Immigrants' Performance, Welfare and Reception: An Economic Analysis for the UK

Francesca Fabbri

University College London

Department of Economics

August 2004

A Thesis submitted for the degree of Ph.D.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

This thesis analyses several aspects of the performance, welfare and reception of immigrants in the UK. The thesis is organised into three parts and contains the introduction and six chapters.

In the introduction, we provide the motivation and establish the background for the analysis performed in the thesis. We discuss recent trends in migration in the UK and describe the data sources used for the analysis.

In *part one* we analyse the economic performance of immigrants, and how this is affected by the level of proficiency in the host country language. We also investigate the association between ethnic concentration and language fluency. In *chapter two*, the effect of language proficiency in English is investigated as a determinant of wages and employment probabilities, taking into account econometric issues such as endogeneity and measurement error. This is followed by *chapter three*, which investigates the association between immigrants' language proficiency and ethnic context.

In *part two*, *chapter four* presents an analysis of health inequalities between ethnic minority immigrants and the native population in England. A model of migration decisions is developed that includes health as a determinant of migration. According to this model, if health and income are positively correlated, immigrants are likely to be positively selected in terms of health. This discussion illustrates the problems deriving from the available measures of health.

Part two deals with issues of "reception". In *chapter five*, the impact of immigration on the UK labour market is investigated. We use aggregate information at regional level. The analysis concentrates on employment and wage effects of immigration. Finally, *chapter six* provides an analysis of the association between ethnic concentration, attitudes of the majority population and the probability of ethnic minorities experiencing racial hostility.

Acknowledgements

I would like to thank Christian Dustmann and Ian Preston for their extraordinarily stimulating and patient supervision all through the years. I would also like to thank them for making me participate to the Research Project on the Labour Market Effects of Migration funded by the Home Office.

For financial support, I gratefully acknowledge the Marie Curie Research Training Grant (Contract no. ERBFMBICT 972810) and the Frederick Bonnart-Braunthal scholarship.

I would like to thank James Nazroo for kindly providing some of the data used in the thesis in its original form which made much of this work possible.

I am grateful to Wendy Carlin for her invaluable advice and support in the last years of my PhD.

I would like to thank Marcos Vera-Hernandez for his help, and, in particular, for his friendship.

Special thanks to Zsuzsa, Filipa, Greg, Iolanda, Joel, Noa, Rob and Tom for their love, encouragement and emotional support.

I am grateful to Sigrid, without whose help the last months of this PhD would have been much harder.

I want to express gratitude to my mother for always encouraging my studies abroad.

Finally, I would like to thank Georg for the endless discussions, and his help with the editing of this work. Mostly, I would like to thank him for the love and immense support he has given me.

Contents

List of Figures

List of Tables

Statement Concerning Conjoint Work

I declare formally that the material in Chapter 5 ("The Local Labour Market Effects of Immigration in the UK"), and Chapter 6 ("Ethnic Concentration, Prejudice and Racial Harassment towards Minorities in the UK") of this thesis is conjoint work with my supervisors, Dr. Christian Dustmann and Dr. Ian Preston. For each of these chapters I claim one third of the credit. The material in Chapter 2 ("Language Proficiency and Labour Market Performance of Immigrants in the UK") and Chapter 3 ("Ethnic Enclaves and Language Proficiency of Immigrants") is conjoint work with my supervisor Dr. Christian Dustmann. For each of these chapters I claim half credit. The remaining work in this thesis is mine.

Francesca Fabbri

Chapter 1

Introduction

According to the latest figures from the 2001 Population Census, some 4.9 million individuals living in the UK, 8.3% of the population, were born in another country. Foreign-born individuals differ from their UK-born peers in education, demographic structure, culture, and skills. These differences may determine the economic success and the degree of social integration of immigrants in the UK. Over time, immigrants may adjust in many respects to natives, due to the accumulation of skills and of information, and the adoption of new habits.

Migration is an increasingly prominent and controversial issue in the political debate[1]. It has implications for multiple areas of government policy, such as education, employment, health and social cohesion. It is an important pre-requisite for migration policy to understand how immigrants perform, and what are the determinants of their achievements in the UK labour market. Similarly important is to evaluate other aspects of their welfare, which greatly depend on the degree of social inclusion that immigrants enjoy in the host country.

Intensive research on migration has been carried forward in the US [2]. Research has also been conducted in other countries, such as Canada, Australia, Germany, and Israel[3].

[1]Political and media discussion has recently focused mainly on asylum seekers and refugees. In the past five years, the number of asylum seekers and refugees in the UK has dramatically increased (see Heath and Hill [2002]). However, they still represent a small fraction of the immigrant population (Dobson et al. [2001]). To our knowledge, there are no data on the social and economic conditions of asylum seekers and refugees in the UK.

[2]See Borjas [1999] for a recent survey on the literature.

[3]See Zimmermann and Bauer [2002] for a collection of papers on the economics of migration with emphasis on empirical and policy-relevant work related to North America, Europe, and Australia.

In contrast, in Britain, relatively less evidence exists on the subject. Most contributions are very recent and analyse mainly the labour market performance of immigrants. Studies on earnings adaptation of immigrants in the UK include Chiswick [1980] and Bell [1997], who find different adaptation rates and entry wage differentials across immigrant groups. Blackaby et al. [1997] and Wheatley Price [2001] investigate differences in the incidence of unemployment between immigrants and natives in the UK. Their findings show that white and non-white immigrants have a lower probability of being employed, compared to white natives. This disadvantage decreases over time for white immigrants, but it does not disappear for non-white immigrants. Finally, Shields and Wheatley Price [2002] analyse the determinants of immigrants' language fluency, and the effect of language on economic outcomes.

The aim of this thesis is to fill some of the gaps in the migration literature for the UK, providing a more comprehensive picture of the welfare and the process of adaptation of immigrants. Throughout the thesis, immigrants are defined as individuals who were born outside the UK. Our main focus is on ethnic minority immigrants. According to the 1991 Population Census, ethnic minorities are people who identify themselves as being of racial or ethnic heritage other than white. In 2000, ethnic minority immigrants represent 49% of the total immigrant population in the UK (Labour Force Survey).

Ethnic minority immigrants mainly come from New Commonwealth countries, and Pakistan. The well-being of ethnic minorities and inter-ethnic relations with the white majority have been subject of much public and political debate. Different education, cultural and demographic backgrounds of ethnic minorities may become disadvantages in some social and economic sectors, and advantages in others. Such differences are likely to be stronger for ethnic minorities than for other white immigrants (who, we will see, mainly come from EU countries). Evidence from Chiswick [1980], and Bell [1997], for example, indicates that most ethnic minority groups have lower wages and lower participation and employment rates than white natives and other immigrants, even after several years of stay in the UK. However, some groups, like the Indians, are often found to outperform white natives.

Several data sets are available which oversample ethnic minorities[4]. These data sets are particularly interesting because they contain much information on the cultural and

[4]As immigrants represent a small portion of the UK population, their sample sizes in survey data are small, which renders results imprecise.

demographic characteristics and the social and economic conditions of ethnic minorities. We will see that information of similar quality is not available for white immigrants.

The thesis is divided into three parts. In the first part, we investigate the determinants of language proficiency, and how it affects the employment probabilities and earnings of non-white immigrants. Unlike the previous literature, we address the problem of endogenous choice of language acquisition and measurement error in language variables. We extend the analysis on language determinants focusing on its correlation with ethnic enclaves. The aim is twofold: to evaluate whether living in an ethnic enclave can prevent immigrants from becoming proficient in English and to understand how language proficiency at arrival can determine the individual's location choice. In the second part, we analyse health differentials between immigrants and natives and try to infer whether immigrants are positively selected in terms of health. Finally, in the third part, we research the quality of the "reception" of immigrants in the UK from two different perspectives. On the one hand, relations between immigrants and natives are likely to be affected by the impact migration may have on the economy of the host country. As one aspect of this impact, we analyse the effect of immigration on the local labour market. On the other hand, to investigate the degree of social cohesion in the UK, we analyse the correlation between ethnic concentration and hostile attitudes towards ethnic minorities, and the probability of ethnic minorities experiencing racial hostility.

The aim of this introduction is to establish a context and present a common background for the following chapters of the thesis. To do that, we first outline recent migration trends for the UK. We then describe the data sets used in the analysis and their characteristics. Finally, we provide an overview of the thesis.

1.1 Migration in the UK

This section presents an outline of the recent migration trends in the UK. Evidence is reported for immigrants as a whole and disaggregated by country of origin, with special focus on ethnic minorities.

In the past 50 years, the United Kingdom has been the destination of large numbers of immigrants coming from New Commonwealth countries. Until the early '60s, citizens of Britain's colonies were not just granted special immigration status (i.e. the right to freely enter, work and settle with their families), but were actively encouraged to migrate

both by the government and employers (such as London Transport and NHS). Indeed Britain, like most other rich and industrialized countries, was facing continuing labour shortages. As a consequence, to match its labour demand, the UK received consecutive tides of immigrants between 1950 and 1970. These immigrants arrived first from the Caribbean, then gradually from India, Pakistan, Bangladesh, Africa and the Far East.

As a consequence of increasing racial tensions and of the falling excess demand for labour in the country, the British government produced three Commonwealth Immigrants Acts (in 1962, 1968 and 1971), which, as immigration control measures, gradually restricted the opportunities for migrants from the New Commonwealth to enter and work in Britain.

After the 1981 Nationality Act, in particular, citizens from Commonwealth countries ceased to have any privileged right to settle in the UK. As a consequence, the new immigration flows of ethnic minorities was mostly based on family reunification[5].

For cultural and demographic reasons, ethnic minorities are likely to differ from the UK native population to a higher extent than white immigrants from countries such as the EU and the US. These differences may hinder their integration process and attract the diffidence and hostility of the majority population, possibly leading to discrimination and racial harassment.

Figure 1.1 outlines the historical pattern of immigration into Britain, using data from the 2000 Labour Force Survey. We focus on the population of working age (men aged 16-64 and women aged 16-59). The figure shows that a large fraction of working age immigrants are recent arrivals. Around 8 per cent of all immigrants have arrived within the last year, and around one third have arrived within the last ten years.

Figure 1.2 charts the year of arrival of immigrant groups from different countries of origin. Immigration flows immediately after the war were predominantly from the Caribbean and Ireland. The 1960s and 1970s saw a large number of arrivals from India and Pakistan. Migration from Bangladesh and China occurred mainly between the late 1970s and the early 1980s. Migration from the European Union developed in a relatively constant flow along the whole period. Migration from the Old Commonwealth countries (including the USA) was higher between 1980 and 1990. A large number of immigrants from the African continent arrived in the 1990s. The diversity of immigrants in the UK,

[5]See Hatton and Wheatley Price [2002] for a comprehensive survey on the UK's international migration experience from after the Second World War.

