details.

of apprenticeship training. We formalize this by testing whether the two coefficients sum to 0. As shown at the bottom of the table, this hypothesis cannot be rejected at any reasonable level of significance. This implies that apprenticeship reduces the probability of being unemployed, but this benefit is lost if the worker's firm closes down. After the worker-firm match is hit by destruction, the job finding rate is no different between the two forms of vocational training.

4.7 Robustness Checks

In this section, we present a number of checks to investigate the sensitivity of the results presented in Table 4.5.

4.7.1 Grouped data

The importance of accounting for dependence in the data is well understood at least since Moulton (1990), and all standard errors in this chapter are adjusted by clustering on the region level. An alternative way of recognizing that the identifying variation is on region-cohort level is to take averages in region-cohort cells, and to repeat the analysis on this aggregated data using averages for all variables which vary within region-cohort cells. Here, we present the corresponding estimates. Table 4.9 shows the estimates for the unemployment and the wage outcome. As expected, the estimates are very similar to the main estimates reported above.

4.7.2 Controlling for occupation-specific fixed effects

If the distribution across occupations differs between the two tracks, one might be worried that our results may partly pick up systematic differences between occupations. We therefore investigate how the results change if one explicitly accounts for occupational fixed effects. The results are found in in Panel B of Table 4.10.[34] They

[34]Panel A replicates the base case results for easy reference.

Table 4.9: Sensitivity: Grouped data

	Indicator: unemployed at least 30 days			Average log daily wages		
	(1)	(2)	(3)	(4)	(5)	(6)
	First stage	OLS	IV	First stage	OLS	IV
apprenticeship training (years)		-0.0318	-0.0741		0.0359	0.0217
		[0.00711]***	[0.0437]*		[0.00654]***	[0.0534]
apprenticeship vacancies (at age 16)	0.494			0.447		
	[0.107]***			[0.121]***		
unskilled market wage (at age 16)	-0.196	0.0146	0.0130	-0.271	-0.0819	-0.0821
	[0.263]	[0.0642]	[0.0674]	[0.268]	[0.0564]	[0.0569]
skilled market wage (at age 16)	0.203	-0.0548	-0.0331	0.190	-0.0855	-0.0807
	[0.543]	[0.157]	[0.162]	[0.536]	[0.124]	[0.122]
unemployment rate (at age 16)	0.412	-0.386	-0.335	0.437	-0.121	-0.107
	[0.501]	[0.127]***	[0.125]***	[0.531]	[0.113]	[0.120]
unskilled vacancies (at age 16)	0.0302	0.00647	0.00760	0.0257	0.00392	0.00421
	[0.0204]	[0.00494]	[0.00523]	[0.0213]	[0.00482]	[0.00499]
skilled vacancies (at age 16)	0.00103	-0.00128	-0.000797	0.00386	-0.00174	-0.00163
	[0.0336]	[0.00709]	[0.00712]	[0.0345]	[0.00852]	[0.00855]
German national (indicator)	1.242	-0.0483	0.00717	1.253	0.0414	0.0601
	[0.120]***	[0.0285]*	[0.0602]	[0.120]***	[0.0236]*	[0.0707]
cohort fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
region trends	Yes	Yes	Yes	Yes	Yes	Yes
calender year (linear)	Yes	Yes	Yes	Yes	Yes	Yes
age (linear)	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1689	1689	1689	1689	1689	1689
First stage F-statistic			21.48			13.67
First stage p-value			0.000			0.000
Mean (dependent variable)	2.182	0.149	0.149	2.188	4.203	4.203
St. dev. (dependent variable)	0.269	0.0698	0.0698	0.282	0.0649	0.0649

Note: Standard errors are clustered on the region level. Data is collapsed into region-cohort cell means. In this specification, age and time fixed effects are replaced by cell-specific mean of age and time. Observations are weighted by cell size to account for the varying precision in estimating the corresponding means.

indicate that controlling for occupational fixed effects reduces the magnitude of the estimates for OLS and IV estimates, although the level of significance of the estimates is unchanged. One interesting difference is that for the log daily wages outcome, the IV coefficient is now very close to zero. This reinforces our earlier conclusion that there are no significant productivity differences between the two groups.

4.7.3 Measurement error in years of apprenticeship training

In this section we investigate one possible source of measurement error in the apprenticeship years data, which concerns the exact date of the transition from apprenticeship training into full-time employment for some of the years in our data: Firms are only required to report the exact end date of an apprenticeship training from 1992 on (Schw-

Table 4.10: Sensitivity analysis

	Indicator: Unemployed at least 30 days		**Average log daily wages**	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
Panel A: Base case results				
Apprenticeship training (years)	-0.0355	-0.105	0.0309	0.0289
	[0.00133]***	[0.0457]**	[0.00133]***	[0.0501]
Observations	242014	242014	218438	218438
First stage F-statistic		25.55		16.43
First stage p-value		0.000		0.000
Panel B: Controlling for occupational fixed effects (two-digit classification)				
Apprenticeship training (years)	-0.0287	-0.0947	0.0273	0.00515
	[0.00132]***	[0.0471]**	[0.00105]***	[0.0444]
Observations	239662	239662	217378	217378
First stage F-statistic		24.94		16.28
First stage p-value		0.000		0.000
Panel C: Measurement error corrected				
Apprenticeship training (years, corrected)	-0.0373	-0.111	0.0331	0.0307
	[0.00136]***	[0.0490]**	[0.00132]***	[0.0533]
Observations	242014	242014	218438	218438
First stage F-statistic		23.30		14.64
First stage p-value		0.000		0.000
Panel D: Excluding inner-German border regions				
Apprenticeship training (years)	-0.0346	-0.118	0.0312	0.0360
	[0.00137]***	[0.0456]**	[0.00139]***	[0.0544]
Observations	223315	223315	201753	201753
First stage F-statistic		28.45		18.19
First stage p-value		0.000		0.000
Panel E: Using log number of apprenticeship places as instrument				
Apprenticeship training (years)	-0.0355	-0.093	0.0309	0.006
	[0.00133]***	[0.047]*	[0.00133]***	[0.056]
Observations	242014	242014	218438	218438
First stage F-statistic		18.19		12.49
First stage p-value		0.000		0.001

Each panel in this table corresponds to a separate sensitivity analysis. Panel A replicates the base case results for easy reference. Panel B additionally controls for occupational fixed effects on a two-digit level. Panel C uses as dependent variable the measurement-error corrected version of apprenticeship training. Panel D replicates the results omitting all areas at the inner-German border. Panel E omits normalization by cohort size. Note: Standard errors reported in brackets, clustered by region. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level. See text for details.

erdt and Bender, 2003). Before that, apprentices who stay on in their training firm after apprenticeship training, and become regular skilled employees in the same firm without any kind of interruption, may be reported as full-time employees for the entire calendar year in which the transition occurs, although in fact they are apprentices in the first part of the year, and regular employees only after the exam. In these cases, we will undercount the number of days spent in apprenticeship training, because we do not observe the days in apprenticeship training in the last year of training. This may introduce some measurement error into the apprenticeship duration data. This is a very limited form of mis-measurement: It does not affect whether we ever see a young adult as apprentice, but it only affects the last calendar year of apprenticeship training. Thus, the undercount consists of either one or six to seven months, depending on the examination date, relative to a full apprenticeship duration of typically three (or three and a half) years.

As a sensitivity check, we use an imputed version of the apprenticeship duration variable. All affected cases share the property that they are seen in the data as apprentice until December 31, and are then reported as skilled employees from January 1 in the same company. Since the apprenticeship terminates with the final exam, which is typically towards the end of January (end of first semester) or during June or July (end of second semester), this transition at the end of the calendar year is a strong signal for this kind of misclassification. We flag these cases, and then increase the duration of apprenticeship training by four months, which is the mean number of months expected in the last year of training between the two exam dates. We then re-run our main specification on this imputed variable, which corrects for the undercount. Results are found in Panel C of Table 4.10.

The estimated coefficients are almost identical for both the wage and the unemployment outcome. The F-statistic is marginally lower, which reduces the precision of the main estimates somewhat, but both the size of the main coefficients and the level of significance is unchanged when we use this imputed version.

4.7.4 Further specification checks

The inner-German border opened up in 1989, within the observation window of this study. It is known that the inner-German border regions experienced differential development from other parts of the country (Buettner and Rincke, 2007). In particular, workers from the eastern part of Germany commuted to West German border areas for better employment prospects in the face of strong economic differentials between east and west. To investigate whether this may have had any effect on our results, we exclude all districts in our sample at the inner-German border. This reduces the number of districts from 141 to 127. The results are shown in Panel D of Table 4.10. The coefficients are again virtually unchanged, which strongly suggests that our results are not affected by any differential behavior in the border regions.

While a normalization seems to be an appropriate way to measure the relevant availability of apprenticeships, a natural question of interest would be to isolate the role of the vacancy availability from the cohort size. To answer that, we re-estimate our main results, but take as instrument the log number of reported vacancies. We report these estimates in Panel E of Table 4.10. Without the normalization, the F-statistic is lower, but the IV coefficient on apprenticeship training in the unemployment equation is of a similar magnitude and the estimate has a p-value of 0.051. The estimate in the wage equation is again very close to 0. This indicates that while the normalization is useful in measuring relative availability, the main source of variation is the number of apprenticeship places.

4.7.5 Functional form

The empirical specification outlined in equations (4.1) and (4.2) imposes a specific functional form on the model. In particular, it imposes that our key variable for apprenticeship, Z_{cj}, enters linearly in the conditional mean of the first stage and the outcome equation. The objective of this section is to semiparametrically investigate whether this functional form is appropriate. For that purpose, we estimate the partially linear model of Robinson (1988), which allows the effect of apprenticeship availability to be completely unrestricted:

$$S_i = f\left(Z_{cj}\right) + \alpha_{2j} + \alpha_{3c} + \alpha_{4t} + \alpha_{5a} + \alpha_6 X_i + \alpha_7 X_{cj} + \alpha_{8j} \cdot c + \epsilon_i \tag{4.3}$$

and similarly for the reduced form equation. Here, $f(\cdot)$ can be an arbitrary smooth function, which we estimate non-parametrically following the two-step approach from Robinson (1988).

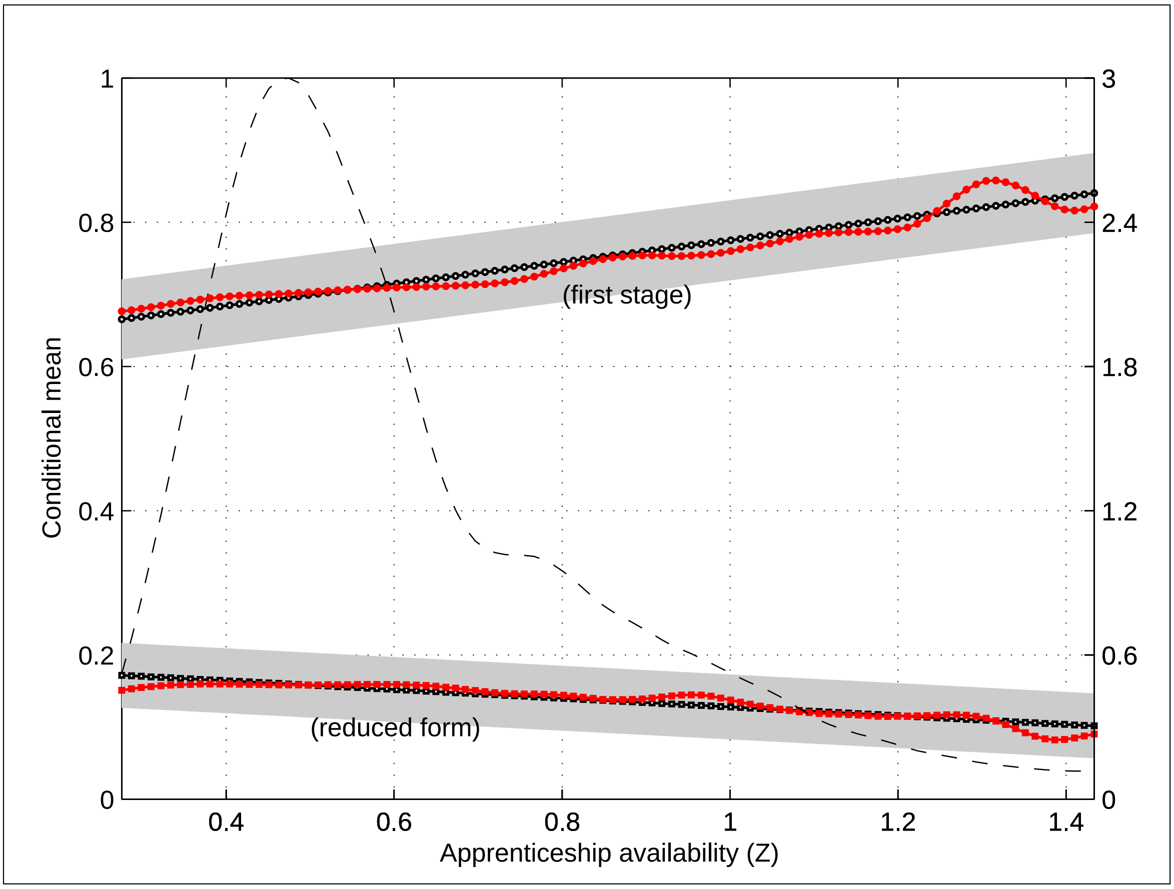
For this exercise, we take the sample of our regression for the outcome unemployment of at least 30 days in the calendar year, for the sample of 24-year olds. We employ a biweight Kernel function and set the bandwidth to $h_z = 0.10$ in this exercise. In a preliminary experiment, we cross-validated the bandwidth and obtained values of at least $h_z = 0.14$, so we substantially undersmooth relative to that, which should help to bring out any non-linearities.[35] The result of this procedure – the nonparametric estimate of $f(\cdot)$ – can be seen in Figure 4.3, together with an estimate of the density of Z for reference. We superimpose the prediction from the linear model for comparison, along with the corresponding 10% confidence interval.[36]

As can be seen from the figure, the nonparametric estimate and the linear prediction are very close together, and even in areas of low density the differences are modest. We conclude that the assumption of linearity appears to be appropriate in this application.

[35] Cross-validation suggests a bandwidth of $h_z = 0.14$ and $h_z = 0.18$ for the first stage and the reduced form, respectively. To reduce computational burden, cross-validation is done in univariate regressions. In the cross-validation, we trim the highest and lowest one percentile in the Z dimension to reduce the effect of outlier values in the cross-validation objective function. In the leave-one-out prediction, we account for the dependence in the data by excluding all points with the same value of Z.

[36] The partially linear model does not separately identify the intercept. We shift the estimates to match the predicted means.

Figure 4.3: Semiparametric analysis: The partially linear model of Robinson (1988)



Note: Graph shows estimates over the first to the 99th percentile range in the Z dimension (apprenticeship availability). Density estimate is rescaled. See text for details.

4.8 Conclusion

The objective of this chapter is to study the relative returns to apprenticeship training relative to vocational schooling. This identifies a parameter which is of substantial interest to policy-makers, who are faced with the decision of how to design vocational education. The diversity of vocational schooling schemes around the world may be seen as evidence that there is no consensus on how these schemes compare. An empirical investigation needs to take account of potentially strong selection effects. We exploit that non-college bound young adults are subject to fluctuation in the availability of apprenticeships. We document that this affects their schooling choice, and leads them to substitute between apprenticeship training and vocational schooling. We use this exposure to estimate the differential return in a rigorous empirical framework, which accounts for permanent region and cohort effects, and allows for region-specific time trends.

Our findings suggest that the skills young people obtain are in fact similar between vocational schooling and apprenticeship training, as measured by wages at ages 23 to 26. That suggests that both schemes have similar productivity effects on participants; the benefits and drawbacks of either form of instruction seem to be balancing out in terms of effects on productivity. That is an important finding in the context of the debate on the relative merits of these alternatives.

At the same time, we find substantial and significant differences in the probability of unemployment. Importantly though, this effect shows a strong age profile. We trace this effect across the age of the young adult and find that it is highest at young ages and then declines rapidly, and becomes insignificant at age 26. Thus, the benefit in terms of lower unemployment rates is a transitory one. This suggests that apprenticeship training provides a benefit to participants in that it improves labor market attachment early in their career. We provide further consistent evidence for this based on firm-closures: When a young adult is hit by a negative shock through a firm closure, the benefit of apprenticeship training is lost, and the job-finding rate between the two groups is no different after that.

In summary, our results indicate that the two forms of vocational preparation deliver similar skills, but the apprenticeship training aides the initial integration of young adults into the labor market. How much weight the policy-maker places on this difference will depend on the emphasis on a smooth school-to-work transition, but the evidence on problems associated with high youth unemployment suggests that these considerations are likely to be important. Traditionally, the comparison between vocational schooling and apprenticeship training focuses primarily on the educational dimension; the results we obtain here underline the relevance of vocational training in the worker-firm matching process.

4.A Appendix

4.A.1 Description of the economy

Consider an economy consisting of regions (denoted by k), in which two products are produced (X_{1k} and X_{2k}), using two labor inputs (apprenticeship graduates, L_k^{appr}, and vocational school graduates, L_k^{school}). Technology is identical across regions (described by production functions $F_1(\cdot)$ and $F_2(\cdot)$) and characterized by constant returns to scale, and positive and diminishing marginal productivities. Assume that products are freely traded across regions, while factors are mobile across sectors but immobile across regions. Product prices $(1, p)$ are determined at the world market, where the price of good 1 has been normalized to 1.

Now consider the educational choice of the young schoolleaver, who decides between apprenticeship and vocational school. Specify utility from the two alternatives for individual i as the present value of future wages:

$$V_i^{appr} = \sum_{t=2}^{T} \left(\frac{1}{1+r} \right)^{t-1} w_k^{appr} + x_k + u_i \tag{4.4}$$

$$V_i^{school} = \sum_{t=2}^{T} \left(\frac{1}{1+r} \right)^{t-1} w_k^{school} \tag{4.5}$$

where r is the discount rate, u_i is a person-specific random utility shock; T denotes the length of the working life, so that the income stream in the main labor market is from $t = 2$ (after the initial training period) to $t = T$ (assuming that income during the training period is negligible). Annual factor rewards for apprenticeship graduates and vocational school graduates, respectively, are denoted by w_k^{appr} and w_k^{school}. Furthermore, x_k is a region-specific parameter which affects the apprenticeship option, which we interpret as local availability of an education-specific factor. Individuals choose the alternative with the higher utility, so that the resulting number of individuals who enter apprenticeship training is

$$
\begin{aligned}
L_k^{appr} &= \sum_i \mathbb{1}\left\{\sum_{t=2}^{T}\left(\frac{1}{1+r}\right)^{t-1}\left(w_k^{appr}-w_k^{school}\right)+x_k+u_i \;>\; 0\right\} && (4.6)\\
&\equiv g_k\left(w_k^{appr}-w_k^{school}\right), && (4.7)
\end{aligned}
$$

where $g_k(\cdot)$ is an educational decision function satisfying $g_k' \geq 0$, which implies a normal factor supply.[37] Given a fixed number of schoolleavers in region k, $\overline{L}_k$, we have $L_k^{school} = \overline{L}_k - L_k^{appr}$.

The economy can be described by the following set of equations:

$$
\begin{aligned}
X_{1k} &= F_1\left(L_{1k}^{appr}, L_{1k}^{school}\right) && (4.8)\\
X_{2k} &= F_2\left(L_{2k}^{appr}, L_{2k}^{school}\right) && (4.9)\\
L_k^{appr} &= L_{1k}^{appr} + L_{2k}^{appr} && (4.10)\\
L_k^{school} &= L_{1k}^{school} + L_{2k}^{school} && (4.11)\\
L_k^{appr} &= g_k\left(w_k^{appr}-w_k^{school}\right) && (4.12)\\
L_k^{school} &= \overline{L_k} - L_k^{appr} && (4.13)
\end{aligned}
$$

Define factor intensities $\rho_{1k} = \frac{L_{1k}^{appr}}{L_{1k}^{school}}$, $\rho_{2k} = \frac{L_{2k}^{appr}}{L_{2k}^{school}}$, and assume $\rho_{1k} > \rho_{2k}$. Define intensive production functions $f_1(\rho_{1k}) = F_1(\rho_{1k}, 1)$, $f_2(\rho_{2k}) = F_2(\rho_{2k}, 1)$. Assume no complete specialization: equilibrium is possible in which all regions produce both goods.

The marginal value of each type of labor must be equal across sectors:

$$
\begin{aligned}
f_1'(\rho_{1k}) &= p\, f_2'(\rho_{2k}) && (4.14)\\
f_1(\rho_{1k}) - \rho_{1k} f_1'(\rho_{1k}) &= p\left(f_2(\rho_{2k}) - \rho_{2k} f_2'(\rho_{2k})\right) && (4.15)
\end{aligned}
$$

[37]Findlay and Kierzkowski (1983) describe an educational choice setting between skilled and unskilled work in which education is provided by a specific factor which is competitively rewarded; their setting leads to an educational decision similar to (4.7).

Now (4.14) and (4.15) determine ρ_{1k} and ρ_{2k}. Since the system (4.14–4.15) is the same across regions, we have that all regions choose the same optimal factor intensities:

$$\rho_{1k} = \rho_1 \tag{4.16}$$

$$\rho_{2k} = \rho_2 \tag{4.17}$$

This in turn pins down factor rewards:

$$w_k^{appr} = w^{appr} = f_1'(\rho_1) \tag{4.18}$$

$$w_k^{school} = w^{school} = f_1(\rho_1) - \rho_1 f_1'(\rho_1). \tag{4.19}$$

Thus, factor price equalization holds: In equilibrium, although the educational decision rule (4.12) is region-specific, there are no differences in factor rewards across regions.

The setting described here is an extension of the model described in Kemp (1964). Using activity analysis, McKenzie (1955, Theorem 2") proves that factor price equalization extends to the variable factor supply case if factor supplies are normal, a requirement satisfied by equations (4.12) and (4.13).

4.A.2 Data appendix

The analysis in this chapter is based on a large administrative data set of individual employment histories for German employees, the IABS, provided by the Institute for Employment Research (IAB). A description of this data can be found in Bender, Haas, and Klose (2000). The sample contains 2% of all employees who are ever subject to social security contributions over the period 1975 to 2001. It excludes data on the self-employed, civil servants, and the military. We limit our analysis to West German males (excluding West Berlin) from the cohorts 1964 to 1975. For each individual, we record the labor market status on a reference day, June 30, of each year. We then define our sample of interest as all those individuals who are in the labor force on June 30. To focus on non-college bound youth we eliminate all individuals who hold a schooling degree from the college-bound schooling track (*Abitur*), or who ever hold a degree from a university (or a university of applied sciences) in our sample. We exclude individuals who enter the labor market later than age 24. As a measure of where the young adult grows up, we record the first employment office district in which he is recorded in the data.

Apprentices in the dual system are are a clearly identified group in the data and can be distinguished from regular workers. Since we observe the full employment history of each sampled individual, it is straightforward to establish whether an individual has ever been an apprentice, and if so, for how long. For this purpose, we compute for each calendar year whether a given individual has had the apprentice status during this year, and if so, for what fraction of the year. We also compute a cumulative version of this variable, which indicates the years of apprenticeship training up to a given age.

We compute the number of days an individual is unemployed during each calendar year, and then define an indicator which takes the value 1 if the individual has been unemployed for at least 30 calendar days. We compute log average daily wages as follows: We divide total wages (in prices of 2000) earned in full-time regular employment on a calendar year basis by the number of days spent in full-time regular employment, and then take logs. We eliminate observations which would imply a wage rate of below 1 Euro per hour (in prices of 2000), assuming an eight hour day. Wage reports in the data are generally top-coded at the social-security contribution limits, but since we focus on young workers and exclude all individuals with a college degree and from the upper schooling track, this is unlikely to be relevant for our sample. — We measure change of occupation as an indicator for a change in occupation on two consecutive reference dates, limiting attention to moves into regular full-time employment. We define industry movers similarly, and both industries and occupations are coded on a two-digit level. A key advantage of this administrative data set relative to survey-based data is that we can expect to have little measurement error; this is especially important for measures of mobility (Kambourov and Manovskii, 2008).

Apprenticeship vacancies are published annually by the German Federal Employment Office. From 1991 onwards, this data is published on the website of the Federal Employment Office. For previous years, we collect the information from annual publications of the Federal Employment Office (see, e.g. BA (1991)). We normalize the vacancy data by an estimate of the number of young people who grow up in each district. For that purpose, we compute cohort sizes at the level of the *Land* based on the number of seventeen year olds, as published by the national Statistical Office. To obtain the number of young people on the finer district level, we split this based on the district level share of 15-19 year olds in each *Land*, for which we use the 1988 shares as reported in BMBF (1992, pp. 206-208).