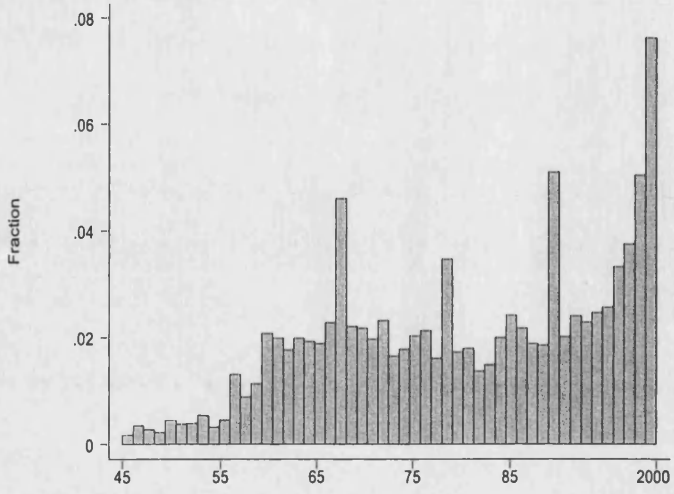


Figure 1.1: Distribution of immigrants by year of entry

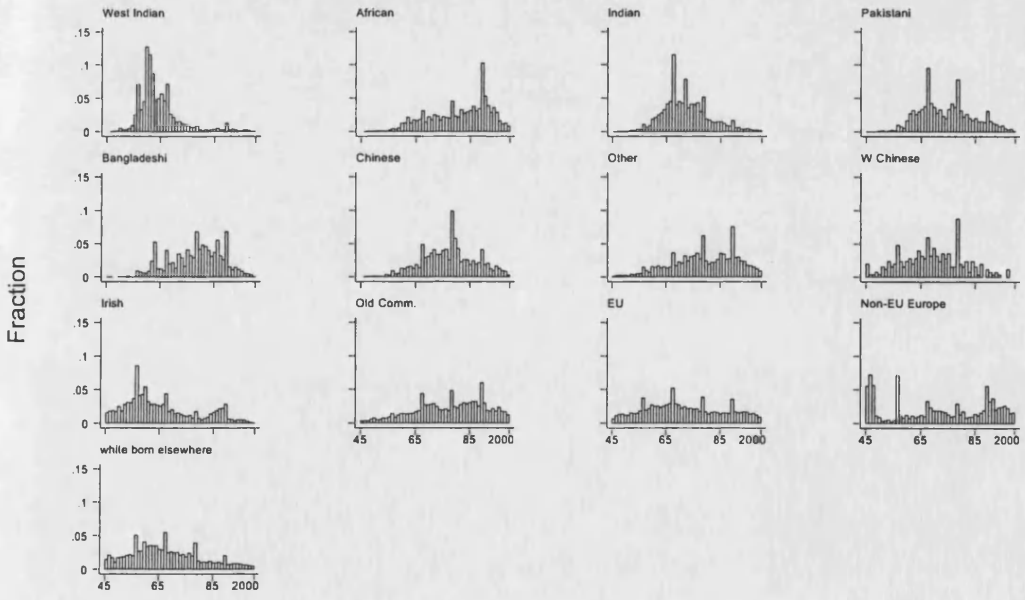


Figure 1.2: Year of arrival by origin status

measured by country of origin, has increased recently.

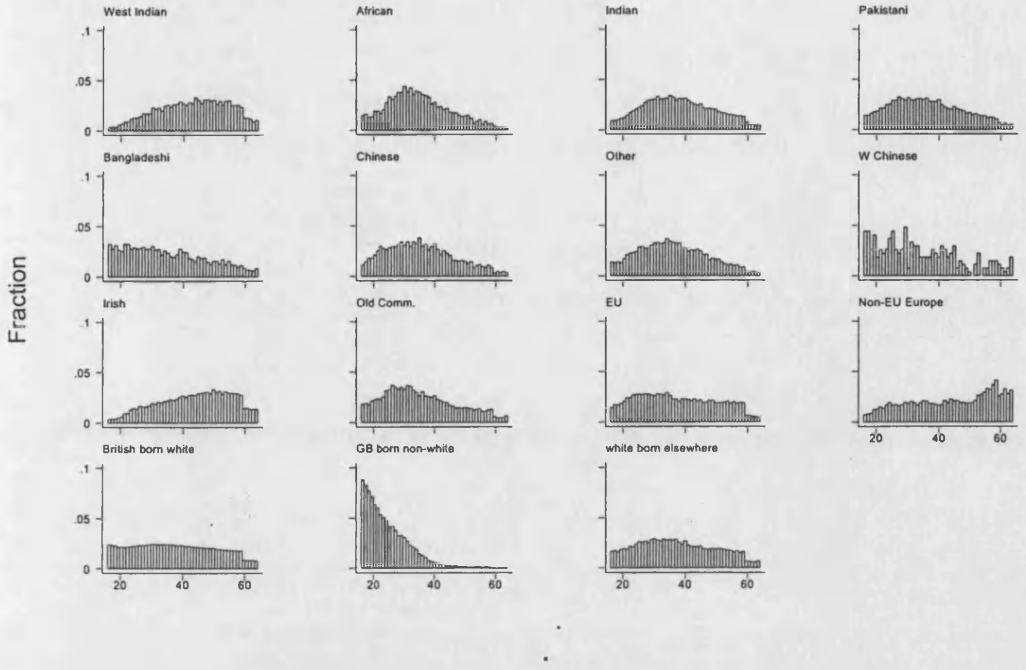


Figure 1.3: Distribution of immigrants by age

Figure 1.3 graphs the age distribution of the various immigrant communities. Since the West Indian community has been in Britain the longest, the age distribution is skewed to the right, with correspondingly fewer arrivals now in their teens or twenties. In contrast the age profiles of African and Bangladeshi immigrants are skewed to the left, with much higher concentrations of individuals in the younger age range, reflecting the more recent entry into Britain. The age profiles of European Union immigrants resemble that of white natives most closely.

Combining the information deriving from the arrival patterns of non-white immigrants and the age distribution of non-white natives, shows that non-white natives are mostly second generation immigrants (that is, children of individuals who were born outside the UK). The possibility to separately identify first and second generation immigrants is important: it provides more information on the individual's cultural background, and on the social and economic adaptation of subsequent generations of immigrants. This information is usually not available for other immigrants.

In Table 1.1, we highlight some basic facts about the various minority groups in Britain. The numbers are taken from the 1979, 1983, and 2000 Labour Force Surveys (LFS), and refer to the population of working age (year of arrival and education data are only available, in full, from 1983 onward).

Table 1.1: Immigrants and UK-born whites in Britain (population of working age)

	Year	UK-born white	UK-born minorities	Immigrants	West Indian	African	Indian	Pakistani	Bangladeshi	Chinese	Other non-white	Irish	Old Comm.	EU	Non-EU	Other white
% of pop.	1979	92.2	0.5	7.3	0.7	0.3	1.2	0.4	0.1	0.2	0.6	1.4	0.4	1	0.4	0.9
	2000	88.3	2.4	9.3	0.4	0.6	1.2	0.7	0.3	0.2	1.2	0.8	0.8	1.4	0.4	1.3
Med. Age	1979	36	19	37	39	30	33	31	37	29	31	42	34	40	53	35
	2000	39	27	38	46	35	43	37	33	37	37	47	34	33	32	39
Med. yrs since mig.	1983	–	–	18	22	11	14	14	10	8	10	26	11	21	35	22
	2000	–	–	19	34	9	25	20	16	14	12	32	11	16	5	28
Med. entry age	2000	–	–	20	16	24	19	19	17	22	23	18	23	19	22	11
% arrive < age15	2000	–	–	34	46	14	30	37	46	25	22	32	33	37	17	57
% grad.(men)	1983	10	3	15	4	20	16	8	11	14	27	5	36	16	11	21
	2000	16	18	21	5	33	23	12	7	31	23	15	26	20	11	32
% No quals.	1983	46	35	47	65	10	42	67	85	47	23	72	22	40	61	30
	2000	14	13	16	38	9	16	35	41	21	12	25	5	9	16	8
% grad.(fem)	1983	4	2	9	1	7	9	4	2	9	13	3	26	10	16	13
	2000	12	16	16	10	12	14	6	6	23	14	14	26	16	14	25
% No quals.	1983	51	31	49	58	38	57	75	91	47	38	61	16	43	50	33
	2000	19	11	19	21	17	28	52	55	15	15	26	5	8	8	10
% in London	1979	10	46	34	60	64	46	17	52	38	57	34	28	29	28	28
	2000	9	45	42	61	73	46	23	62	49	67	34	37	31	56	33
% marry same	1979	99	33	91	82	81	90	94	98	80	57	99	97	98	99	98
	2000	99	58	89	66	74	89	93	98	72	66	98	96	96	95	97

Notes: All figures population weighted. Married includes cohabitees and is conditional on being married

In the first three columns of the table, we report figures for British-born whites, British-born ethnic minorities (who are almost certainly second generation immigrants of ethnic minority origin), and individuals who are foreign-born. The next columns split the foreign-born into groups of various origins.

The immigration flows outlined in Figures 1.1 and 1.2 have shaped the demographic patterns observed in Table 1.1. In 1979, around 7% of the working-age population were born outside Britain. Ethnic minority natives account only for 0.5% of the population. The largest immigrant community in Britain in 1979 were those of Irish origin, some 1.4% of the population, or around 0.6 million individuals. Next came members of the Indian and European Union communities, each accounting for around 1% of the working age population.

By the year 2000, the total immigrant stock had risen to around 9% of the working age population. Ethnic minority natives accounted for 2.5% of the population and immigrants represented 49% of the immigrant population. The largest immigrant group were now individuals born in the European Union (outside Ireland), at around 1.4 % of the population, followed by immigrants from India. The shares of immigrants from sub-Saharan Africa, Pakistan, Bangladesh and the Old Commonwealth countries of Australia, New Zealand and the USA, all grew over this period, whilst the shares of immigrants from the Caribbean and Ireland fell[6].

The median age of the immigrant population is very similar to that of UK-born whites in both 1979 and 2000. The median age of UK-born ethnic minorities is much lower, which is explained by the immigration patterns of the foreign-born ethnic minority individuals. This population ages considerably between 1979 and 2000.

Information on the year of arrival is not available for the 1979 LFS. The first year for which this information was recorded is 1983. We report in the table the median years since migration for the total immigrant population, and distinguish between different origin groups, for the years 1983 and 2000. The average immigrant had already spent around 18 years in Britain in 1983 and around 19 years by 2000. This average conceals some large differences across the various groups, reflecting the history and geographic

[6]Notice that the change in the composition of the immigrant population of working age was not only due to immigration and demographic developments, but may also have been due to return migration. Hatton and Wheatley Price [2002] show that the trends in total emigration from and immigration to the UK often cross each other during the 1964-1994 period.

pattern of immigration into Britain over the past 50 years. Members of the West Indian community have been in the UK the longest, around 34 years on average in 2000. They are followed by the Irish and Indian communities, with 32 and 25 years of residence in 2000 respectively. More recent ethnic minority immigrants are from Pakistan and Bangladesh, with 20 and 16 median years since migration, respectively. The most recent immigrants, on average, now come from European countries currently outside the European Union.

We report in the next panel the age at which immigrants enter the UK. The numbers show that the median age of arrival of the working-age population residing in Britain in the year 2000 was around 20. Again, there is large variation across the various immigrant groups that we identify. In particular, the median age of arrival for immigrants from most ethnic groups is below 20. Looking at the distribution of age at entry, we find that 90% of immigrants resident in the year 2000 came to Britain before the age of 30. Around one third of all immigrants arrived as children, before the age of 15. Again there is considerable heterogeneity across the different groups. Nearly half of all Caribbean and Bangladeshi immigrants arrived as children, compared with less than a fifth of immigrants from Africa. With the exception of the whites born elsewhere group, the fraction of child immigrants has risen over time, presumably, in part, because the families of original immigrants become eligible for settlement.