4.A.3 Further tables

Table 4.11: Apprenticeship vacancies and educational choice

	apprenticeship training (years)	**vocational degree (indicator)**
	OLS	OLS
	(1)	(2)
apprenticeship vacancies (at age 16)	0.492	0.0197
	[0.0974]***	[0.0251]
cohort fixed effects	Yes	Yes
region fixed effects	Yes	Yes
region trends	Yes	Yes
age fixed effects	Yes	Yes
year fixed effects	Yes	Yes
labor market controls	Yes	Yes
Observations	242014	242014
Mean (dependent variable)	2.182	0.796
St. dev. (dependent variable)	1.239	0.403

Note: Standard errors reported in brackets, clustered by region. * indicates significance at 10%, ** indicates significance at 5%, *** indicates significance at 1% level. See text for details.

Chapter 5

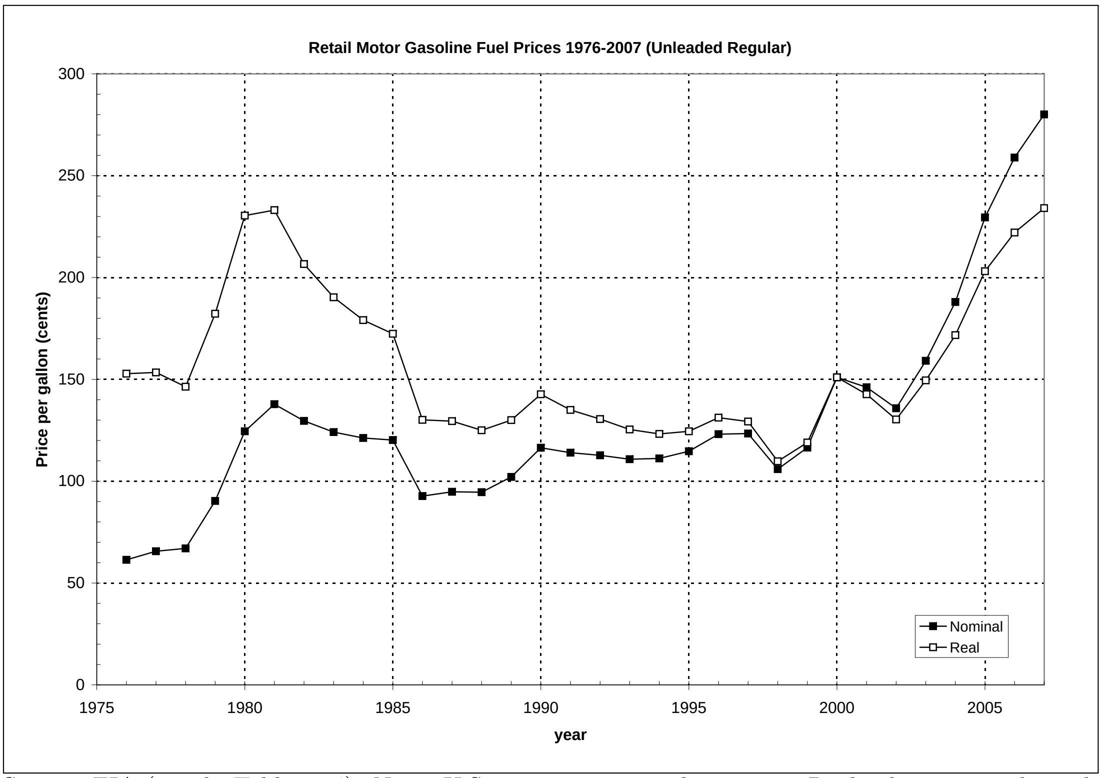
Measuring the Price Responsiveness of Gasoline Demand

5.1 Introduction

This chapter describes a new method for estimating a demand function for gasoline and the welfare costs of changes in gasoline prices. The method is also applicable to other goods. In the U.S., as in many other countries, the price of gasoline rose rapidly from 1998 until mid 2008. Figure 5.1 shows the how the average price of gasoline in the U.S. has varied over the last three decades. Prices began rising steeply in about 1998 following a period of price stability that began in about 1986. Between March 2007 and March 2008, the average gasoline price increased by 25.7 percent in nominal terms.[1] In real terms, gasoline prices reached levels similar to those seen during the second oil crisis of 1979-1981. Although prices have decreased since mid 2008, due at least in part to the global economic downturn, many observers expect prices to rise again in the future as economic activity increases. The measurement of the welfare

[1] Own calculation based on EIA (2008b, Table 9.4).

Figure 5.1: Retail Motor Gasoline Price 1976-2007 (Unleaded Regular)



Source: EIA (2008b, Table 5.24). Note: U.S. city average gasoline prices. Real values are in chained (2000) dollars based on GDP implicit price deflators. See source for details.

consequences of price changes begins with estimating the demand function for the good in question. This is often done by using a linear model in which the dependent variable is the log of demand and the explanatory variables are the logs of price and income. This model is easy to interpret because it gives constant income and price elasticities, but it is rejected by our data. Table 5.1 presents the results of estimating a constant elasticity model of gasoline demand for a class of households in the U.S. The data are from the National Household Travel Survey (NHTS). We describe the NHTS further in Section 5.3. RESET specification tests reject the constant-elasticity model. Further analysis that is described in Section 5.4 reveals that adding an interaction term to the constant elasticity model does not correct the specification error. This motivates us to use nonparametric estimation methods. Hausman and Newey (1995) also used nonparametric methods to estimate gasoline demand. Deviations from the

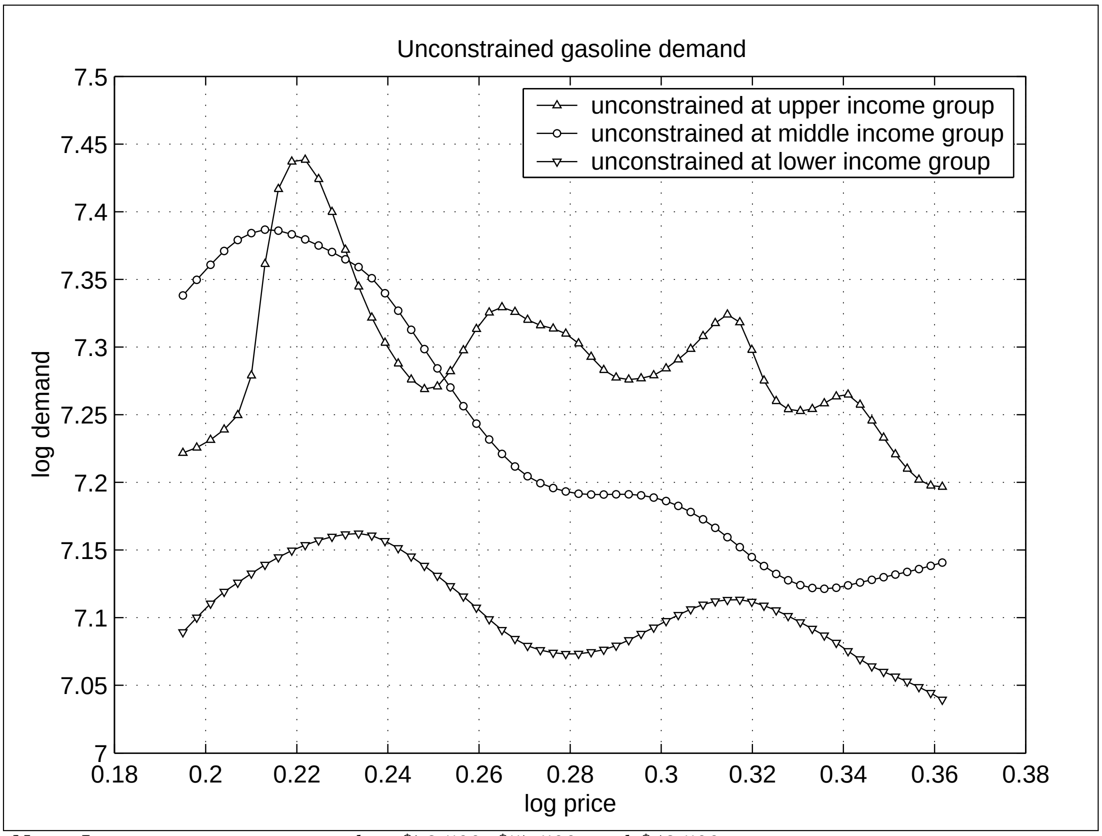
Table 5.1: OLS regression and specification test

dependent variable: log gasoline demand	
log price	-0.885
	[0.157]**
log income	0.292
	[0.015]**
constant	4.226
	[0.166]**
N	5,257
RESET3: F-stat	5.522
RESET3: p-value	0.004**
RESET4: F-stat	4.034
RESET4: p-value	0.007**

Note: Dependent variable is log of annual household gasoline demand in gallons. * indicates significance at 5%, ** indicates significance at 1% level. The bottom panel reports results from the F-test of two Ramsey RESET specification tests. RESET3 refers to including second and third polyonomials of the predicted values of the dependent variable, and RESET4 refers to including second to fourth polynomials. See text for details.

constant-elasticity model are not simply a technical concern. It is likely to matter greatly how peoples' responses to prices vary according to the price level and over the income distribution. Therefore, a flexible modeling approach such as nonparametric regression seems attractive. However, nonparametric regression can yield implausible and erratic estimates. Figure 5.2 shows nonparametric estimates of gasoline demand as a function of price at three points across the income distribution. The estimates are obtained from the NHTS data. Details of the estimation method are presented in Section 5.2 of this chapter. The figure gives some overall indication of downward sloping demand curves with slopes that differ across the income distribution but there are parts of the estimated demand curves that are upward sloping and, therefore, implausible. We interpret the implausible shapes of the curves in Figure 5.2 as indicating that fully nonparametric methods are too imprecise to provide useful estimates of gasoline demand functions with our data. One way of dealing with this problem is to impose a parametric form such as log-log linearity on the demand function. But any parametric form is essentially arbitrary and, as will be discussed further in Sec-

Figure 5.2: Unconstrained nonparametric demand estimates



Note: Income groups correspond to \$72,500, \$57,500, and \$42,500.

tion 5.4, may be misspecified in ways that produce seriously erroneous results. As a compromise between the desire for flexibility and the need for structure, one may use a semiparametric model, such as a partially-linear or single-index model. These impose parametric restrictions on some aspects of the function of interest but leave other parts unrestricted. In this chapter, we take a different approach and impose structure through shape restrictions based on economic theory. Specifically, we impose the Slutsky restriction of consumer theory on an otherwise fully nonparametric estimate of the demand function. We show that this approach yields well-behaved estimates of the demand function and price responsiveness across the income distribution while avoiding the use of arbitrary and possibly misspecified parametric or semiparametric models. We implement our approach by making use of a kernel-type estimator in which observations are weighted in a way that ensures satisfaction of the Slutsky restrictions. This maintains the flexibility of nonparametric regression while using restrictions of economic theory to avoid implausible estimation results. The constrained nonparametric estimates are consistent with observed behavior and provide intuitively plausible, well-behaved descriptions of price responsiveness across the income distribution. One important use of demand function estimates is to compute deadweight loss (DWL) measures of tax policy interventions. For some interventions, we show that reliance on the unrestricted nonparametric estimate results in DWL estimates that have incorrect signs and are, therefore, nonsensical. Our constrained estimator deals with this problem in a way that is consistent with economic theory. We find that there is substantial variation in price sensitivity across both price and income. In particular, we find that price responses are non-monotonic in income. Our estimates indicate that households at the median of the income distribution respond more strongly to an increase in prices than do households at the lower or upper income group. We do not speculate on why this is the case, but we show that it implies that our DWL measure is typically higher at the median of the income distribution that in the lower or upper income group. Section 5.2 explains our approach to nonparametric estimation of demand functions and DWL subject to the Slutsky shape restrictions. Section 5.3 describes the NHTS data. Section 5.4 presents the estimates of the demand function and shows how price responsiveness varies across the income distribution. Section 5.4 also presents the DWLs associated with several price changes and shows how they vary across the income distribution. Section 5.5 concludes.