In the second column of the table we report respective numbers of ethnic minority individuals who are born in the UK. While only 0.5 percent of the working-age population in the UK were non-white natives in 1979, this number has risen to 2.4 percent 20 years later. This is about half of all ethnic minority individuals in Britain (some 800,000 people).

Table 1.1 also outlines the differential levels of educational attainment between immigrants, white UK-born individuals, and ethnic minority natives, and across immigrant groups. It is apparent that the immigrant community as a whole is generally more educated than native whites. Among males, in 1983, only 10 percent of UK-born whites had graduated, while this is the case for 15 percent of the immigrant population. By 2000, the percentage of graduates in the UK-born white population had increased to 16 percent, and to 21 percent in the immigrant population. At the lower end of the education distribution, the relative numbers are quite similar: 46 and 47 percent of the white natives, and the foreign-born population had no educational qualification in

1983; these numbers have dramatically decreased for both populations, to 14 and 16 percent respectively. This indicates a significant improvement in the lower end of the skill distribution of immigrants to the UK. When we look at educational attainment for male immigrants for the various origin groups, we see that there have been significant improvements for nearly all groups at the lower end of the skill distribution.

On the other hand, there are stark differences in the percentages of graduates, according to country of birth, and in particular among ethnic minorities. While, for instance, only 4 (5) percent of individuals from the West Indies had graduated in 1983 (2000), 16 (23) percent of immigrants from India had a degree. The black African, Indian and Chinese groups contain many more graduates than UK-born whites and a correspondingly lower share of those with no qualifications. In 2000, around one third of the African and Chinese immigrant population living in Britain had a degree, compared to sixteen per cent of UK-born whites. In contrast, the West Indian, Pakistani, and particularly, the Bangladeshi communities contained fewer graduates than the national average and many more individuals with no formal qualifications.

In 2000, around 40% of all Bangladeshis had no formal qualifications, compared to 14 per cent of UK-born whites and 9 per cent of those in the black African group. Whilst the West Indian immigrant community does relatively badly in terms of educational attainment, it is the only ethnic group, including UK-born whites, where women do better than men. The proportion of female West Indian women with a degree is close to the national average and the share of West Indian women with no qualifications is below the national average. In contrast, the share of women in the Bangladeshi and Pakistani communities with no qualifications is more than twice the national average. For females, the differences across years and origin groups are similar, but the levels are generally lower.

Comparing white and non-white natives, in 2000 there are less white natives with a degree (16% of males and 12% of females) than minority natives (18% of males and 16% of females). Furthermore, more white natives do not have any formal qualification than minority natives, and this difference is particularly significant among females (19% for whites and 11% of non-whites).

A very interesting feature revealed by Table 1.1 is the high concentration of immigrants, as well as the UK-born ethnic minorities, in the capital. In 2000, London contained around 9% of the total population, but more than 40% of all immigrants, and

45% of UK-born ethnic minorities.

Comparing 2000 to 1979, the geographic concentration in the capital appears to have increased. Only the Pakistani, Irish and European groups are less concentrated in the capital, though members of these groups are still more than twice as likely to live in the capital compared to UK-born whites.

The bottom two rows of Table 1.1 highlight the proportion of each group who have married within the same ethnic/immigrant group. Around 10% of immigrants have married outside their ethnic group. It is apparent that marriage across ethnic lines is much more common amongst ethnic minority natives, nearly half of whom, if in a relationship, are married or cohabiting with someone from a different ethnic group. Amongst immigrants, marriage or cohabitation with someone from outside the immigrant/ethnic group is quite common amongst members of the West Indian and Chinese communities and less so in the Pakistani and Bangladeshi communities.

To summarise, the share of white and non-white immigrants in the UK population has grown steadily from the 1950's onwards. Ethnic minority immigrants, ethnic minority natives and white natives do not share the same individual and demographic characteristics. Furthermore, the same characteristics vary across different ethnic groups. This suggests that analyses should be performed taking into account of such differences.

1.2 Description of the Data Sources

This thesis focuses on a fraction of the immigrant population: ethnic minorities. There are various reasons for this choice, partly based on the quality of the available data. In the UK, immigrants and ethnic minorities represent a small part of the overall population. As a consequence, sample size of immigrants and ethnic minorities in standard surveys is often not sufficient to allow meaningful econometric research to be performed. However, ethnic minorities have recently been the subject of extensive sociological and epidemiological research[7], which have led to the design of surveys covering issues specific to ethnic minorities and aiming to cast light upon their social and economic conditions. To provide a suitable number of observations, ethnic minorities are oversampled in these surveys.

The Family and Working Lives Survey (FWLS) was been collected between 1994

[7]For a comprehensive survey, see Modood and Berthoud [1997].

and 1995 and was conducted by the Department for Education and Employment. It contains retrospective data on adults aged between 16 and 69. It consists of a main sample of 9,139 individuals plus a"boost" sample of 2098 individuals belonging to four racial minority groups: Black Caribbeans, Indians, Pakistanis and Bangladeshis. The data provides information on earnings, education, nationality, and ethnic group. An important feature of the FWLS is that it contains information on English language skills. Foreign born respondents assess their level of speaking, reading and writing ability.

The Fourth National Survey on Ethnic Minorities (FNSEM) is a cross-sectional survey, which was carried out between 1993 and 1994 by the Policy Studies Institute. Individuals included are aged 16 or more. It consists of a main sample of 5196 ethnic minority respondents and an independent comparison sample of 2867 white individuals. The survey provides information on social and economic conditions of the respondents. Similarly to the FWLS, it provides information on the language fluency of foreign born respondents, as perceived by the interviewer, plus data on which language the interview was conducted. In addition, in the FNSEM, ethnic characteristics and inter-ethnic relations are investigated in detail. The minority sample contains data on quality of inter-ethnic relations (as perceived by the respondents), racial hostility, and consequent precautionary measures. In contrast, the white respondents' sample contains information on self-assessed degree of prejudice towards ethnic minorities. Finally, it provides data on the health status of respondents.

The Health Survey for England (HSE) comprises a series of annual surveys, commissioned by the Department of Health, of which the 1999 survey is the ninth. All nine surveys have covered the adult population aged 16 and over living in private households in England. The aim of the HSE is to provide information on the health of the population in England. The 1999 survey was the first to increase the representation of minority ethnic adults and children from Black Caribbean, Indian, Pakistani, Bangladeshi, Chinese and Irish communities. Besides information on health and social and economic conditions, it provides information on the cultural backgrounds of the respondents.[8]

The Labour Force Survey (LFS) is a continuous household survey, conducted by the Office for National Statistics (ONS). The LFS has been running since Spring 1992 in its

[8]Note that Black Africans are not included in any of these surveys. The reason for the exclusion advanced by the authors of the surveys is based on the fact that, differently from the other ethnic groups, Black Africans are not a homogeneous group.

present form although a LFS has been carried out in the UK since 1973. Between 1973 and 1983 a biennial survey was carried out in the Spring. In 1984 the survey became annual. In Spring 1992, for the first time, the data were made available quarterly, with a quarterly sample size approximately equivalent to that of the previous annual data, thus becoming the Quarterly Labour Force Survey. Each quarter interviews are achieved at about 59,000 addresses with about 138,000 respondents. A core of questions covering household, family structure, basic housing information and demographic details of individuals in the households is included in every survey, together with non-core questions which vary from quarter to quarter.

The Census of Population is a questionnaire survey of the United Kingdom population held every ten years. The aim of the Census is to describe the state of the country, providing a comprehensive spatial coverage. The data used for this thesis are 1971, 1981 and 1991 (these are also the only ones available electronically). They contain information on total population, gender, age, marital status, country of birth, economic activity, employment status and various household characteristics. Additional information can be found in the more detailed 1991 version, like ethnic group, qualifications and weekly hours worked. The information is available only in selected cross-tabulations of aggregate data for geographical areas of the United Kingdom which broadly correspond to administrative areas. This implies a limited use of the data if further disaggregation is required in the analysis.

Both the FWLS and the FNSEM include geographical identifiers at ward level which allow us to combine them with the Population Census data sets. As a consequence, we are able to derive detailed demographic information about the area where the respondent lives[9]. On the other hand, the 1999 HSE and the LFS contain only information at wider spatial level. This limits the scope of the analysis when detailed area information is needed.

1.3 Overview of the Thesis

The welfare of immigrants is affected by numerous factors, one of which is the economic performance in the host country. Immigrants' economic performance is the main focus of most of the economics literature on migration in the UK. While providing further

[9]Chapter 3 provides a comprehensive explanation.

insights into the role of immigrants in the UK labour market, this thesis also aims to analyse their welfare from other perspectives. The analysis benefits from the use of the numerous data sources described above.

The thesis is divided into three main parts. In *part one*, we analyse the economic performance of immigrants and look at how immigrants' proficiency in the language of the host country affects their labour market outcomes. We also investigate what the determinants of language proficiency are, focusing in particular on ethnic density. In *part two* we try to determine whether immigrants are positively selected in terms of health, and whether inequalities in health exist between immigrants and natives. In *part three*, we investigate issues regarding the quality of reception of immigrants in the UK.

Part one is organised in two chapters. In *chapter two*, we explain how language proficiency, as a component of the human capital, contributes to the labour market outcome of immigrants. The problems of measurement error and unobserved heterogeneity related to the language variable are discussed and addressed. Findings show that language proficiency provides a substantial advantage in employment probability and earnings. This advantage appears to be underestimated if measurement error is not taken into account. Results also suggest that the bias deriving from measurement error is larger than the bias deriving from unobserved heterogeneity.

Chapter three investigates the relationship between language fluency and ethnic concentration. The presence of enclaves may prevent immigrants from interacting with the majority community, contributing to social exclusion. Ethnic enclaves may offer economic, social and cultural opportunities to immigrants which they may struggle to find elsewhere especially if they are not fluent in English. In particular, higher levels of ethnic concentration may correspond to lower incentives to learn English. We develop a model of human capital which considers the potential effect of ethnic density and its change over time on the acquisition of language proficiency. The model also contemplates the case that, at arrival, immigrants' settlement choice is affected by the level of fluency. As such, it is a generalisation of models presented by previous literature. Our findings show that ethnic concentration has no significant effect on the language acquisition of male immigrants, but has a negative and significant effect on the language acquisition of female immigrants. For neither sample, we find evidence of sorting based on the level of language proficiency at arrival.

Health is an important measure of the social integration and well-being of individuals in a society. In *part two* of the thesis, health inequalities between immigrants and their UK-born peers are investigated. In *chapter four* we present a simple model where health determines the migration decisions of the individual and the family. The model predicts that immigrants are positively selected if the gain from migration is increasing in health. Health differentials between first generation immigrants, second generation immigrants, and natives are estimated. Findings suggest that immigrants are likely to be positively selected. However, results depend on the health measure considered.

In *part three*, we investigate aspects on the quality of reception of immigrants in the UK. Attitudes of the majority population towards minorities, the quality of inter-ethnic relations, and the deriving degree of social cohesion are key determinants of the social and economic integration process of immigrants in the host country. The last two chapters contribute to the discussion from two perspectives. According to existing theories, natives may expect immigrants to be a threat to the economic, social, and political hegemony of their community. Racial prejudice is then interpreted as a response to the consequent anticipated competition for scarce resources (such as, for example, employment and welfare support), and possibly as a manifestation of individual preferences against immigration.