5.2 Shape Restrictions and the Estimation of Demand and Deadweight Loss

We begin this section by describing our approach to estimating the demand function subject to the Slutsky shape restriction. Then we describe how we estimate the DWL of a tax-induced price increase. The Slutsky condition is an inequality constraint on the demand function. Our method for estimating the demand function nonparametrically subject to this constraint is adapted from Hall and Huang (2001), who present a nonparametric kernel estimator of a conditional mean function subject to a monotonicity constraint. We replace their monotonicity constraint with the Slutsky condition. To describe our estimator, let Q, P, and Y, respectively, denote the quantity of gasoline demanded by an individual, the price paid, and the individual's income. We assume that these variables are related by

$$Q = g\left(P, Y\right) + U \tag{5.1}$$

where g is a function that satisfies smoothness conditions and the Slutsky restriction but is otherwise unknown, and U is an unobserved random variable satisfying $E\left(U|P = p, Y = y\right) = 0$ for all p and y. Our aim is to estimate $g\left(p, y\right)$ nonparametrically subject to the Slutsky constraint

$$\frac{\partial g(p,y)}{\partial p} + g(p,y)\frac{\partial g(p,y)}{\partial y} \leq 0. \tag{5.2}$$

The data are observations $\{Q_i, P_i, Y_i : i = 1, ..., n\}$ for n randomly sampled individuals. A fully nonparametric estimate of g that does not impose the Slutsky restriction can be obtained by using the Nadaraya-Watson kernel estimator (Nadaraya, 1964; Watson, 1964). The properties of this estimator are summarized in Härdle (1990). We call it the unconstrained nonparametric estimator, denoted by $\hat{g}_U$, because it is not constrained by (5.2). The estimator is

$$\hat{g}_U(p,y) = \frac{1}{nh_p h_y \hat{f}(p,y)} \sum_{i=1}^{n} Q_i K\left(\frac{p - P_i}{h_p}\right) K\left(\frac{y - Y_i}{h_y}\right) , \tag{5.3}$$

where

$$\hat{f}(p,y) = \frac{1}{nh_p h_y} \sum_{i=1}^{n} K\left(\frac{p - P_i}{h_p}\right) K\left(\frac{y - Y_i}{h_y}\right) , \tag{5.4}$$

K is a bounded, differentiable probability density function that is supported on [-1,1] and is symmetrical about 0, and h_p and h_y are bandwidth parameters.

Owing to the effects of random sampling errors, $\hat{g}_U$ does not necessarily satisfy (5.2) even if g does satisfy this condition. Following Hall and Huang (2001), we solve this problem by replacing $\hat{g}_U$ with the weighted estimator

$$\hat{g}_C(p,y) = \frac{1}{h_p h_y \hat{f}(p,y)} \sum_{i=1}^{n} w_i Q_i K\left(\frac{p-P_i}{h_p}\right) K\left(\frac{y-Y_i}{h_y}\right) , \tag{5.5}$$

where $\{w_i : i = 1, ..., n\}$ are non-negative weights satisfying $\sum_{i=1}^{n} w_i = 1$ and the subscript C indicates that the estimator is constrained by the Slutsky condition. The weights are obtained by solving the optimization problem

$$\underset{w_1,...,w_n}{\text{minimize}} : \; D(w_1, ..., w_n) \tag{5.6}$$

subject to

$$\frac{\partial \hat{g}_C(p_j,y_j)}{\partial p} + \hat{g}_C(p_j,y_j)\frac{\partial \hat{g}_C(p_j,y_j)}{\partial y} \leq 0, \quad j = 1, ..., J, \tag{5.7}$$

$$\sum_{i=1}^{n} w_i = 1 , \tag{5.8}$$

and

$$w_i \geq 0; \; i = 1, ..., n, \tag{5.9}$$

where $\{p_j, y_j : j = 1, ..., J\}$ is a grid of points in the (p, y) plane. The objective function is the following measure of the 'distance' of the weights from the values $w_i = 1/n$ corresponding to the Nadaraya-Watson estimator:

$$D(w_1, ..., w_n) = n - \sum_{i=1}^{n} (n w_i)^{1/2} \tag{5.10}$$

When $w_i = 1/n$ for all $i = 1, ..., n$, $\hat{g}_C(p_j, y_j) = \hat{g}_U(p_j, y_j)$ for all $j = 1, ..., J$. Thus, the weights minimize the distance of the constrained estimator from the unconstrained one. The constraint is not binding at points (p_j, y_j) that satisfy (5.2). In the empirical application described in Section 5.4, we solve (5.6) by using the nonlinear programming algorithm E04UCF from the NAG Library. The bandwidths are selected using a method that is described in Section 5.4. In some applications, it may be desirable to impose the restriction that the good in question is normal. This can be done by adding the constraints $\partial \hat{g}_C(p_j, y_j)/\partial y \geq 0$ to (5.6), but we do not take this step here.

We now describe our method for estimating the DWL of a tax. Let $E(p)$ denote the expenditure function at price p and some reference utility level. The DWL of a tax that changes the price from p^0 to p^1 is

$$L(p^0, p^1) = E(p^1) - E(p^0) - (p^1 - p^0) g \left[p^1, E(p^1) \right]. \tag{5.11}$$

We estimate this by

$$\hat{L}(p^0, p^1) = \hat{E}(p^1) - \hat{E}(p^0) - (p^1 - p^0) \hat{g} \left[p^1, E(p^1) \right], \tag{5.12}$$

where $\hat{E}$ is an estimator of the expenditure function and $\hat{g}$ may be either $\hat{g}_U$ or $\hat{g}_C$. We obtain $\hat{E}$ by solving the differential equation

$$\frac{d\hat{E}(t)}{dt} = \hat{g} \left[p(t), \hat{E}(t) \right] \frac{dp(t)}{dt}, \tag{5.13}$$

where $\left[p(t), \hat{E}(t) \right]$ $(0 \leq t \leq 1)$ is a price-(estimated) expenditure path. We solve this equation along a grid of points by using Euler's method (Ascher and Petzold, 1998). We have found this method to be quite accurate in numerical experiments.

Inference with the constrained estimator $\hat{g}_C$ is difficult because the estimator's asymptotic distribution is very complicated in regions where (5.2) is a binding constraint (strict equality). However, if we assume that (5.2) is a strict inequality in the population, then violation of the Slutsky condition by $\hat{g}_U$ is a finite-sample phenomenon, and we can use $\hat{g}_U$ to carry out asymptotically valid inference. We use the bootstrap to obtain asymptotic joint confidence intervals for $g(p, y)$ on a grid of (p, y) points and to obtain confidence intervals for L. The bootstrap procedure is as follows.

1. Generate a bootstrap sample $\{Q_i^\star, P_i^\star, Y_i^\star : i = 1, ..., n\}$ by sampling the data randomly with replacement.

2. Use this sample to estimate $g(p, y)$ on a grid of (p, y) points without imposing the Slutsky constraint. Also, estimate L. Denote the bootstrap estimates by $\hat{g}_U^\star$ and $L^\star$.

3. Form percentile confidence intervals for L by repeating steps 1-2 many times. Also, use the bootstrap samples to form joint percentile-t confidence intervals for g on the grid of points $\{p_j, y_j : j = 1, ..., J\}$.

The joint confidence intervals at a level of at least $1 - \alpha$ are

$$\hat{g}_U(p_j, y_j) - z_\alpha(p_j, y_j)\hat{\sigma}(p_j, y_j) \leq g(p_j, y_j) \leq \hat{g}_U(p_j, y_j) + z_\alpha(p_j, y_j)\hat{\sigma}(p_j, y_j), \quad (5.14)$$

where

$$\hat{\sigma}^2(p, y) = \frac{B_K}{[nh_p h_y \hat{f}(p, y)]^2} \sum_{i=1}^{n} \hat{U}_i^2 K\left(\frac{p - P_i}{h_p}\right) K\left(\frac{y - Y_i}{h_y}\right), \quad (5.15)$$

$$\text{with } B_K = \int K(v)^2 dv \;\; and \;\; \hat{U}_i = Q_i - \hat{g}_U(P_i, Y_i) \; , \quad (5.16)$$

is a consistent estimate of $\mathrm{Var}[\hat{g}_U(p, y)]$. The coefficient $z_\alpha(p_j, y_j)$ is chosen following the approach in Härdle and Marron (1991) for computing joint confidence intervals. For this purpose, we partition the grid into intervals of $2h_p$. Within each of these M neighborhoods, $z_\alpha(p_j, y_j)$ is the solution to

$$P^\star\left[\frac{|\hat{g}_U^\star(p_j, y_j) - \hat{g}_U(p_j, y_j)|}{\hat{\sigma}^\star(p_j, y_j)} \leq z_\alpha(p_j, y_j)\right] = 1 - \beta \; , \quad (5.17)$$

where $P^{\star}$ is the probability measure induced by bootstrap sampling, and $\hat{\sigma}^{\star}(p, y)$ is the version of $\hat{\sigma}(p, y)$ that is obtained by replacing $\hat{U}_i$, P_i, and Y_i in (5.15) by their bootstrap analogs, and β is a parameter. We then choose β such that the simultaneous size in each neighborhood equals $1 - \frac{\alpha}{M}$. As Härdle and Marron (1991) show using the Bonferroni inequality, the resulting intervals over the full grid form simultaneous confidence intervals at a level of at least $1 - \alpha$. Hall (1992) shows that the bootstrap consistently estimates the asymptotic distribution of the Studentized form of $\hat{g}_U$. It is necessary to undersmooth $\hat{g}_U$ and $\hat{g}_U^{\star}$ (that is, use smaller than asymptotically optimal bandwidths) in (5.14) and step 2 of the bootstrap procedure to obtain a confidence interval that is centered at g. We discuss bandwidth selection in Section 5.4.

5.3 Data

Our analysis is based on the 2001 National Household Travel Survey. The NHTS was sponsored by the Bureau of Transportation Statistics and the Federal Highway Administration. The data were collected through a telephone survey of the civilian, non-institutionalized population of the U.S. The survey was conducted between March 2001 and May 2002 (ORNL, 2004, Ch. 3). The telephone interviews were complemented with written travel diaries and odometer readings.

The variables used in our study are annual gasoline consumption, the gasoline price, and household income. Gasoline consumption is derived from odometer readings and estimates of the fuel efficiencies of vehicles. Details of the computations are described in (ORNL, 2004, Appendices J and K). The gasoline price for a given household is the average price in dollars per gallon, including taxes, in the county where the household is located. This price variable is a county average, rather than the price actually paid by a household. It precludes an intra-county analysis (see Schmalensee and Stoker (1999)) but does capture variation in prices consumers face in different regions.

Household income in dollars is available in 18 groups. In our analysis, we assign each household an income equal to the midpoint of its group. The highest group, consisting of incomes above \$100,000, is assigned an income of \$120,000.[2] To investigate how price responsiveness of gasoline demand varies across the income distribution, we focus on three income levels of interest: a middle income group at \$57,500, which corresponds to median income in our sample, a low income group (\$42,500), which corresponds to the first quartile and a high income group (\$72,500)[3]. To obtain gasoline demand at the household level, we aggregate vehicle gasoline expenditure in dollars and gasoline consumption in gallons over multi-car households. We divide the household gasoline expenditure by the quantity of gasoline consumed to obtain the household's gasoline price. We do not investigate the errors-in-variables issues raised by the use of county-average prices or the interval censoring issues raised by the grouping of household incomes in the data. These potentially important issues are left for future research.

We exclude from our analysis households where the number of drivers is zero or whose income, gasoline cost, or annual gasoline consumption is not reported. We also exclude households that are located in Hawaii. In addition, we restrict our sample to households with a white respondent, two or more adults, and at least one child under 16 years of age. We take vehicle ownership as given and do not investigate how changes in prices affect vehicle purchases or how vehicle ownership varies across the income distribution (Poterba (1991), West (2004), Bento, Goulder, Henry, Jacobsen, and von Haefen (2005), Bento, Goulder, Jacobsen, and von Haefen (2009)). The results of Bento, Goulder, Henry, Jacobsen, and von Haefen (2005) indicate that over 95 percent of the reduction in gasoline demand in response to price changes is due to changes in miles traveled rather than fleet composition. We limit attention to vehicles that use

[2] Assuming log-normality of income, we have estimated the corresponding mean and variance by using a simple tobit model, right-censored at \$100,000. Excluding households with very high incomes above \$150,000, the median income in the upper group corresponds to about \$120,000.

[3] The income point \$72,500 occupies the 59.6-63.3th percentile. This point was chosen to avoid the problems created by the interval nature of the income variable which becomes especially important in the upper quartile of the income distribution: income brackets are relatively narrow (with widths of \$5,000) up to \$80,000, but substantially wider for higher incomes. However, estimates using higher quantiles yielded similar results and did not change our conclusions on price responsiveness across the income distribution.

gasoline as fuel, rather than diesel, natural gas, or electricity. The resulting sample consists of 5,257 observations. Table 5.2 shows summary statistics.