In this context, the aim in *chapter five* is to evaluate how migration affects the UK labour market. In particular, we estimate the effects of immigration on employment and wages. We evaluate the impact of immigration separately on different skill and demographic groups. Results point towards positive wage effects of immigration, even if not always statistically convincing. Similarly, the effect on employment is negative, but mostly non significant. Theoretical literature suggests that there are realistic routes by which immigration can affect labour market outcomes, but the absence of any long run impact is by no means implausible or inconsistent with theory for the case of an open economy with a large heterogeneous traded goods sector such as the UK.

In *chapter six*, we investigate the determinants of racial harassment. Racial harassment is an important aspect in which inter-ethnic relations manifest themselves. Previous theories are not exclusive, but rather outline different aspects in which harassment manifests itself. As a consequence, empirical evidence does not provide a definite answer as to how ethnic contexts affects inter-racial relations. In the chapter, we focus on the association between ethnic concentration, prejudices towards minorities, and the

probability of ethnic minority individuals experiencing racial hostility. We consider precautionary behaviour of individuals worried of being harassed as a possible determinant of the incidence of harassment. We find that racial harassment is not an extreme form of negative prejudice, as much of the previous literature assumed. In particular, results show that, as ethnic concentration increases, the incidence of harassment and precautionary behaviour decrease. In contrast, racial prejudices seem to be higher in areas at higher ethnic density.

Part I

Language and Economic Performance

Chapter 2

Language Proficiency and Labour Market Performance of Immigrants in the UK

Abstract:

This chapter investigates the determinants of language proficiency, and the effect of language on earnings and employment probabilities of non-white immigrants. We address the problem of endogenous choice of language acquisition and measurement error in language variables. Our results show that language acquisition, employment probabilities, as well as earnings differ widely across non-white immigrants, according to their ethnic origin. Language proficiency has a positive effect on employment probabilities, and lack of English fluency leads to earnings losses.

2.1 Introduction

According to the 2000 Labour Force Survey, immigrants (defined as individuals who are born outside the UK) account for around 9 per cent of the working age population of Britain. Immigrants are heavily concentrated in Metropolitan areas. In 2000, London contained around 9 per cent of the total population of the UK, but more than 40 per cent of all immigrants. The ethnic origin of immigrants in the UK is diverse, with the largest group being born elsewhere in the European Union, followed by immigrants from India, the Old Commonwealth, Pakistan, and Africa (see Introduction to the thesis for more details).

A number of recent studies analyse various aspects of labour market behaviour of ethnic minorities, and compare outcomes with those of the majority population (see e.g. Blackaby et al. [1994], Blackaby et al. [1997], and Clark and Drinkwater [2000]). In much of this literature, however, no attempt is made to distinguish between immigrant and British born minorities. But many important questions are specifically related to first generation immigrants, who constitute a significant fraction of minorities in the UK. As we have shown in the Introduction, about 66 percent of ethnic minorities of working age were born abroad.

There are few papers that investigate the economic assimilation of immigrants. The earliest study is by Chiswick [1980], who uses data from the 1972 General Household Survey (GHS). His main finding is that, while white immigrants have very similar earnings patterns to native-born individuals, earnings of immigrants from ethnic minority groups are about 25 percent lower, other things the same. This gap is not decreasing with time of residence in the UK. In a more recent paper, Bell [1997] uses also data from the GHS, but he pools waves between 1973 to 1992. Like Chiswick, he finds that white immigrants are doing well. While white immigrants have an initial wage advantage, compared to native workers, immigrants from the West Indies and India have an earnings disadvantage, but wage differentials between this group and white natives decrease with the time spent in the UK. Shields and Wheatley Price [1998] use data from the British Labour Force Survey. They emphasise the different assimilation patterns between foreign and native born minority individuals.

It may be in the interest of the host country to support the process of economic assimilation. To achieve this, it is important to understand the factors that determine

the economic performance of minority immigrants. In this chapter, we concentrate on one specific human capital factor, which is important not only for immigrants' economic assimilation, but also for their social integration: proficiency in the host country language. Recent analyses for the US, Canada, Australia, Israel, and Germany show that fluency and literacy in the dominant host country language are important components for explaining immigrants' labour market success (see, e.g., Rivera-Batiz [1990], Chiswick [1991], Dustmann [1994], Chiswick and Miller [1995], Chiswick et al. [1997], and Berman et al. [2000]). Work by Shields and Wheatley Price [2001] indicates that language is also positively related to occupational success of some immigrant groups in the UK.

We analyse the determinants of fluency and literacy in the host language for immigrants belonging to ethnic minority groups, and how it relates to their labour market performance. We first investigate factors influencing the acquisition of language by the immigrant, such as education, age, and years of residence in the host country. We distinguish between education received in the host- and in the home countries.

We then analyse the extent to which language ability influences the labour market outcomes of immigrants. We focus on its effect on employment probabilities, and on the level of earnings. Estimates of language coefficients in straightforward regressions are bedeviled by two problems. First, as pointed out by Borjas [1994], the choice to acquire proficiency in a foreign language may be endogenous. Second, as stressed by Dustmann and van Soest [2001], language measures usually reported in survey data may suffer substantially from measurement error. The bias induced by these two problems points in opposite directions. We attempt to address both problems in this chapter, and propose estimators which may help to reduce, or eliminate the bias. We combine an IV estimator that eliminates the bias due to measurement error with a matching estimator that addresses the problem of endogenous choice of language acquisition. Our results suggest that measurement error leads to a downward bias in the estimates of language on employment probabilities and earnings, and that the true effects are larger than what OLS estimates reveal. These results are in line with findings for other countries (see Dustmann and van Soest [2001]).

Our best estimates suggest that fluency in English increases employment probabilities by about 22 percentage points. This estimate is 5 percentage points higher than the OLS estimate. Furthermore, OLS estimates show that proficiency in English is associated with 18-20 percent higher earnings. Again, our estimator that takes account of

both measurement error and endogenous selection indicates that effects are larger, but estimates are not significant, probably due to the small sample sizes.

We base our analysis on data from two UK surveys on ethnic minorities: the Fourth National Survey on Ethnic Minorities (FNSEM), which has been collected between 1993 and 1994, and the Family and Working Lives Survey (FWLS), which has been collected between 1994 and 1995. Both data sets consist of two subsamples. The FWLS contains a main sample of the entire UK population, and a "boost" sample of individuals belonging to ethnic minorities. The FNSEM contains a main sample of respondents belonging to ethnic minorities, and a reference sample of individuals belonging to the white majority population. Both surveys include questions on social and economic conditions of the interviewees, and measures on language proficiency. Information in the two data sets is complementary. For instance, while the FNSEM only reports spoken language proficiency, the FWLS contains also information about reading and writing skills. Also, the FNSEM distinguishes between education acquired in home- and host economy, information which is not available for the FWLS. Using two data sets allows us to conduct comparable analyses to check the robustness of the results obtained.

The data sources we use for this analysis are to our knowledge the only data sets for the UK that contain information about immigrants' language proficiency, as well as information on employment status and earnings. They are restricted to ethnic minority immigrants, and our results do therefore not necessarily generalise to the overall population of immigrants in the UK. Furthermore, results have to be evaluated subject to our ability to address the endogenous choice of language acquisition with the information available in the data, and the relatively small sample sizes, in particular for the earnings analysis. Despite these limitations, our analysis provides interesting insight into the relationship between language, and economic outcomes for a large group of the UK's immigrant population.

The structure of the chapter is as follows. Section 2.2 develops the estimation equations. Section 2.3 describes the data sets, and gives some descriptive statistics. Section 2.4 investigates language determinants. Section 2.5 analyses how language proficiency affects employment probabilities, and earnings. Section 2.6 summarises the results and concludes.

2.2 Language and Labour Market Outcomes

The literature on migrants' earnings assimilation distinguishes between human capital which is specific to the host country, human capital which is specific to the home country, and human capital which is equally productive in both countries. Typically, immigrants enter the host country with skills which are only of limited use in the host economy, which results in an initial earnings disadvantage (see Chiswick [1978]). After immigration, migrants transfer home country specific human capital into general or host country specific human capital, and acquire additional skills which are specific to the host country economy. The intensity of this process determines the speed of economic assimilation.

Language capital is an important component of host country human capital. It is also very specific to the host economy, since it is usually not transferable to the migrant's home economy. Standard human capital models may serve as a basis to formulate empirical specifications explaining the determinants of language capital (see Dustmann [1999]). In such models, human capital is produced by investing time and other inputs. The cost of production equals forgone earnings, plus the cost of other input goods. A simple equilibrium condition states that investment into human capital production is set such that the marginal cost equals the marginal benefit from the discounted future enhanced earnings potential. The production potential may differ across individuals according to their ability to acquire knowledge, and it may depend on the stock of human capital acquired in the past. The benefit of any acquisition of host country specific human capital depends, in addition, on the length of the period over which it is productively put into use.

Investment into language capital should therefore depend on its potential future benefits, on the cost of acquisition, and on the individual's efficiency in producing it. Chiswick and Miller [1995] provide an extensive discussion on the variables which represent these factors. Variables which measure the immigrant's efficiency in acquiring language capital are the level of education upon immigration, and the age at immigration (since the learning potential may deteriorate over the life cycle). The cost of acquiring the host country language depends on the distance of the migrant's mother tongue to the dominant majority language, which may be captured by country of origin dummies. Clearly, this last variable picks up a variety of other factors which affect language proficiency, like different degrees of immigrant selection across countries (see

Borjas [1985] and Borjas [1987]). Assuming that all migrations are permanent, the time period over which language capital is productive depends on the migrant's age at entry. Accordingly, those who migrate at younger age should have a higher incentive to acquire language capital. Its acquisition may, in addition, depend on the extent to which individuals are exposed to the language of the majority population. As noted by Chiswick and Miller [1995], a variable which measures exposure is the time of residence abroad.

2.2.1 Language, Earnings, and Employment Probabilities

When analysing the effect of language on labour market outcomes, two problems may occur. First, the choice of learning the host country language may be endogenous, and related to variables which affect outcomes. This may lead to an upward bias of estimated language effects on economic outcomes. Second, unsystematic measurement error may lead to a downward bias of the effect of language on earnings. Numbers presented in Dustmann and van Soest [2001] on repeated language information for the same individual suggest that measurement error is substantial in self-reported language measures. In fact, in their data, more than half of the within individual variation in language responses is due to measurement error. Their results suggest that the downward bias induced by measurement error overcompensates the upward bias induced by unobserved heterogeneity.

To give a causal interpretation to the language coefficient, we need to deal with both sources of bias. We first discuss the problem of the endogenous choice of language acquisition. Assume for the moment that the language variable is measured without error. Then the problem is that those individuals who have chosen to obtain proficiency in the English language may differ from those individuals who have chosen not to do so. If these differences affect outcomes (in our case, employment or earnings) other than through language, a comparison in outcomes of the two groups does not produce an unbiased estimate of the causal effect of language proficiency.

We define the parameter we would like to obtain as the difference in outcomes for an individual of being proficient and non-proficient, after having made the choice of acquiring language proficiency.[1] Denoting these two potential outcomes by y_i^1 and y_i^0,

[1] An alternative parameter of interest is the difference in outcomes of being proficient and non-proficient in the English language for individuals who have chosen not to learn the language. See Dearden et al. [2002] for a discussion of the two parameters.

and proficiency in English by $l_i = 1$, where i is an index for individuals, this parameter is given by

$$E(y_i^1 - y_i^0 | l_i = 1).$$

This mean effect of language proficiency on outcomes for those who have decided to learn the foreign language is often referred to as the effect of "treatment on the treated" (see Heckman et al. [1998]). The problem we face in retrieving this parameter is that we do not observe individuals who decided to learn the host country language, but then refrained from doing so. In other words, the counterfactual $\mathrm{E}(y_i^0 | l_i = 1)$ is not observed. What we observe instead is $\mathrm{E}(y_i^0 | l_i = 0)$. If individuals who have, and who have not chosen to learn the language differ in characteristics related to wages, $\mathrm{E}(y_i^1 - y_i^0 | l_i = 1)$ $\neq$ $\mathrm{E}(y_i^1 | l_i = 1)$ - $\mathrm{E}(y_i^0 | l_i = 0)$.

To estimate the mean effect of language on outcomes for those who have chosen to learn the language, we use a matching type approach. Suppose that we observe a vector of conditioning variables x_i, sufficient to control for the endogenous choice of learning the English language. Then the expectation of the outcome with no language proficiency is conditionally independent of the decision to learn the language, i.e. $\mathrm{E}(y_i^0 | x_i, l_i = 0)$ $= \mathrm{E}(y_i^0 | x_i, l_i = 1)$. Under this conditional independence assumption, we can use the outcome of those who are not proficient in the English language to estimate the counterfactual outcome of those who are proficient, were they not proficient. The parameter of interest is then given by $\mathrm{E}(y_i^1 | l_i = 1, x_i)$ - $\mathrm{E}(y_i^0 | l_i = 0, x_i)$, which can be obtained from the data.

If x_i is multi-dimensional, this amounts to comparing individuals with the same cell distribution in terms of the variables in x_i. This requires large data sets, and discretisation of continuous variables in x. Rosenbaum and Rubin [1983] show that, if the conditional independence assumption is fulfilled, then it suffices to match on the propensity score $\mathrm{P}(l_i = 1 | x_i) = P(x_i)$ (the probability of being proficient in English, conditional on characteristics x_i), which reduces the matching index to one dimension.

It is important to ensure that individuals are only matched for those x_i commonly observed for proficient, and non-proficient individuals (i.e. who have a common support in x). If, for instance, there are values of x_i where only proficient individuals are observed - in other words, $P(x_i) = 1$ for some values of x_i - the conditional expectation of $\mathrm{E}(y_i^0 | l_i = 0, x_i)$ is not defined. Heckman et al. [1997] show that, if the common support condition is not fulfilled, then the matching approach may lead to seriously biased estimates.

We use a propensity score estimator, which ensures that the support conditions are fulfilled. We estimate the propensity score for being proficient in the English language using a simple logit model. We estimate the conditional expectation of the counterfactual using a Gaussian kernel, and match observations by nearest neighbour matching, based on the propensity score. We disregard individuals for which the absolute difference in the propensity score to the nearest neighbour in the control sample is not small enough. We then compute the mean difference between the treatment group and the constructed counterfactual. We estimate $\gamma^M = \int \mathrm{E}(y_i^1 - \mathrm{E}(y_i^0|P(x_i), l_i = 0)|l_i = 1)dF(P(x))$, where $\mathrm{E}(y_i^0|P(x_i), l_i = 0)$ is estimated using a Gaussian Kernel on those who are not proficient in the English language. Finally, we compute standard errors by bootstrap, using 500 repetitions.

A second problem we face is that there is measurement error in the self-reported language indicator. To address the measurement error problem, we use a two stage approach, which is based on the following idea. Suppose we had an instrument I_i, which has the properties that (i) it is independent of the outcome, conditional on x_i and l_i and (ii) it explains variation in l_i (in other words, $\mathrm{E}(l_i|I_i = r)$ is a non-trivial function of r, where r is in the support of I). These conditions correspond to the rank and order conditions for instrumental variable estimation. Let the instrument I_i be binary (in our case, another measure of language). Then an estimator which corrects for individual heterogeneity (using the matching approach) and measurement error (using an IV argument) is given by

$$\gamma^{MI} = \frac{\mathrm{E}(y_i|I_i = 1, x_i) - E(y_i|I_i = 0, x_i)}{\mathrm{Prob}(\tilde{l}_i = 1|I_i = 1, x_i) - \mathrm{Prob}(\tilde{l}_i = 1|I_i = 0, x_i)}, \tag{2.1}$$

where $\tilde{l}_i$ is the measured binary language variable. To estimate this parameter, we proceed in two stages. In the first stage, we compute the numerator of (2.1) by propensity score matching, using the binary instrument I_i (which is the interview language) instead of the language variable. In a second step, we re-scale this parameter. We compute the denominator as the difference in the predicted probabilities of our language measure (using a linear probability model) for the two outcomes of the instrument.[2] We then

[2]The intuition is as follows. The numerator is the change in the outcome variable if the instrument switches from zero to one; the denominator is the change in the probability of being proficient if the instrument switches from zero to one. It is easy to show that the expression in the denominator is equal to the change in the probability of being proficient in the true language measure if the instrument

compute the ratio of the two to obtain an estimate of the effect of language on outcomes, which takes account of both endogenous choice and measurement error. To compute the standard errors, we use bootstrapping.

The matching approach is based on the idea that the observable characteristics are sufficient to explain any relationship the choice of learning the language has on the outcome if non-proficient in English. In both data sets, we observe individual specific characteristics (like education, age, origin) and minority concentration in the area. Education should be correlated with otherwise unobserved determinants of the choice to acquiring language proficiency, like innate ability. In the two data sets, some information about family and household characteristics is available. For the FNSEM, we include marital status, number of children, and partner characteristics. In the FWLS, we only observe marital status and number of children, but we have information on some self-assessed abilities, like mental arithmetic, and finding an address on a map.

As instrument to address the measurement error problem, we use information as to whether the interview was held wholly in English, partly in English, or wholly in the individual's mother tongue. The survey has a screening stage during which the majority of the participants to the survey are contacted (by phone) by interviewers [3]. During this screening stage, interviewers attempt a preliminary fluency assessment. In case of poor fluency, respondents are assigned an interviewer from the same ethnic (and language) minority. The ethnic minority interviewers do not participate to the screening stage. Therefore, some ethnic minority respondents are not matched to same ethnicity interviewers and their interview must be held in English. This assignment is independent of the ethnic minority interviewer (i.e., the choice of language in the interview was made by someone different from the final interviewer)[4].

For those respondents who are matched with same ethnicity interviewers, if the interview was not conducted in English, the interviewer is still asked to "attempt a conversation to assess ability". This provides some "independence" in the assessment of the two variables, which supports the validity of the instrument. However, it is not

switches from zero to one, as long as the instrument is not correlated with the measurement error. The ratio of the two is then the change in the outcome variable if the true language variable switches from zero to one. See Heckman [1997] for a discussion of similar estimators.

[3]The initial screening interview took place among those interviewees living in all areas with a minority density above 0.5%, which represent 97% of the sample.

[4]See Smith and Prior [1996] for a complete description of the survey procedures.

possible to completely rule out that the instrument and the measurement error are not correlated[5]. In this case, the IV estimation will only reduce, but not eliminate, the bias. In particular, the estimates we obtain are a lower bound of the "true" coefficient.

2.3 The Data

The Family and Working Lives Survey (FWLS) has been collected in 1994 and 1995. It is a retrospective survey on adults aged between 16 and 69, including 9000 respondents and their partners. It contains a "boost" sample of about 2000 individuals belonging to four racial minority groups: Black Caribbeans, Indians, Pakistanis and Bangladeshis. The data provides information on earnings, education, nationality, language skills and other background characteristics. Of the 2388 people forming the minority sample in the main and "boost" sample, 68% (1639) are foreign born.

The Fourth National Survey on Ethnic Minorities (FNSEM) is also a cross- sectional survey, which has been carried out between 1993 and 1994. Individuals included are aged 16 or more, and of Caribbean, Indian, Pakistani, Bangladeshi, or Chinese origin. There are 5196 observations in the minority sample, and 2867 observations in the independent comparison sample of white individuals. Similarly to the FWLS, more than 77% (4019) of the individuals in the ethnic minority sample are foreign born.

The FWLS identifies the ward where the individual lives.[6] It is therefore possible to match this data set with the 1991 Population Census to construct a variable on the ethnic concentration on ward level.[7] The FNSEM does not contain geographical identifiers; therefore, matching with the Census data is not possible. However, it contains grouped information on ethnic concentration at ward level, obtained by the authors of the survey from the 1991 Census.

Both data sets provide information on earnings. The FWLS reports weekly gross (before tax) earnings, while the FNSEM reports grouped gross weekly earnings. Both data sets report the main activity of the individual (e.g. full-time or part-time paid

[5]This may happen, for instance, if the answer to the question of how much of the interview was run in English affects the evaluation on the respondent's language ability.

[6]In the UK, a ward is the smallest geographical area identified in the Population Census. According to the 1991 census, the mean population within a ward is 5459 individuals, and the median is 4518.

[7]We define ethnic concentration as the ratio of the number of individuals belonging to ethnic minorities over the total population living in the ward. See footnote Table 2.3 for details.

work, full-time education, unemployed, etc.).

Table 2.1: Census 1991 - Ethnic Immigrant Composition in the UK

	Immigrants Perc. on UK Pop.	Ethnic composition	Ethnic composition without Africans
Caribbean	0.56	18.19	23.41
Indian	0.84	27.57	35.49
African	0.68	22.31	-
Bangladeshi	0.22	7.09	9.13
Pakistani	0.47	15.46	18.89
South East Asian	0.29	9.37	12.06
Total	3.06	100	100

Note: Ethnic concentration of own group. Immigrant concentration at arrival. Change in immigrant concentration between time at arrival and time of the survey.

The sample design of the two surveys differs substantially. The ethnic minority sample of the FWLS was selected by screening addresses in areas where the ethnic minority population, according to the 1991 Census, was more than 3% of the local population. The selection in the FNSEM was more complex, considering wards with any percentage of ethnic minorities on the population and oversampling Bangladeshis to obtain a sufficient sample size. For more details, see Appendix 1 in Modood and Berthoud [1997], and Smith and Prior [1996].

Table 2.1 shows the percentage of immigrants belonging to ethnic minorities with respect to the overall population in the UK (column 1), and the ethnic composition within the group of ethnic immigrants. Numbers are based on the 1991 Census. Table 2.2 gives the ethnic composition of the two surveys. Both surveys do not include Black African immigrants, and the FWLS does not include the Chinese minority. In the last column of Table 2.1, we report respective numbers in the census, excluding Africans. Comparing the two tables, it appears that both surveys tend to oversample the South Asian groups (Indians, Pakistanis and Bangladeshis). Also, the two surveys differ in the ethnic composition of the respondents: Bangladeshis amount to 31% in the FWLS and 14% in the FNSEM, Indians to 19% in the FWLS and 24% in the FNSEM and African Asians to 8% in the FWLS and 17% in the FNSEM.