Table 5.2: Sample descriptives

log gasoline demand	7.168
	[0.679]
log price	0.287
	[0.057]
log income	10.954
	[0.613]
N	5,257

Note: Table shows means and standard deviations.

5.4 Estimates of Demand Responses

5.4.1 The constant elasticity model

We begin by using ordinary least squares to estimate the following log-log linear demand model:

$$\log Q = \beta_0 + \beta_1 \log P + \beta_2 \log Y + U; \quad E\left(U|P=p, Y=y\right) = 0. \tag{5.18}$$

This constant elasticity model is one of the most frequently estimated (e.g., Dahl (1979), Hughes, Knittel, and Sperling (2008)). It has been criticized on many grounds (e.g., Deaton and Muellbauer (1980)) but its simplicity and frequent use make it a useful parametric reference model. Later in this section, we compare the estimates obtained from model (11) with those obtained from the nonparametric analysis.

The estimates of the coefficients of (5.18) are shown in Table 5.1. They imply a price-elasticity of demand of -0.88 and an income elasticity of 0.29. These estimates are similar to those reported by others. Hausman and Newey (1995) report estimates of -0.81 and 0.37, respectively, for price and income elasticities based on U.S. data collected between 1979 and 1988. Schmalensee and Stoker (1999) report price elasticities between -0.72 and -1.13 and income elasticities between 0.12 and 0.33, depending on the survey year and control variables, in their specifications without regional fixed effects. Yatchew and No (2001) estimate an income elasticity of 0.27 using Canadian data for 1994-1996 and a model that does not include the price of gasoline. West (2004) reports a mean price elasticity of -0.89 using 1997 data.

Although the estimates we obtain from model (5.18) are similar to those reported by others, there is evidence that (5.18) is misspecified. We tested (11) for misspecification with Ramsey's (1969) RESET test. This test consists of adding powers of the predicted values of $\log Q$ to the model, re-estimating the resulting augmented model, and testing the hypothesis that the coefficients of the additional regressors are zero. Rejection of this hypothesis indicates that the original model is misspecified. We carried out this test twice, once with the squares and cubes of the predicted $\log Q$ values added to the model (RESET3 in Table 5.1) and once with the squares, cubes, and fourth powers of the $\log Q$'s added (RESET4). As can be seen from Table 5.1, both versions of RESET reject model (5.18) at the 0.05 level. Thus, we conclude that model (5.18) is misspecified.

West (2004) found evidence for dependence of the price elasticity on income. Accordingly, we added the interaction term $(\log P)(\log Y)$ to model (5.18). The resulting augmented model is also rejected at the 0.05 level by the RESET tests. Conceivably adding further powers and interactions of $\log P$ and $\log Y$ would yield a model that is not rejected by RESET. However, this kind of informal specification search leads to inconsistent estimators whose properties are unknown. Nonparametric estimators, by contrast, are consistent.

5.4.2 Unconstrained nonparametric estimates

Our unconstrained nonparametric estimates of the demand function, $\hat{g}_U$, are displayed in Figure 5.2. They were obtained by using the Nadaraya-Watson kernel estimator with a biweight kernel. In principle, the bandwidths h_p and h_y can be chosen by applying least-squares cross-validation (Härdle, 1990) to the entire data set, but this yields bandwidths that are strongly influenced by low-density regions. To avoid this problem, we used the following method to choose h_p and h_y. We are interested in $g(p, y)$ for y values corresponding to our three income groups and price levels between the 5th and 95th percentiles of the observed prices. We defined three price-income rectangles consisting of prices between the 5th and 95th percentiles and incomes within 0.5 of each income level of interest (measured in logs). We then applied least-squares cross-validation to each price-income rectangle separately to obtain bandwidth estimates appropriate to each rectangle. This procedure yielded $(h_p, h_y) = (0.0431, 0.2143)$ for the lower income group, (0.0431, 0.2061) for the middle income group, and (0.0210, 0.2878) for the upper income group. The estimation results are not sensitive to modest variations in the dimensions of the price-income rectangles. As was discussed in Section 5.2, $\hat{g}_U$ and $\hat{g}_U^{\star}$ must be undersmoothed to obtain properly centered confidence intervals. To this end we multiplied each of the foregoing bandwidths by 0.8 when computing confidence intervals.

Figure 5.2 shows several instances in which the nonparametric estimate of the (Marshallian) demand function is upward sloping. This anomaly is also present in the results of Hausman and Newey (1995). The theory of the consumer requires the compensated demand function to be downward sloping. Combined with a positive income derivative, an upward-sloping Marshallian demand function implies an upward-sloping compensated demand function and, therefore, is inconsistent with the theory of the consumer. At the median income, our nonparametric estimate of $\partial g/\partial y$ is positive over the range of prices of interest except for the two lowest grid points. Therefore, the nonparametric estimates are inconsistent with consumer theory. As is discussed in more detail in Section 5.4.3, we believe this result to be an artifact of random sampling errors and the consequent imprecision of the unconstrained nonparametric estimates. This motivates the use of the constrained estimation procedure, which increases estimation precision by imposing the Slutsky condition.

5.4.3 Nonparametric estimates under the Slutsky condition

Figure 5.3 shows the nonparametric estimates of the demand function, $\hat{g}_C$, at each of the three income levels of interest (solid dots). These estimates are constrained to satisfy the Slutsky condition and were obtained using the methods described in Section 5.2. For comparison, the figure also shows the unconstrained nonparametric estimates, $\hat{g}_U$ (open dots). The solid lines in Figure 5.3 connect the endpoints of joint 90% confidence intervals for $g(p, y)$. These were obtained using the bootstrap procedure described in Section 5.2.

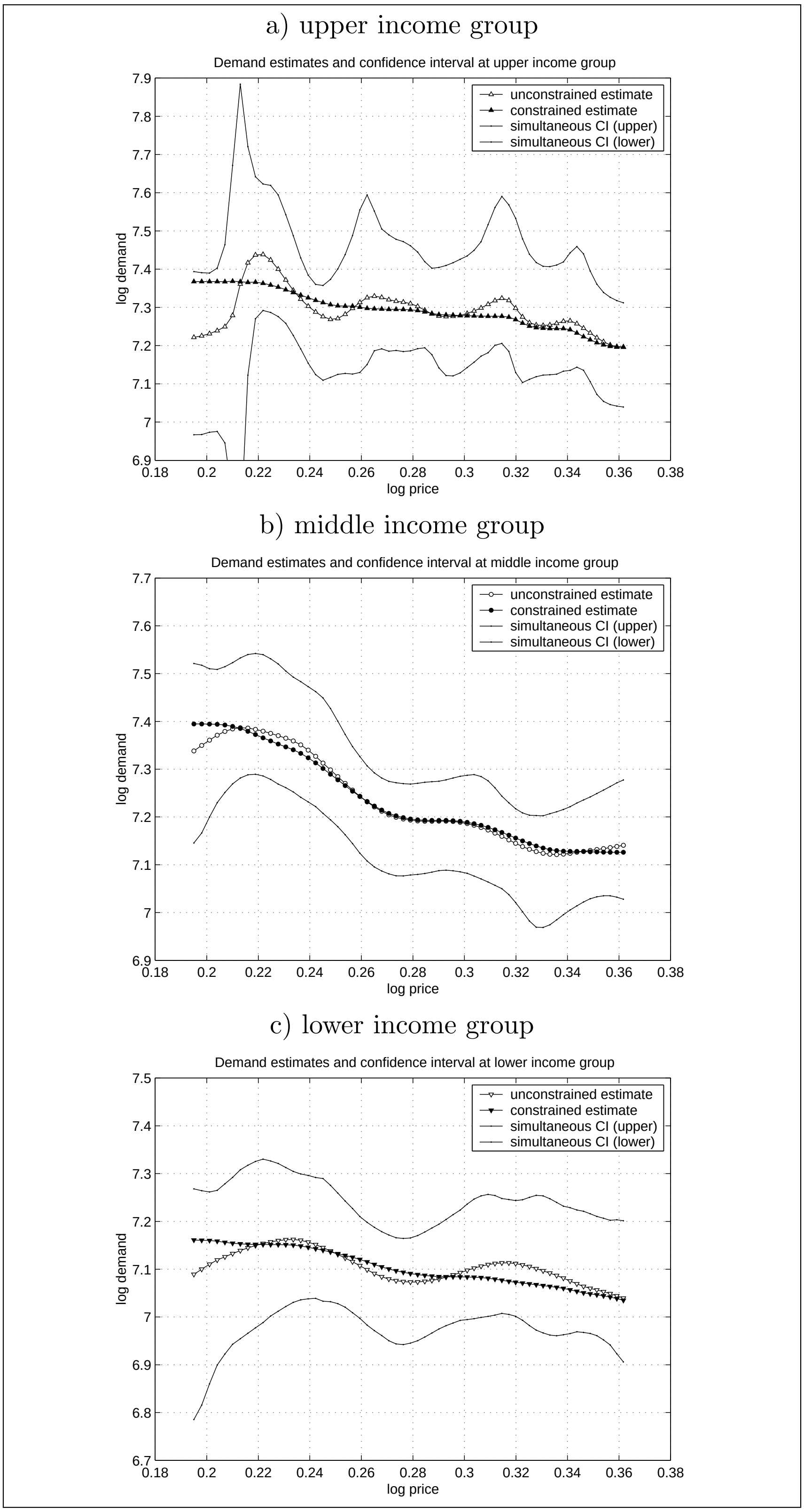
In contrast to the unconstrained estimates, the constrained estimates are downward sloping everywhere. The constrained estimates are also less wiggly than the unconstrained ones. In contrast to ad hoc 'ironing procedures' for producing monotonic estimates, $\hat{g}_C$ is consistent with the theory of the consumer and everywhere differentiable. This is important for estimation of DWL. The 90% confidence bands shown in Figure 5.3 contain both the constrained and unconstrained estimates. This is consistent with our view that the anomalous behavior of the unconstrained estimates is due to imprecision of the unconstrained estimator. It also indicates that the Slutsky constraint is consistent with the data.

The results in Figure 5.3 indicate that the middle income group is more sensitive to price changes than are the other two groups. In particular, the slope of the constrained estimate of g is noticeably larger for the middle group than for the other groups. This, in turn, suggests that the DWL of a tax increase is larger for the middle income group than for the others. We investigate this further in Section 5.4.4.

5.4.4 Estimates of deadweight loss

We now investigate the DWLs associated with several increases in gasoline taxes. The increases considered in the literature typically are quite large and often out of the support of the data. We start with an intervention that moves prices from the 5th to the 95th percentile of the price distribution in our sample. Historically observed tax

Figure 5.3: Demand estimates and simultaneous confidence intervals at different points in the income distribution



Note: Income groups correspond to \$72,500, \$57,500, and \$42,500. Confidence intervals shown refer to bootstrapped symmetrical, studentized simultaneous confidence intervals with a confidence level of 10%, based on 10,000 replications. See text for details.

changes in the U.S. tend to be much smaller than this, possibly due to the political difficulty of implementing large tax increases. To reflect the kind of intervention a legislature might actually consider, we also look at smaller interventions in which the price increases by \$0.05. As is well known, DWL increases with the square of the tax rate (e.g., Auerbach (1985)), so the DWL estimates are very different for the two types of interventions.

We compute DWL as follows. Over the range of the intervention, we evaluate the Marshallian demand estimates presented in the previous section for the three estimators (parametric, unconstrained nonparametric, and constrained nonparametric) on a grid of 61 points. We then use this demand estimate and the corresponding derivatives to compute the expenditure function and DWL by following the methods described in Section 5.2.

We study DWL relative to tax paid, which we interpret as a 'price' for raising tax revenue. We refer to this measure as relative DWL. Results are shown in Table 5.3.[4] Each panel of the table corresponds to one intervention. Intervention I moves prices from the 5th to the 95th price percentile in our data. The differences in the demand estimates between the different estimation methods translate into differences in relative DWLs. Comparing across income levels, the log-log linear model estimates relative DWL to be almost identical for the three income groups and indicates that the cost of taxation is about 7.6% of revenue raised for intervention I, irrespective of income level. In contrast, the constrained nonparametric estimates indicate that the cost of taxation is higher for the middle income group than for the other two groups. This result is consistent with our earlier finding that the middle income group is more responsive to price changes than are the other groups. The result also illustrates how the functional form assumptions of the parametric model affect estimates of consumer behavior and the effects of taxation.