Table 2.2: Ethnic Immigrant Composition in Survey data

	FWLS		FNSEM		
	No.	Perc.	No.	Perc.	Perc.
Black Caribbean	265	16.17	698	18.20	17.37
Indian	314	19.16	971	25.32	24.17
Afro-Asian	123	7.50	656	17.11	16.32
Bangladeshi	512	31.24	550	14.34	13.68
Pakistani	425	25.93	960	25.05	23.89
Chinese	-	-	184	-	4.58
Total	1639	100	4019	100	100

Both surveys contain information on language. In the FWLS, language ability is self-assessed. The individual is first asked whether s/he speaks English as mother tongue. If not, the individual is asked to self-assess proficiency in speaking, reading, and writing English on a 5 point scale. The FNSEM contains two variables which are related to language proficiency: first, the interviewer's evaluation of the individual's spoken language ability, on a 4 point scale. Second, information about what fraction of the interview was held in English. In all areas with a minority density above 0.5% (which includes 97% of the sample individuals), there was an initial screening interview with the interviewee. In the case of poor fluency, the interviewers were chosen to be fluent in the language of the respondents. During the interview, interviewers decided about the extent to which English could be used in the interview, and we have information as to whether the interview was held wholly in English, partly in English, or wholly in the individual's mother tongue.

In Table 2.9 we display the responses to self-assessed (FWLS) or interviewer-assessed (FNSEM) language questions for the two data sets, broken down according to ethnic origin. The general pattern is similar for the two data sets.

For the empirical analysis, we re-define the language indicators in the two surveys as dichotomous variables. For the FWLS, this variable assumes the value 1 if the individual reports language fluency or literacy as "quite well" or "very well", or reports English as a first language. For the FNSEM, it is equal to 1 if individuals fall in the categories "fairly" or "fluently". We use the information on the interview language in the FNSEM as an instrument for measurement error. Our instrument is equal to one if the interview was done in English only.

Table 2.3 explains the variables used for the analysis, and presents summary statistics. The mean values on language indicate that the percentage of individuals who speak the English language at least fairly (or quite well) is very similar in the two samples. Percentages for reading and writing in English (available in the FWLS) are slightly lower.

About 51% (FWLS) and 56% (FNSEM) of the sample populations are in the labour force. Of those, 70% (FWLS) and 75% (FNSEM) are employed. These numbers are remarkably similar for the two data sets.

The mean value of weekly wages in the FWLS is £239.17, considering both part and full-time workers. Mean weekly wages are reported in the FNSEM as a grouped variable. The mean weekly gross wage is £240, which is similar to the mean wage in the FWLS.[8]

The average education level is slightly higher in the FNSEM than in the FWLS, with 12.7% graduates in the former sample, and only 7.2% in the latter sample. Furthermore, there is a slightly higher percentage of individuals with no qualification in the FWLS (56.8%) than in the FNSEM (53.3%).[9]

The average ethnic minority concentration at ward level amounts, in both samples, to more than 16% (the average ward concentration in the FNSEM is obtained by taking the average of the mid-point values of the grouped variable, since the information is available only in intervals). The considerable difference in the sample designs is reflected only by the larger standard deviation indicated in the FNSEM.

In Table 2.10, we break down means of the age at immigration, year of immigration, and age for the various ethnic groups. In the FWLS, individuals are on average four years younger than in the FNSEM, and have immigrated at a younger age. The immigration patterns for the various ethnic groups are similar in both data sets, and correspond to the migration patterns indicated in the Introduction and by Bell [1997] and Hatton and Wheatley Price [2002]: Black Caribbeans arrivals are concentrated in the late 1950's and early 1960's, whereas Indians, African Asians and Pakistanis arrived mainly during the 1970's, and Bangladeshis towards the end of the 1970's. Consistent with their shorter

[8]Information on earnings is grouped in the FNSEM. To obtain this number, we estimate a grouped regression model on a constant, and compute the mean of the prediction (see Stewart [1983]).

[9]We construct the education variables following a classification by Dearden [1999]. The variable Degree defines University degree or post-graduate diploma; the variable Alev stands for A-Levels or higher vocational degree; the variable OlevCSE includes O-levels, middle or lower vocational degrees, and miscellaneous qualifications.

stay, Bangladeshis are the youngest group, whereas Black Caribbeans are the oldest on average.

Table 2.3: Variable Description and Sample Characteristics

Variable	FWLS		FNSEM		Description
	Mean	S.D.	Mean	S.D.	
Speak	0.709	0.454	0.691	0.462	Dummy=1 if spoken English is good or very good
Read	0.671	0.469	-	-	Dummy=1 if read English is good or very good
Write	0.641	0.479	-	-	Dummy=1 if written English is good or very good
LabFo	0.511	0.500	0.559	0.469	Dummy=1 if in Labour Force
empl	0.703	0.457	0.749	0.433	Dummy=1 if employed (conditional on LabFo=1)
Wgearn	239.175	432.809	240.049	-	Weekly gross earnings
Male	0.468	0.499	0.505	0.500	Dummy=1 if male
Age	38.347	13.588	42.604	14.407	Age
YSM	20.404	10.313	21.367	10.001	Years of residence in the UK
Married	0.726	0.446	0.776	0.417	Dummy=1 if married
nchild	1.937	1.793	1.654	1.761	Number of children in household
Degree*	0.072	0.258	0.127	0.333	Dummy=1 if university degree
Alev*	0.129	0.335	0.109	0.312	Dummy=1 if A Levels or equivalent
OlevCSE*	0.231	0.422	0.230	0.421	Dummy=1 if O Levels or equivalent
Noqual	0.568	0.495	0.533	0.499	Dummy=1 if no qualification
Ethcon	0.168	0.153	0.166	0.189	Ward ethnic minority concentration**
Caribbean	0.1620	0.369	0.178	0.383	Dummy=1 if Black Caribbean
Indian	0.1860	0.389	0.245	0.429	Dummy=1 if Indian
Afro-asian	0.0838	0.277	0.169	0.375	Dummy=1 if African Asian
Pakistani	0.255	0.436	0.218	0.413	Dummy=1 if Pakistani
Chinese	-	-	0.048	0.214	Dummy=1 if Chinese
Bangladeshi	0.318	0.466	0.142	0.349	Dummy=1 if Bangladeshi

* Definitions follow Dearden (1999). ** Defined as the ratio of own ethnic minority individuals over the total population.

2.4 Language Determinants

After eliminating all the observations with missing values in the variables of interest, we are left with 1589 observations in the FWLS sample, and 3732 observations in the FNSEM sample.

Table 2.4 reports coefficient estimates and robust standard errors from linear probability models, where the indicator variable equals one if the individual is proficient in the respective language component.[10] Comparing results on spoken language for the two data sets shows that the signs of regressors are equal for both samples in most cases, and the sizes of the coefficients are likewise similar (although the coding of the fluency variables differs slightly). Males have a significantly higher probability to be fluent in the majority language. The effect of age (which corresponds to the effect of age at entry, since we condition on years of residence) is negative and strongly significant. Years of residence has the expected positive effect, which decreases with time in the host country. All these results are consistent with findings for other countries. For the FWLS, the effect of these variables is similar for all three components of language capital.

The effect of the education variables is quite strong for fluency (the comparison group are individuals who report to have no qualification). For instance, for the FWLS (FNSEM) individuals with O-levels or equivalent have a 29 (22) percentage points higher probability of being fluent in English.

Speaking fluency may largely be acquired by exposure to the host country language, while writing and reading in a foreign language is a skill which is more difficult to obtain. Acquisition requires a more systematic way of learning, and the general level of schooling obtained may enhance the efficiency of acquiring this component of language capital. This is reflected by our results, which indicate that educational background variables have larger coefficients for reading and writing skills.

Education may be partly obtained in the host country. Since those who wish to enter the educational system in the UK are likely to have acquired some language skills, this leads to a classical simultaneity bias.

The FNSEM allows us to distinguish between education obtained in the UK and abroad. We have re-estimated the language equation, distinguishing between education obtained overseas, and in the UK. Results are reported in the last column of Table 2.4. We denote by F educational achievements obtained abroad, and by E educational achievements obtained in the UK.[11] The effect of overseas qualifications on language fluency is very similar to the effect of education obtained in the UK.

[10]We have also estimated probit models. Marginal effects, evaluated at the sample means, are almost identical to the coefficients we report in the tables.

[11]The variable "Edegree" predicts outcomes perfectly. Estimations are performed on the sample of non-degree holders.

Table 2.4: Language determinants, Linear Probability Models

	FWLS						FNSEM			
	Speaking		Reading		Writing		Speaking			
							All Qualifications		UK/non UK Q	
Variable	Coeff	S.E.	Coeff	S.E.	Coeff	S.E.	Coeff	S.E.	Coeff	S.E.
Const	0.616**	0.083	0.639**	0.084	0.640**	0.085	0.778**	0.053	0.872**	0.055
male	0.105**	0.019	0.109**	0.019	0.082**	0.019	0.144**	0.012	0.152**	0.012
age	-0.013**	0.004	-0.014**	0.004	-0.018**	0.004	-0.024**	0.002	-0.030**	0.002
age^2/100	0.010*	0.005	0.010*	0.005	0.016**	0.005	0.014**	0.002	0.019**	0.002
YSM	0.021**	0.003	0.012**	0.004	0.012**	0.004	0.023**	0.002	0.027**	0.002
YSM^2/100	-0.036**	0.010	-0.014	0.010	-0.018**	0.010	-0.027**	0.005	-0.034**	0.006
degree	0.308**	0.037	0.415**	0.038	0.457**	0.038	0.400**	0.019	–	–
Alevtea	0.303**	0.028	0.362**	0.029	0.421**	0.029	0.275**	0.019	–	–
OlevCSE	0.299**	0.023	0.337**	0.023	0.380**	0.023	0.223**	0.015	–	–
Edegree	–	–	–	–	–	–	–	–	–	–
EAlevtea	–	–	–	–	–	–	–	–	0.190**	0.023
EOlevCSE	–	–	–	–	–	–	–	–	0.182**	0.019
Fdegree	–	–	–	–	–	–	–	–	0.461**	0.023
FAlevtea	–	–	–	–	–	–	–	–	0.234**	0.029
FOlevCSE	–	–	–	–	–	–	–	–	0.195**	0.018
married	-0.047*	0.023	-0.053*	0.024	-0.039	0.024	0.004	0.015	0.006	0.016
nchild	-0.016**	0.006	-0.012*	0.006	-0.018**	0.006	-0.005	0.003	-0.006*	0.003
Indian	0.249**	0.030	0.230**	0.030	0.223**	0.030	0.089**	0.021	0.087**	0.021
Afro-asian	0.241**	0.037	0.236**	0.038	0.215**	0.038	0.232**	0.022	0.258**	0.023
Pakistani	0.137**	0.025	0.075**	0.025	0.074**	0.025	-0.021	0.019	-0.019	0.020
Caribbean	0.373**	0.036	0.396**	0.037	0.435**	0.037	0.454**	0.024	0.482**	0.025
Chinese	–	–	–	–	–	–	0.071*	0.031	0.069*	0.034
ethcon	-0.468**	0.091	-0.316**	0.093	-0.181	0.093	-0.208**	0.031	-0.215**	0.032
No. of Obs.	1589		1589		1589		3732		3552	
Obs. Prob.	0.710		0.646		0.641		0.691		0.675	

Base Category: No educational qualification, Bangladeshi. Ethnic concentration for FNSEM at midpoints. Robust standard errors are reported. *: Significant at 5 percent level. **: Significant at 1 percent level.

The variable "nchild" measures the number of children in the household. Chiswick and Miller [1995] suggest that children may have counteracting effects on language: first, they may act as a translator between the parent and the English speaking community (thus reducing incentives to learn the foreign language). Second, they may enhance

exposure to the majority population by forcing the parent to cope with institutional matters, like school and parents of native friends of children. Our results indicate that children coefficients are negative for both data sets, and for all language components.[12]

There are large differences in the level of language proficiency across different ethnic groups. Results of both data sets indicate that Bangladeshis, the base group, are dominated by nearly all other ethnic groups, except for Pakistanis in the FNSEM.

The variable "ethcon" measures ethnic concentration at ward level. It is strongly associated with language proficiency for both data sets. Results from the FWLS indicate that an increase in the ethnic density by 1 percentage point is associated with a 0.47 percentage point decrease in the probability to be fluent in the dominant language. The negative association with reading and writing skills is slightly smaller. Results from the FNSEM also indicate a negative association, but the size of the coefficient is only half as large as that for the FWLS[13]. These results are in line with findings for the US, Canada and Israel (see Chiswick [1994], and Chiswick and Miller [1995]).

2.5 Language and Economic Outcomes

2.5.1 Employment Probabilities

Language proficiency is likely to be a decisive factor in determining employment probabilities. Language may help to acquire information about optimal job search strategies. Migrants who are not sufficiently proficient in the dominant language may have difficulties to convince prospective employers of their qualifications. Also, many jobs, for instance in the service sector, require communication skills. Likewise, literacy in the dominant language is a crucial prerequisite for many unskilled occupations.

To understand the association between employment probabilities and language, we consider individuals who are in the labour force, and we distinguish between those who are in work, and those who are not employed, but who are actively seeking for a job.[14]

[12]We have also estimated models where we interact number of children with gender. The children variable is positive (though insignificant) for males, but negative (and significant for the FWLS data) for females.

[13]See Chapter 3 for a more detailed investigation on this relationship.

[14]This follows the ILO definition of unemployment. According to the ILO definition, people are considered as unemployed if aged 15 years or older, who are without work, but available to start within

Table 2.5: Employment probabilities, Linear Probability Models

	FWLS						FNSEM			
	1		2		3		4		5	
							All Qualifications		UK/nonUK Q	
Variable	Coeff	S.E.	Coeff	S.E.	Coeff	S.E.	Coeff	S.E.	Coeff	S.E.
Const	-0.052	0.169	-0.082	0.169	-0.087	0.169	0.101	0.116	0.105	0.118
male	-0.128**	0.034	-0.123**	0.034	-0.125**	0.034	-0.080**	0.019	-0.079**	0.019
married	0.175**	0.042	0.176**	0.041	0.178**	0.042	0.167**	0.025	0.168**	0.025
nchild	-0.035**	0.011	-0.034**	0.011	-0.034**	0.011	-0.026**	0.006	-0.026**	0.006
degree	0.047	0.053	0.019	0.055	0.018	0.055	0.107**	0.026	–	–
Alevtea	0.008	0.045	-0.016	0.047	-0.017	0.047	0.121**	0.027	–	–
OlevCSE	-0.064	0.039	-0.084*	0.040	-0.086*	0.040	0.071**	0.022	–	–
Edegree	–	–	–	–	–	–	–	–	0.103**	0.034
EAlevtea	–	–	–	–	–	–	–	–	0.116**	0.030
EOlevCSE	–	–	–	–	–	–	–	–	0.069**	0.025
Fdegree	–	–	–	–	–	–	–	–	0.082**	0.032
FAlevtea	–	–	–	–	–	–	–	–	0.067	0.040
FOlevCSE	–	–	–	–	–	–	–	–	0.052*	0.025
age	0.029**	0.009	0.030**	0.009	0.030**	0.009	0.016**	0.006	0.016**	0.006
$age^2/100$	-0.039**	0.012	-0.040**	0.011	-0.040**	0.011	-0.024**	0.007	-0.024**	0.007
YSM	0.002	0.007	0.004	0.006	0.003	0.007	0.003	0.004	0.003	0.004
$YSM^2/100$	0.001	0.017	-0.000	0.017	-0.001	0.017	-0.004	0.010	-0.003	0.010
Caribbean	0.105	0.059	0.094	0.059	0.089	0.059	0.126**	0.039	0.127**	0.039
Afro-asian	0.128*	0.057	0.131*	0.057	0.125*	0.057	0.182**	0.035	0.183**	0.035
Indian	0.172**	0.049	0.173**	0.048	0.166**	0.049	0.177**	0.033	0.183**	0.033
Pakistani	0.064	0.045	0.071	0.045	0.066	0.045	0.024	0.033	0.029	0.033
Chinese	–	–	–	–	–	–	0.250**	0.046	0.243**	0.047
speak	0.147**	0.046	–	–	0.049	0.062	0.171**	0.025	0.169**	0.025
write	–	–	0.164**	0.042	0.133*	0.057	–	–	–	–
N. of Obs.	839		839		839		2100		2100	

Base Category: No educational qualification, Bangladeshi. Robust standard errors are reported. *: Significant at 5 percent level. **: Significant at 1 percent level.

Our samples consist of 839 individuals for the FWLS, and 2100 individuals for the FNSEM. Our dependent variable takes the value 0 if the individual is unemployed and seeking a job or claiming benefits, and the value 1 if the individual works full- or part-time. Explanatory variables are the demographic and human capital characteristics available in the two data sets, including a dummy variable for the level of language proficiency. The results are reported in Table 2.5. For the FWLS, we report results conditioning on fluency only, and on fluency and written literacy.

Most coefficient estimates for the two data sets are very similar. Males have a significantly lower probability of being employed (13 percentage points in the FWLS, and 8 percentage points in the FNSEM). Being married increases employment probabilities by about 18 (17) percentage points. Having children influences, on the other hand, the employment probability negatively. These effects are consistent with evidence for British (male) natives (see Nickell [1980]).

For the FWLS, education coefficients are mostly insignificant. For the FNSEM, education coefficients are significant, and in the expected order of magnitude. In the last columns of Table 2.5, we run regressions which distinguish between education levels acquired in the UK, and in the home country. The coefficients on the UK educational degrees seem slightly larger than the coefficients on education acquired at home. However, we can not reject the null hypothesis that the coefficients are equal (neither in isolation, nor jointly).

Age is positively associated with employment probabilities, and the age profile is concave. Conditional on age, the time of residence in the UK does not have a significant effect on employment probabilities, for both the FWLS and the FNSEM. Indians, Afro-Asians and Chinese have higher probabilities of being employed than Pakistanis and Bangladeshis. Again, Bangladeshis seem to be the most disadvantaged group.

The coefficients on the language variables are quite large, and similar for the two data sets. English fluency is associated with a 15 (17) percentage point higher employment probability, using the FWLS (FNSEM) data. The coefficients are highly significant.

The FWLS data distinguishes between speaking, writing and reading abilities – information which is not available in most data sets on migrants' language abilities. One may argue that proficiency in the spoken language alone is not sufficient to affect labour

the next two weeks, and who have actively sought employment at some time during the previous four weeks.

market outcomes, but that writing skills are likewise needed. The positive coefficient of the fluency variable may then simply reflect the correlation between these two components of language capital. To investigate this point, we have included an indicator for writing abilities (columns 2), and both speaking and writing variables (columns 3). The effect of writing proficiency (unconditional on fluency) is slightly higher. When including both indicator variables, we find that writing abilities are associated with a 13 percentage point increase in employment probabilities, while speaking ability alone increases this probability by only 5 percentage points. The latter effect is not significant. This suggests that literacy in the dominant majority language, in addition to fluency, is important to obtain a job.

2.5.2 Employment, Endogenous Choice and Measurement Error

The above results suggest that language proficiency has a positive impact on employment probabilities. As we discussed above, however, the estimated coefficients may be seriously biased due to endogenous choice and measurement error. Furthermore, the effect of language on employment may be different for males and females. In this section, we address these issues. We estimate different models, addressing both these problems, and using the pooled sample, and males and females separately. We report the results in Table 2.6.

Table 2.6: Employment and Language

Specification		All	Males	Females	All	Males	Females
		FNSEM			**FWLS**		
1: OLS	Coeff	**0.170**	**0.166**	**0.172**	**0.147**	**0.190**	**-0.007**
	S.E.	0.025	0.024	0.041	0.046	0.037	0.070
2: Prop. Match.	Coeff	**0.102**	**0.102**	**0.133**	**0.100**	**0.112**	**-0.140**
	S.E.	0.049	0.060	0.103	0.117	0.123	0.120
3: Prop. Match.	Coeff	**0.223**	**0.261**	**0.141**	–	–	–
Measurement Error	S.E.	0.071	0.094	0.113	–	–	–

Robust standard errors are reported for specification 1; bootstrapped standard errors (based on 500 repetitions) are reported for specifications 2 and 3.

In the first row, we replicate our OLS results (based on the same specification as in Table 2.5), where we also report estimates for males and females separately. For the FNSEM data, the language coefficient is very similar for males and females, and significantly different from zero for both groups. For the FWLS, the coefficient for males is slightly larger than the coefficient for the pooled sample, while the coefficient for females is practically zero.

The second row reports results using the propensity score matching estimator, as we have explained in Section 2.2. Coefficients decrease slightly, which is compatible with unobserved ability being still present in the simple regression in row 1.

In the last row, we report results from estimations implementing the two stage estimator which takes account of measurement error (see (2.1) above). Coefficient estimates increase quite substantially. The results suggest that measurement error in the language variable leads to a substantial downward bias in estimated parameters.

Altogether, these results indicate that measurement error and endogenous choice bias the estimates of language effects in opposite directions. Our results suggest that the true effect of language on employment probabilities is substantial, and possibly larger than simple OLS estimates suggest. Overall, the results we obtain from the estimator which controls for measurement error suggest that fluency increases the probability that a male individual is employed, given that he looks for a job, by around 26 percentage points. The estimate for females is smaller, and not significant.

2.5.3 Earnings

We now turn to the effect of language on weekly gross earnings. Neither sample provides information on the number of hours worked per week, and we therefore consider only individuals who are working full-time.

In the FWLS, the dependent variable is the natural logarithm of gross (before tax) weekly earnings. The earnings variable in the FNSEM is gross weekly earnings, which is reported in categorical form (16 categories). In both samples there is a considerable percentage of working individuals who do not report their earnings (28% in the FNSEM and 45% in the FWLS).

To check the extent to which attrition is non-random, we compare the means of the language variables, origin dummies, the educational variables and other individual characteristics for individuals who do, and who do not report earnings. Results are

presented in Table 2.11. We also report the t-statistics for testing whether the means of the variables are significantly different. In some cases, we reject the null hypothesis of equal means, but there seems to be no systematic pattern of attrition across the two data sets.