[4]Confidence intervals for the unconstrained and the parametric model are reported in Table 5.5 in the Appendix 5.A.

Table 5.3: Relative Deadweight Loss estimates

	Income	DWL (as % of tax paid)		
		unconstrained	constrained	log-log
		(1)	(2)	(3)
Intervention I (\$1.215 – \$1.436)	\$72,500	10.09 %	10.18 %	7.59 %
	\$57,500	10.09 %	11.92 %	7.58 %
	\$42,500	6.40 %	6.70 %	7.56 %
Intervention II (\$1.22 – \$1.27)	\$72,500	4.20 %	3.27 %	1.80 %
	\$57,500	3.08 %	4.50 %	1.80 %
	\$42,500	-1.33 %	0.72 %	1.79 %
Intervention III (\$1.27 – \$1.32)	\$72,500	-1.06 %	0.84 %	1.73 %
	\$57,500	6.42 %	5.74 %	1.73 %
	\$42,500	3.86 %	2.82 %	1.72 %
Intervention IV (\$1.32 – \$1.37)	\$72,500	-3.02 %	0.49 %	1.67 %
	\$57,500	2.61 %	2.07 %	1.66 %
	\$42,500	-2.23 %	0.77 %	1.66 %

Note: For each intervention, the price change considered is indicated in round brackets (in U.S. dollars). Intervention I corresponds to moving prices from the 5th to the 95th percentile in the data. Interventions II, III and IV each increase price by five U.S. cents. Deadweight Loss is shown as percentage of tax paid after the (compensated) intervention. See text for details.

We also estimate the DWLs associated with taxes that increase the price by \$0.05 from several different initial values. Intervention II increases the price from \$1.22 to \$1.27, Intervention III from \$1.27 to \$1.32, and Intervention IV from \$1.32 to \$1.37. The results are shown in Table 5.3. The DWLs obtained from the log-log linear parametric model of the demand function are virtually constant across incomes. The DWLs obtained from the unconstrained nonparametric estimate of demand are sometimes negative. This anomalous result occurs because, due to random sampling errors, the unconstrained estimate of the demand function does not decrease monotonically and does not satisfy the integrability conditions of consumer theory. The constrained nonparametric model yields DWL estimates that are positive and, in some cases, more than double those obtained from the parametric model.

One can also study DWL relative to income so as to reflect the household's utility loss relative to available resources. The results for this analysis are shown in Table 5.4. The estimates from the parametric model and constrained nonparametric model give

Table 5.4: Deadweight Loss estimates relative to household income

	Income	DWL (relative to income) * 10^4		
		unconstrained (1)	constrained (2)	log-log (3)
	\$ 72,500	4.11	4.14	3.01
Intervention I	\$ 57,500	4.89	5.69	3.54
(\$ 1.215-1.436)	\$ 42,500	3.80	3.97	4.37
	\$ 72,500	0.43	0.34	0.18
Intervention II	\$ 57,500	0.41	0.59	0.21
(\$ 1.22-1.27)	\$ 42,500	-0.20	0.11	0.26
	\$ 72,500	-0.11	0.09	0.17
Intervention III	\$ 57,500	0.74	0.67	0.20
(\$ 1.27-1.32)	\$ 42,500	0.54	0.40	0.24
	\$ 72,500	-0.32	0.05	0.16
Intervention IV	\$ 57,500	0.29	0.23	0.18
(\$ 1.32-1.37)	\$ 42,500	-0.32	0.11	0.23

Note: For each intervention, the price change considered is indicated in round brackets (in U.S. dollars). Intervention I corresponds to moving prices from the 5th to the 95th percentile in the data. Interventions II, III and IV each increase price by five U.S. cents. Deadweight Loss is shown relative to baseline income. See text for details.

different indications of the effects of the tax increase across income groups. The parametric estimates indicate that the relative utility loss increases as income decreases. However, the constrained nonparametric estimates indicate that the relative utility loss is greater for the middle income group than for the other groups.

5.5 Conclusions

Simple parametric models of demand functions can yield misleading estimates of price sensitivity and welfare measures such as DWL, owing to misspecification. Fully nonparametric estimation of demand reduces the risk of misspecification but, because of the effects of random sampling errors, can yield imprecise estimates with anomalous properties such as non-monotonicity. This chapter has shown that these problems can be overcome by constraining nonparametric estimates to satisfy the Slutsky condition of economic theory. This stabilizes the nonparametric estimates without the need for parametric or other restrictions that have no basis in economic theory.

We have implemented this approach by using a modified kernel estimator that weights the observations so as to satisfy the Slutsky restriction. To illustrate the method, we have estimated a gasoline demand function for a class of households in the U.S. We find that some simple parametric specifications are rejected by a specification test, whereas a fully nonparametric estimate of the demand function is non-monotonic. In contrast, the estimate that is constrained to satisfy the Slutsky condition is well-behaved. Moreover, the constrained nonparametric estimates show patterns of price sensitivity that are very different from those of the simple parametric model. We find price responses vary non-monotonically with income. In particular, we find that low- and high-income consumers are less responsive to changes in gasoline prices than are middle-income consumers.

We have also computed the DWLs of several increases in the price of gasoline. We find that the unconstrained nonparametric estimates sometimes yield negative DWLs, which are inconsistent with economic theory and presumably caused by imprecision of the unconstrained estimates. The constrained nonparametric estimates of DWL are positive and, in many cases, quite different from those obtained with the parametric model. Mirroring the results on price responsiveness, the DWL estimates are highest for middle income groups. These results illustrate the usefulness of nonparametrically estimating demand functions subject to the Slutsky condition.

5.A Appendix

Table 5.5: Confidence intervals for DWL measures

	Income	DWL (as % of tax paid)				DWL (relative to income) * 10^4			
		unconstrained		log-log		unconstrained		log-log	
		lower	upper	lower	upper	lower	upper	lower	upper
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	\$ 72,500	1.24 %	23.03 %	4.81 %	10.33 %	1.36	9.09	1.99	4.04
Intervention I	\$ 57,500	-1.50 %	17.69 %	4.82 %	10.17 %	0.02	8.69	2.35	4.70
(\$ 1.215-1.436)	\$ 42,500	-7.24 %	15.21 %	4.81 %	10.28 %	-3.50	9.54	2.90	5.88
	\$ 72,500	-5.34 %	16.39 %	1.17 %	2.43 %	-0.50	1.63	0.11	0.25
Intervention II	\$ 57,500	-3.07 %	9.90 %	1.17 %	2.40 %	-0.37	1.36	0.13	0.28
(\$ 1.22-1.27)	\$ 42,500	-10.31 %	3.18 %	1.17 %	2.42 %	-1.56	0.51	0.17	0.36
	\$ 72,500	-11.26 %	6.37 %	1.13 %	2.34 %	-1.14	0.74	0.11	0.23
Intervention III	\$ 57,500	0.52 %	14.40 %	1.13 %	2.30 %	0.11	1.63	0.13	0.26
(\$ 1.27-1.32)	\$ 42,500	-1.99 %	12.78 %	1.13 %	2.33 %	-0.19	1.75	0.16	0.33
	\$ 72,500	-18.92 %	0.87 %	1.09 %	2.25 %	-2.11	0.32	0.10	0.21
Intervention IV	\$ 57,500	-1.81 %	7.77 %	1.09 %	2.22 %	-0.17	0.86	0.12	0.24
(\$ 1.32-1.37)	\$ 42,500	-8.29 %	2.45 %	1.09 %	2.24 %	-1.20	0.41	0.15	0.30

Note: For each intervention, the price change considered is indicated in round brackets (in U.S. dollars). Intervention I corresponds to moving prices from the 5th to the 95th percentile in the data. Interventions II, III and IV each increase price by five U.S. cents. Table shows confidence intervals corresponding to estimates reported in Tables 5.3 and 5.4. Confidence intervals are computed with an undersmoothed bandwidth, based on 5,000 replications. See text for details.

Chapter 6

Conclusion

This thesis applies microeconometric methods to understand determinants and effects of individual behavior relating to both educational choices and consumer demand. The work reported in Chapter 2 adds to our understanding of intergenerational effects of education by emphasizing the role of potential transmission channels in translating the effect of maternal education to the child. The results indicate that children from more educated mothers benefit from additional investments in a number of dimensions. The chapter also documents how the home environment in which children grow up is affected by maternal education, and shows that the effects of mother's education persist into early adulthood.

The work in Chapter 3 documents how mobility investments during higher education affect future international labor market mobility, and highlights that mobility programs can have long-lasting effects. The outcomes are measured one to five years after graduation, which is still relatively early in the individuals' labor market careers. In future research, it would be of interest to investigate how the effect evolves throughout the career of these university graduates. In particular, it would be of interest to investigate what fraction of the individuals who have moved abroad following their graduation eventually return to their home country.

Chapter 4 compares two types of vocational training forms which are of particular relevance for non-college bound youth. This question is important to policy-makers because governments can substitute between apprenticeship training and full-time vocational school, and the heterogeneity across countries in the design of vocational schooling systems is evidence for that. The key result in this work is that (former) apprentices have an advantage compared to vocational school graduates, but this advantage is transitory and fades out over time.

The evidence presented in Chapter 5 indicates that imposing economic restrictions is an attractive alternative to functional form assumptions or other semiparametric estimators. The constraint removes the erratic variation in the unconstrained estimate, and can be thought of as a potential substitute to bandwidth smoothing. Interestingly, the results indicate that the price responsiveness of gasoline demand varies non-monotonically with income: price responsiveness is highest for the middle income group, and lower for both the low-income and the high-income group. Since many of the frequently chosen functional forms impose a form of monotonicity on the price effect, these approaches would not be able to accurately represent the pattern found using the nonparametric approach. While we apply this method to study gasoline demand, the approach taken can be applied much more widely in the context of the study of consumer demand.

Bibliography

Acemoglu, D., and J.-S. Pischke (1998): "Why Do Firms Train? Theory and Evidence," *The Quarterly Journal of Economics*, 113(1), 79–119.

Adda, J., C. Dustmann, C. Meghir, and J.-M. Robin (2006): "Career Progression and Formal versus On-the-Job Training," *IZA Discussion Paper*, 2260.

Aizer, A. (2004): "Home Alone: supervision after school and child behavior," *Journal of Public Economics*, 88, 1835–1848.

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.

Antonovics, K. L., and A. S. Goldberger (2005): "Does Increasing Women's Schooling Raise the Schooling of the Next Generation? Comment," *The American Economic Review*, 95(5), 1738–1744.

Arkes, J. (2005): "Using Unemployment Rates As Instruments to Estimate Returns to Schooling," Unpublished manuscript.

Ascher, U. M., and L. R. Petzold (1998): *Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations*. SIAM.

Auerbach, A. J. (1985): "The Theory of Excess Burden and Optimal Taxation," in *Handbook of Public Economics*, ed. by A. J. Auerbach, and M. Feldstein, chap. 2, pp. 61–127. Elsevier.

BA (1991): *Berufsberatung 1989/90. Ergebnisse der Berufsberatungsstatistik*. Bundesanstalt für Arbeit, Nürnberg.

——— (2008): "Schlüsselverzeichnis für die Angaben zur Tätigkeit in den Meldungen zur Sozialversicherung," Federal Employment Agency.

Bartik, T. J. (1991): *Who Benefits from State and Local Economic Development Policies?* W.E. Upjohn Institute for Employment Research, Kalamazoo.

BEA (2008): "National Economic Accounts, Gross Domestic Product: Current-dollar and "real" GDP," U.S. Department of Commerce, Bureau of Economic Analysis. April 30, 2008. Available at `http://www.bea.gov/national/xls/gdplev.xls`.

Becker, G. S. (1962): "Investment in Human Capital: A Theoretical Analysis," *The Journal of Political Economy*, 70(5), 9–49.

——— (1993): *Human Capital. A Theoretical and Empirical Analysis, with Special Reference to Education.* The University of Chicago Press, Chicago and London, 3 edn.

Behrman, J. R., and M. R. Rosenzweig (2002): "Does Increasing Women's Schooling Raise the Schooling of the Next Generation?," *The American Economic Review*, 92(1), 323–334.

——— (2005): "Does Increasing Women's Schooling Raise the Schooling of the Next Generation? Reply," *The American Economic Review*, 95(5), 1745–1751.

Bender, S., A. Haas, and C. Klose (2000): "IAB Employment Subsample 1975-1995 Opportunities for Analysis Provided by the Anonymised Subsample," *IZA Discussion Paper*, 117.

Bento, A. M., L. H. Goulder, E. Henry, M. R. Jacobsen, and R. H. von Haefen (2005): "Distributional and Efficiency Impacts of Gasoline Taxes: An Econometrically Based Multi-Market Study," *The American Economic Review*, 95(2), 282–287.

Bento, A. M., L. H. Goulder, M. R. Jacobsen, and R. H. von Haefen (2009): "Distributional and Efficiency Impacts of Increased U.S. Gasoline Taxes," *American Economic Review*, forthcoming.

Bernal, R., and M. P. Keane (2006): "Child Care Choices and Children's Cognitive Achievement: The Case of Single Mothers," Manuscript, Northwestern University.

Berndt, E. R., and Z. Griliches (1990): "Price Indexes for Microcomputers: An Exploratory Study," *NBER Working Paper*, 3378.

Bertrand, M., E. Duflo, and S. Mullainathan (2004): "How much should we trust Differences-in-Differences Estimates?," *The Quarterly Journal of Economics*, 119(1), 249–275.

BJOERKLUND, A., M. LINDAHL, AND E. PLUG (2006): "The Origins of Intergenerational Associations: Lessons from Swedish Adoption Data," *The Quarterly Journal of Economics*, 121(3), 999–1028.

BLACK, S. E., P. J. DEVEREUX, AND K. G. SALVANES (2005): "Why the Apple Doesn't Fall Far: Understanding the Intergenerational Transmission of Human Capital," *The American Economic Review*, 91(1), 437–449.

BLANCHFLOWER, D. G., AND L. M. LYNCH (1994): "Training at Work: A Comparison of U.S. and British Youths," in *Training and the Private Sector*, ed. by L. M. Lynch, NBER Comparative Labor Market Series, chap. 8, pp. 233–260. The University of Chicago Press, Chicago and London.

BLAU, D., AND J. CURRIE (2003): "Preschool, Day Care, and After School Care: Who's Minding the Kids?," Manuscript.

BLOMQUIST, S., AND M. DAHLBERG (1999): "Small Sample Properties of LIML and Jackknife IV Estimators: Experiments with Weak Instruments," *Journal of Applied Econometrics*, 14(1), 69–88.

BLUNDELL, R., L. DEARDEN, AND C. MEGHIR (1996): "The Determinants and Effects of Work-Related Training in Britain," The Institute of Fiscal Studies.

BMBF (1992): *Berufsbildungsbericht 1992*, vol. 31 of *Grundlagen und Perspektiven für Bildung und Wissenschaft*. Federal Ministry of Education and Research, Bad Honnef.

BONNAL, L., S. MENDES, AND C. SOFER (2002): "School-to-work transition: apprenticeship versus vocational school in France," *International Journal of Manpower*, 23(5), 426–442.

BORJAS, G. J. (1987): "Self-Selection and the Earnings of Immigrants," *The American Economic Review*, 77(4), 531–553.

BOUND, J., J. GROEN, G. KÉZDI, AND S. TURNER (2004): "Trade in university training: cross-state variation in the production and stock of college-educated labor," *Journal of Econometrics*, 121(1-2), 143–173.

BOUND, J., D. A. JAEGER, AND R. M. BAKER (1995): "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak," *Journal of the American Statistical Association*, 90(430), 443–450.

Briedis, K., and K.-H. Minks (2004): "Zwischen Hochschule und Arbeitsmarkt. Eine Befragung der Hochschulabsolventinnen und Hochschulabsolventen des Prüfungsjahres 2001," *HIS Projektbericht*, 2004(March).

Buettner, T., and J. Rincke (2007): "Labor Market Effects of Economic Integration: The Impact of Re-Unification in German Border Regions," *German Economic Review*, 8(4), 536–560.

Cameron, S. V., and J. J. Heckman (2001): "The Dynamics of Educational Attainment for Black, Hispanic, and White Males," *Journal of Political Economy*, 109(3), 455–499.

Cameron, S. V., and C. Taber (2004): "Estimation of Educational Borrowing Constraints Using Retruns to Schooling," *Journal of Political Economy*, 112(1), 132–182.

Card, D. (1993): "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," *NBER Working Paper Series*, 4483.

——— (1999): "The Causal Effect of Education on Earnings," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 3A, chap. 30, pp. 1801–1863. Elsevier.

Carneiro, P., C. Crawford, and A. Goodman (2007): "Which Skills Matter," in *Practice Makes Perfect: The Importance of Practical Learning*, ed. by D. Kehoe. Social Market Foundation, forthcoming.

Carneiro, P., and J. J. Heckman (2003): "Human Capital Policy," in *Inequality in America: What Role for Human Capital Policies*, ed. by J. J. Heckman, and A. Krueger. MIT Press.

Carneiro, P., J. J. Heckman, and D. V. Masterov (2005): "Labor Market Discrimination and Racial Differences in Premarket Factors," *Journal of Law and Economics*, 48(1).

Carneiro, P., J. J. Heckman, and E. Vytlacil (2005): "Understanding What Instrumental Variables Estimate: Estimating Marginal and Average Returns to Education," Revised, May 20, 2005.

——— (2006): "Estimating Marginal and Average Returns to Education," Manuscript.

Chamberlain, G. (1984): "Panel Data," in *Handbook of Econometrics*, ed. by Z. Griliches, and M. D. Intriligator, vol. II, chap. 22, pp. 1247–1318. Elsevier.

Chevalier, A. (2004): "Parental Education and Child's Education: A Natural Experiment," *IZA Discussion Paper*, 1153.

Chevalier, A., C. Harmon, V. O'Sullivan, and I. Walker (2005): "The Impact of Parental Income and Education on the Schooling of Their Children," *IZA Discussion Paper*, 1496.

CHRR (2002): "NLSY79 Child & Young Adult Data Users Guide," Center for Human Resource Research. A Guide to the 1986–2000 Child Data, 1994–2000 Young Adult Data. The Ohio State University, Columbus, Ohio.

Coleman, J. S., E. Q. Campbell, C. J. Hobson, J. McPartland, A. M. Mood, F. D. Weinfeld, and R. L. York (1966): *Equality of Educational Opportunity*. US Department of Health, Education & Welfare. Office of Education, Washington, DC.

Commission of the European Communities (2002): "Commission's Action Plan for skills and mobility," *Official Journal*, (13/02/2002).

Council of the European Communities (1987): "Council Decision of 15 June 1987 adopting the European Community Action Scheme for the Mobility of University Students (Erasmus)," *Official Journal*, L166(25/06/1987), 20–24, cited after http://eur-lex.europa.eu.

Cunha, F., and J. Heckman (2006): "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," Manuscript, University of Chicago.

——— (2007): "The Technology of Skill Formation," *IZA Discussion Paper*, 2550.

Cunha, F., J. J. Heckman, L. Lochner, and D. V. Masterov (2005): "Interpreting the Evidence on Life Cycle Skill Formation," *NBER Working Paper Series*, 11331, National Bureau of Economic Research.

Currie, J. (2007): "Healthy, Wealthy, and Wise? Socioeconomic Status, Poor Health in Childhood, and Human Capital Development," Manuscript.

Currie, J., and E. Moretti (2002): "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings and Longitudinal Data," *NBER Working Paper Series*, 9360, National Bureau of Economic Research.

——— (2003): "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings," *The Quarterly Journal of Economics*, 118(4), 1495–1532.

Currie, J., and D. Thomas (1995): "Does Head Start Make a Difference?," *The American Economic Review*, 85(3), 341–364.

——— (2001): "Early Test Scores, Socioeconomic Status, School Quality and Future Outcomes," *Research in Labor Economics*, 20, 103–132.

DAAD (1992): *Studienland EG. Geförderte Kooperationsprogramme deutscher Hochschulen 1992/93*. Deutscher Akademischer Austauschdienst, Arbeitsstelle ERASMUS.

Dahl, C. A. (1979): "Consumer Adjustment to a Gasoline Tax," *The Review of Economics and Statistics*, 61(3), 427–432.

Dahl, G. (2002): "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets," *Econometrica*, 70(6), 2367–2420.

Dahl, G., and L. Lochner (2006): "The Impact of Family Income on Child Achievement," Manuscript.

Deaton, A., and J. Muellbauer (1980): *Economics and consumer behavior*. Cambridge University Press, reprinted 1999.

Dreher, Axel, and Poutvaara, Panu (2005): "Student Flows and Migration: An Empirical Analysis," *IZA DP*, 1612.

Duckworth, A. L., and M. E. P. Seligman (2005): "Self-discipline outdoes IQ in predicting academic performance of adolescents," *Psychological Science*, 16(12), 939944.

Dustmann, C. (2004): "Parental background, secondary school track choice, and wages," *Oxford Economic Papers*, 56, 209–230.

Dustmann, C., and U. Schönberg (2007): "The Effect of Expansions in Maternity Leave Coverage on Children's Long-Term Outcomes," Manuscript. University of Rochester.

EIA (2008a): "Annual Energy Review 2007," Energy Information Administration, U.S. Department of Energy.

——— (2008b): "Monthly Energy Review, April 2008," Energy Information Administration, U.S. Department of Energy.

Federal Institute for Vocational Education and Training (2006): "Charts on Vocational Education," in German. available online at `http://www.bibb.de/dokumente/pdf/a22_ausweitstat_schaubilder_heft-2006.pdf`.

——— (2008): "Training Profiles," Available online at http://www.bibb.de/en/ausbildungsprofile-start.htm.

Federal Statistical Office (2007): "Bildung und Kultur: Berufliche Schulen," Fachserie 11 Reihe 2. Schuljahr 2006/07.

Fersterer, J., J.-S. Pischke, and R. Winter-Ebmer (2008): "Returns to Apprenticeship Training in Austria: Evidence from Failed Firms," Manuscript, London School of Economics.

Findlay, R., and H. Kierzkowski (1983): "International Trade and Human Capital: A Simple General Equilibrium Model," *The Journal of Political Economy*, 91(6), 957–978.

Fitzenberger, B., A. Osikominu, and R. Völter (2005): "Imputation Rules to Improve the Education Variable in the IAB Employment Subsample," *Discussion Paper*, 3, Federal Employment Agency.

Freeman, R. B. (2009): "What Does Global Expansion of Higher Education Mean for the US?," *NBER Working Paper*, 14962.

Fryer, R. G. J., and S. D. Levitt (2004): "Understanding the Black-White Test Score Gap in the First Two Years of School," *The Review of Economics and Statistics*, 86(2), 447464.

——— (2006): "The Black-White Test Score Gap Through Third Grade," *American Law and Economics Review*, 8(2), 249–281.

——— (2007): "Testing for Racial Differences in the Mental Ability of Young Children," Manuscript, Harvard University.

Galindo-Rueda, F. (2003): "The Intergenerational Effect of Parental Schooling: Evidence from the British 1947 School Leaving Age Reform," Manuscript, London School of Economics.

Gehmlich, Volker (2000): *Der DAAD begleitet ERASMUS.* Deutscher Akademischer Austauschdienst (DAAD).

Goldin, C. (2001): "The Human-Capital Century and American Leadership: Virtues of the Past," *The Journal of Economic History*, 61(2), 263–292.

Greenwood, Michael J. (1975): "Research on Internal Migration in the United States: A Survey," *Journal of Economic Literature*, 13(2), 397–433.

——— (1985): "Human Migration: Theory, Models, and Empirical Studies," *Journal of Regional Science*, 25(4), 521–544.

——— (1997): "Internal Migration in Developed Countries," in *Handbook of Population and Family Economics*, ed. by M. R. Rosenzweig, and O. Stark, vol. 1, pp. 647–720. Elsevier.

Groen, J. A. (2004): "The effect of college location on migration of college-educated labor," *Journal of Econometrics*, 121(1-2), 125–142.

Guellec, Dominique, and Cervantes, Mario (2001): "International Mobility of Highly Skilled Workers: From Statistical Analysis to Policy Formulation," in *International Mobility of the Highly Skilled*, pp. 71–98. OECD.

Hall, P. (1992): *The Bootstrap and Edgeworth Expansion.* Springer.

Hall, P., and L.-S. Huang (2001): "Nonparametric Kernel Regression Subject to Monotonicity Constraints," *The Annals of Statistics*, 29(3), 624–647.

Hall, R. E. (1971): "The Measurement of Quality Change from Vintage Price Data," in *Price Indexes and Quality Change*, ed. by Z. Griliches, pp. 240–271. Harvard University Press.

Hansen, K. T., J. J. Heckman, and K. J. Mullen (2004): "The effect of schooling and ability on achievement test scores," *Journal of Econometrics*, 121, 39–89.

Härdle, W. (1990): *Applied Nonparametric Regression*, vol. 19 of *Econometric Society Monograph Series.* Cambridge University Press.

Härdle, W., and J. S. Marron (1991): "Bootstrap Simultaneous Error Bars for Nonparametric Regression," *The Annals of Statistics*, 19(2), 778–796.

Hausman, J. A., and W. K. Newey (1995): "Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss," *Econometrica*, 63(6), 1445–1476.

Haveman, R., and B. Wolfe (1995): "The Determinants of Children's Attainments: A Review of Methods and Findings," *Journal of Economic Literature*, 33(4), 1829–1878.

Heckman, J. (1993): "Assessing Clinton's Program on Job Training, Workfare, and Education in the Workplace," *NBER Working Paper*, 4428.

Heckman, J., and R. Robb (1985): "Using Longitudinal Data to Estimate Age, Period and Cohort Effects in Earnings Equations," in *Cohort Analysis in Social Research. Beyond the Identification Problem*, ed. by W. M. Mason, and S. E. Fienberg, pp. 137–150. Springer-Verlag.

Heckman, J. J., R. L. Roselius, and J. A. Smith (1993): "U.S. Education and Training Policy: A Re-evaluation of the Underlying Assumptions Behind the "New Consensus"," *Harris School Working Paper Series*, 93.04.

Heckman, J. J., J. Stixrud, and S. Urzua (2006): "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 24(3), 411–482.

Heckman, J. J., and E. J. Vytlacil (2007): "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast Their Effects in New Environments," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6B, chap. 71, pp. 4875–5143. Elsevier.

Hicks, John R. (1932): *The Theory of Wages*. Macmillan.

Holmlund, H., M. Lindahl, and E. Plug (2006): "Estimating Intergenerational Schooling Effects: A Comparison of Methods," Manuscript.

Hoxby, C. (1997): "How the Changing Market Structure of U.S. Higher Education Explains College Tuition," *NBER Working Paper Series*, 6323, National Bureau of Economic Research.

Hughes, J. E., C. R. Knittel, and D. Sperling (2008): "Evidence of a Shift in the Short-Run Price Elasticity of Gasoline Demand," *The Energy Journal*, 29(1), 93–114.

Imbens, G. W., and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–475.

Isserstedt, Wolfgang and Schnitzer, Klaus (2002): *Internationalisierung des Studiums. Ergebnisse der 16. Sozialerhebung des Deutschen Studentenwerks*. Bundesminsterium für Bildung und Forschung.

Jacobson, L. S., R. J. LaLonde, and D. G. Sullivan (1993): "Earnings Losses of Displaced Workers," *The American Economic Review*, 83(4), 685–709.

Jahr, V., and U. Teichler (2000): "Employment and Work of Former ERASMUS Students," in *Socrates 2000 Evaluation Study*, ed. by U. Teichler, J. Gordon, and F. Maiworm, chap. 5, pp. 99–117. Study for the European Commission.

Jencks, C., and M. Phillips (1998): *The Black-White Test Score Gap*. Brookings Institution Press.

Johnson, M., and M. P. Keane (2007): "A Dynamic Equilibrium Model of the U.S. Wage Structure, 1968-1996," Manuscript.

Kambourov, G., and I. Manovskii (2008): "Occupational Specificity of Human Capital," Manuscript. University of Toronto.

Kane, T. J., and C. E. Rouse (1993): "Labor Market Returns to Two- and Four-Year College: Is a Credit a Credit and Do Degrees Matter?," *Working Paper*, 311, Industrial Relations Section, Princeton University.

Kemp, M. C. (1964): *The Pure Theory of International Trade*. Prentice-Hall.

Kling, J. R. (2001): "Interpreting Instrumental Variables Estimates of the Returns to Schooling," *Journal of Business & Economic Statistics*, 19(3), 358–364.

KMK (2006): *Rahmenvereinbarung über die Berufsfachschulen 35*. Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany, Available online at www.kmk.org/doc/beschl/RVBFS06-10-10.pdf.

——— (2008): *The Education System in the Federal Republic of Germany 2006*. Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany, available online at `http://www.kmk.org/dossier/dossier_en_ebook.pdf`.

Knudsen, E. I., J. J. Heckman, J. L. Cameron, and J. P. Shonkoff (2006): "Economic, Neurobiological and Behavioral Perspectives on Building America's Future Workforce," *NBER Working Paper*, 12298.

KODRZYCKI, YOLANDA K. (2001): "Migration of Recent College Graduates: Evidence from the National Longitudinal Survey of Youth," *New England Economics Review*, 1.

KRUEGER, A. B. (1999): "Experimental Estimates of Education Production Functions," *The Quarterly Journal of Economics*, 114(2), 497–532.

LOWELL, B. LINDSAY (2007): "Trends in International Migration Flows and Stocks, 1975-2005," *OECD Social, Employment and Migration Working Papers*, 58.

MAIWORM, F., W. STEUBE, AND U. TEICHLER (1993): "ERASMUS Student Mobility Programmes 1989/90 in the View of their Coordinators," *Werkstattberichte, Uni Kassel*, 41.

MALAMUD, O., AND C. POP-ELECHES (2008): "General Education vs. Vocational Training: Evidence from an Economy in Transition," *NBER Working Paper*, 14155.

MAURIN, E., AND S. MCNALLY (2005): "Vive la Révolution! Long term returns of 1968 to the angry students," unpublished manuscript, February 2005.

MCKENZIE, L. W. (1955): "Equality of Factor Prices in World Trade," *Econometrica*, 23(3), 239–257.

MCLANAHAN, S. (2004): "Diverging Destinies: How Children Are Faring Under the Second Demographic Transition," *Demography*, 41(4), 607–627.

MOULTON, B. R. (1990): "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units," *The Review of Economics and Statistics*, 72(2), 334–338.

MÜSSIG-TRAPP, P., AND K. SCHNITZER (1997): *Vorbereitung auf Europa durch Mobilität und Internationalisierung des Studiums*. Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie.

NADARAYA, E. (1964): "On Estimating Regression," *Theory of Probability and its Applications*, 9(1), 141–142.

NCCE (2008a): "The Cooperative Education Model," National Commission for Cooperative Education, available online at www.co-op.edu.

——— (2008b): "Trends in Higher Education and the Need for the Integration of Learning and Work," National Commission for Cooperative Education. Available online at www.co-op.edu/resources.htm.

NCES (2008): "Percent of students who earn various percentages of their total course credits in CTE," National Center for Education Statistics. U.S. Department of Education.

Neal, D. (2005): "Why has the black-white skill convergence stopped?," Manuscript.

OECD (2002): *International Mobility of the Highly Skilled*. OECD Publishing, Paris.

OECD (2006): *OECD Factbook - Economic, Environmental and Social Statistics*. OECD Publishing, cited after http://www.sourceoecd.org/factbook.

OECD (2008): "Education at a Glance," Paris.

Oosterbeek, H., and D. Webbink (2007): "Wage effects of an extra year of basic vocational education," *Economics of Education Review*, 26, 408419.

——— (2009): "Does studying abroad induce a brain drain?," *Economica*, forthcoming.

Oreopoulos, P., M. Page, and A. H. Stevens (2008): "The Intergenerational Effects of Worker Displacement," *Journal of Labor Economics*, 26(3), 455–483.

Oreopoulos, P., M. E. Page, and A. H. Stevens (2003): "Does Human Capital Transfer from Parent to Child? The Intergenerational Effects of Compulsory Schooling," November 2003.

ORNL (2004): "2001 National Household Travel Survey. User Guide," Oak Ridge National Laboratory. Available at `http://nhts.ornl.gov/2001/`.

Plug, E. (2004): "Estimating the Effect of Mother's Schooling on Children's Schooling Using a Sample of Adoptees," *The American Economic Review*, 94(1), 358–368.

Plug, E., and W. Groot (1998): "Apprenticeship versus vocational education: Exemplified by the Dutch situation," Unpublished manuscript, Amsterdam School of Economics.

Poterba, J. N. (1991): "Is the Gasoline Tax Regressive?," *NBER Working Paper*, 3578.

Ramsey, J. (1969): "Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis," *Journal of the Royal Statistical Society. Series B (Methodological)*, 31(2), 350–371.

Robinson, P. M. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56(4), 931–954.

Rothenberg, T. J. (1971): "Identification in Parametric Models," *Econometrica*, 39(3), 577–591.

Ryan, P. (1998): "Is Apprenticeship Better? A Review of the Economic Evidence," *Journal of Vocational Education & Training*, 50(2), 289–329.

——— (2001): "The School-to-Work Transition: A Cross-National Perspective," *Journal of Economic Literature*, 39(1), 34–92.

Sacerdote, B. (2002): "The Role of Nature versus Nurture in Determining Economic Outcomes," *The American Economic Review*, 92(2), 344–348, Papers and Proceedings of the One Hundred Fourteenth Annual Meeting of the American Economic Association.

Schmalensee, R., and T. M. Stoker (1999): "Household Gasoline Demand in the United States," *Econometrica*, 67(3), 645–662.

Schwerdt, W., and S. Bender (2003): "Was tun Lehrlinge nach ihrer Ausbildung? Eine Analyse mit dem Linked Employer-Employee-Datensatz des IAB," *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung*, 1, 46–59.

Smith, A. (1988): "Erasmus - eine Investition in die Zukunft der Gemeinschaft," in *Erasmus - Wege zum Studium*. University of Cologne, cited in Gehmlich (2000).

Sollogoub, M., and V. Ulrich (1999): "Les jeunes en apprentissage ou en lycée professionnel. Une mesure quantitative et qualitative de leur insertion sur le marché du travail," *Économie et Statistique*, 323(3), 31–52.

Solon, G. (1999): "Intergenerational Mobility in the Labor Market," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 3, chap. 29, pp. 1760–1800. Elsevier.

Soskice, D. (1994): "Reconciling Markets and Institutions: The German Apprenticeship System," in *Training and the Private Sector*, ed. by L. M. Lynch, NBER Comparative Labor Market Series, chap. 1, pp. 25–60. The University of Chicago Press, Chicago and London.

Staiger, D., and J. H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65(3), 557–586.

STOCK, J. H., J. H. WRIGHT, AND M. YOGO (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business & Economic Statistics*, 20(4), 518–529.

SULLIVAN, D., AND T. VON WACHTER (2007): "Mortality, Mass-Layoffs, and Career Outcomes: An Analysis using Administrative Data," *NBER Working Paper*, 13626.

TODD, P. E., AND K. I. WOLPIN (2003): "On the Specification and Estimation of the Production Function for Cognitive Achievement," *The Economic Journal*, 113, F3–F33.

——— (2006): "The Production of Cognitive Achievement in Children: Home, School and Racial Test Score Gaps," Manuscript, University of Pennsylvania.

TOPEL, R. H., AND M. P. WARD (1992): "Job Mobility and the Careers of Young Men," *The Quarterly Journal of Economics*, 107(2), 439–479.

TROLTSCH, K., L. ALEX, R. VON BARDELEBEN, AND J. G. ULRICH (1999): "Jugendliche ohne Berufsausbildung. Eine BIBB/EMNID-Untersuchung," Federal Ministry of Education and Research.

WATSON, G. S. (1964): "Smooth Regression Analysis," *Sankhyā: The Indian Journal of Statistics, Series A*, 26(4), 359–372.

WERWATZ, A. (2002): "Occupational Mobility after Apprenticeship – How Effective is the German Apprenticeship System?," *Konjunkturpolitik*, 48(3–4), 279–303.

WEST, S. E. (2004): "Distributional effects of alternative vehicle pollution control policies," *Journal of Public Economics*, 88, 735–757.

WINKELMANN, R. (1996): "Employment Prospects and Skill Acquisition of Apprenticeship-Trained Workers in Germany," *Industrial and Labor Relations Review*, 49(4), 658–672.

WITTE, J. C., AND A. L. KALLEBERG (1995): "Matching Training and Jobs: The Fit between Vocational Education and Employment in the German Labour Market," *European Sociological Review*, 11(3), 293–317.

YATCHEW, A., AND J. A. NO (2001): "Household Gasoline Demand in Canada," *Econometrica*, 69(6), 1697–1709.

ZVS (2009): "zvs online: FAQ," Zentralstelle für die Vergabe von Studienplätzen. Available online at `http://www2.zvs.de/index.php?id=24`.