Our final sample sizes for the earnings analysis are 254 individuals for the FWLS data, and 920 individuals for the FNSEM data. Results of straightforward log wage regressions are presented in Table 2.7, where we use the least squares estimator for the FWLS, and the least squares estimator at the midpoints for the FNSEM.[15]

As regressors, we include the same set of variables as in the employment regressions. Coefficient estimates on most variables are roughly similar for the two data sets. Males have a significant earnings advantage, compared to females. Having a degree more than doubles earnings, compared to holding no qualification. O-levels (or equivalent) alone increase earnings by about 17 (FWLS) or 24 (FNSEM) percent.[16]

In the last column, we use again the more detailed educational information in the FNSEM, and decompose educational attainments into overseas and UK qualifications. We find that the coefficients on UK qualifications are larger than overseas ones, but the joint null hypothesis that degrees acquired abroad have a significantly different effect on earnings than degrees acquired in the UK is rejected at the 5 percent level. Coefficients are only significantly different for A levels or equivalent degrees.

The coefficients on the ethnicity dummies indicate significant wage differences between ethnic groups. Like in the language and employment equations, Bangladeshis are the most disadvantaged group. Conditional on education, age and years of residence, their wages are 66 percent lower than those of the most successful group, the Chinese (FNSEM). In both data sets the earnings of Caribbeans are about 35 percent higher than Bangladeshis.

We find large and significant coefficients for the English fluency variables. The point estimates in the FNSEM and FWLS are quite similar, and indicate that English language proficiency is associated with about 21 (FNSEM) or 23 (FWLS) percent higher wages. Again, we use writing proficiency as an additional indicator for language proficiency (see columns 2 and 3). Interestingly, and different from the employment equation, fluency

[15]We have also estimated grouped regression models for the FNSEM (where the boundaries are transformed by taking logs). Results are almost identical.

[16]We compute here and in the following percent differences in earnings as $(e^{\hat{\beta}} - 1) * 100$, where $\hat{\beta}$ is the estimated parameter on the variable to which the discussion refers.

seems to be more important for wages than literacy.

Table 2.7: Earnings Regressions

	FWLS						FNSEM			
	1		2		3		4		5	
							All Qualifications		UK/nonUK Q	
Variable	Coeff	S.E.	Coeff	S.E.	Coeff	S.E.	Coeff	S.E.	Coeff	S.E.
Cons	3.551**	0.411	3.577**	0.412	3.546**	0.413	3.843**	0.243	3.809**	0.249
male	0.238**	0.072	0.251**	0.071	0.238**	0.072	0.107**	0.039	0.115**	0.039
married	-0.010	0.088	-0.008	0.089	-0.008	0.089	0.176**	0.051	0.160**	0.051
degree	0.786**	0.104	0.788**	0.106	0.781**	0.106	0.671**	0.048	–	–
Alevtea	0.206*	0.090	0.202*	0.093	0.201*	0.093	0.384**	0.051	–	–
OlevCSE	0.169	0.091	0.172	0.091	0.166	0.092	0.156**	0.043	–	–
Edegree	–	–	–	–	–	–	–	–	0.607**	0.056
EAlevtea	–	–	–	–	–	–	–	–	0.351**	0.054
EOlevCSE	–	–	–	–	–	–	–	–	0.120*	0.050
Fdegree	–	–	–	–	–	–	–	–	0.504**	0.066
FAlevtea	–	–	–	–	–	–	–	–	0.132	0.078
FOlevCSE	–	–	–	–	–	–	–	–	0.094	0.050
age	0.038	0.023	0.036	0.023	0.038	0.023	0.019	0.012	0.021	0.013
age^2	-0.045	0.029	-0.042	0.029	-0.044	0.029	-0.022	0.015	-0.025	0.015
YSM	0.026	0.015	0.030*	0.014	0.027	0.015	0.033**	0.007	0.032**	0.007
YSM^2	-0.050	0.035	-0.035	0.035	-0.050	0.036	-0.051**	0.019	-0.050*	0.020
Caribbean	0.302*	0.132	0.327*	0.130	0.301*	0.132	0.279**	0.076	0.301**	0.077
Afro-asian	0.081	0.125	0.109	0.123	0.083	0.125	0.224**	0.068	0.259**	0.068
Indian	0.311**	0.113	0.329**	0.112	0.310**	0.113	0.157*	0.069	0.206**	0.069
Pakistani	0.239*	0.118	0.251*	0.118	0.239*	0.119	0.025	0.072	0.066	0.073
Chinese	–	–	–	–	–	–	0.408**	0.083	0.416**	0.085
speak	0.204	0.115	–	–	0.171	0.161	0.180**	0.055	0.192**	0.055
write	–		0.149	0.103	0.040	0.145	–	–	–	–
No. of Obs.	254		254		254		920		920	

Base Category: No educational qualification, Bangladeshi. Robust standard errors are reported. *: Significant at 5 percent level. **: Significant at 1 percent level.

2.5.4 Earnings, Endogenous Choice and Measurement Error

Besides measurement error and endogenous choice of acquiring language proficiency, an additional difficulty with investigating earnings is non-random selection into the work-

force. Non-participation is large among minority immigrants, and in particular among females. It is likely that participation is selective, and correlated with the choice to acquiring language proficiency, thus biasing parameter estimates.

The conventional way to address non-random selection is to model the selection process and the earnings equation simultaneously. A simple estimator is a two step estimator which conditions earnings on the (generalised) residual from the first step auxiliary participation equation. To implement this approach requires identifying assumptions. We experimented with a number of possible exclusion restrictions. We are not confident about the validity of most exclusion restrictions that are feasible given the information in our data.[17]

We therefore refrain from estimating a joint model. To the extent that the participation choice is due to observables, our matching approach takes care of this problem. For any remaining selection, our strategy is to interpret the coefficients on the language variable as bounds, which is possible under some plausible assumptions. As we have seen in the last section, language has a positive effect on employment probabilities, and simple regressions show that it has also a positive effect on participation. If we are willing to assume that unobservables, which affect the participation probability, are positively correlated with unobservables which affect earnings, then the estimate of the language coefficient in an earnings regression on participants only is downward biased, compared to the hypothetical coefficient for the overall population. The intuition is simple: those individuals who are not proficient in the English language, but participate nevertheless, must be drawn from the upper part of the ability distribution to compensate for their language deficiencies, thus inducing a downward bias in the estimated language coefficient[18]. Accordingly, we can interpret the coefficient estimates we obtain on the sample

[17]For females, we considered to use variation in religious believes (conditional on origin) as an instrument for participation. The idea is that some religions may impose a strict role behaviour on females more than others, and religion may thus explain variation in participation. The FNSEM data distinguishes between Sikh, Hindu, Muslim, Christian, and no religion. These variables are jointly significant in an auxiliary first step participation regression. The generalised residual was not significant in the earnings regression, and hardly changed the language coefficient.

[18]More formally, suppose that the latent participation index p^* is linear in l, with $p_i^* = \alpha_0 + \alpha l_i + u_i$, and that the individual participates if $p_i^* > 0$. Suppose that the outcome equation is given by $y_i = \gamma_0 + \gamma l_i + v_i$, and assume that u_i and v_i are jointly normally distributed, with variances 1 and σ_v^2 and correlation coefficient ρ. Then selection could be accounted for by adding the generalised residual $E(v_i|p_i^* > 0) = \lambda(c_i)$ to the estimation equation, where $\lambda(c_i) = \phi(c_i)/\Phi(c_i)$, with ϕ and Φ being the

of participants as lower bounds of the effect of language on earnings.

In Table 2.8, we report results for the pooled sample, and for males and females separately. Splitting the sample into males and females leads to very small sample sizes, in particular for the FWLS, and most of our estimates are quite imprecise. We should therefore interpret results with care.

In the first row, we report the Least Squares results. While for the FWLS, coefficients for males and females are quite similar, the language coefficient using the FNSEM data is much larger for females than for males. Coefficient estimates for the FWLS are however not significant, with large standard errors for the separated samples.

Table 2.8: Earnings and Language

Specification		All	Males	Females	All	Males	Females
Estimation			**FNSEM**			**FWLS**	
1: OLS	Coeff	**0.180**	**0.121**	**0.354**	**0.204**	**0.173**	**0.167**
	S.E.	0.055	0.063	0.120	0.115	0.180	0.121
2: Prop. Match.	Coeff	**0.281**	**0.238**	**0.463**	**0.101**	–	–
	S.E.	0.108	0.103	0.186	0.174	–	–
3: Prop. Match.	Coeff	**0.356**	**0.460**	**0.844**	–	–	–
Measurement Error	S.E.	0.324	0.272	0.844	–	–	–

Robust standard errors are reported for specifications 1; bootstrapped standard errors (based on 500 repetitions) are reported for specifications 2,3.

In the second row, we report results from the propensity score estimator. Coefficients for both males and females are larger relative to the simple OLS estimator. This seems to be contrary to what endogenous choice of language acquisition would predict. However, as we discussed above, non-random participation may lead to downward biased estimates of language coefficients. The matching estimator corrects for participation selection, as long as it is on observables, and may therefore reduce the downward bias due to selective participation. Sample sizes for the FWLS data when we distinguish between males and females became too small for this estimator, and we only report results for the FNSEM.

density and distribution function of the standard normal, and $c_i = \alpha_0 + \alpha l_i$. We obtain the estimation equation $y_i = \gamma_0 + \gamma l_i + \sigma_v \rho \lambda(c_i) + e_i$. Omission of $\lambda(c_i)$ results in a biased estimate for γ. The expectation of the error term when omitting λ, conditional on l_i, is $\rho \sigma_v \mathrm{E}(\lambda(c_i)|l_i)$. Since λ decreases in c_i, the bias is downward for $\rho > 0$ and $\alpha > 0$.

In row 3, we implement our estimator which accounts for measurement error in addition. For females, the coefficient estimate becomes very large, and is estimated with very low precision. For males, coefficient estimates increase by factor 2, but the coefficient is not significant at the 5 percent level. Sample sizes are too small to draw robust conclusions from this evidence. We may however interpret the increase in coefficients when correcting for measurement error as evidence that measurement error leads to downward biased estimates also here.

2.6 Conclusion

Based on two recent surveys, we analyse the determinants of English language fluency for ethnic minority immigrants in the UK, and the effect of language on labour market outcomes. We also investigate the effect of other characteristics on language acquisition, and employment and earnings.

We find that in simple regressions, language proficiency is associated with higher employment probabilities and with higher earnings. Language effects may be under- or overestimated, due to endogenous choice of learning the language, and measurement error. We address both these issues. We use a matching estimator to address the endogenous choice of language acquisition. We combine our matching estimator with an IV type estimator to eliminate the downward bias due to measurement error, using information about the interview language for identification. Our results indicate that the bias induced by the two problems points in opposite directions, and that the effect of language on outcomes is larger than suggested by simple regression estimators. While OLS estimates indicate that language fluency increases employment probabilities by 17 percentage points, estimates that address both selection and measurement error suggest an increase by about 22 percentage points. Our analysis on earnings is less conclusive. OLS estimates suggest an earnings advantage of those who are proficient in English of about 18-20 percent. Estimates based on the estimator that addresses both endogenous selection and measurement error are insignificant.

The validity of our matching approach depends on our believes about whether the set of matching variables eliminates the problem of endogenous selection. The set of conditioning variables available to us includes indicators that are likely to be correlated with unobserved ability that sorts individuals into groups of those who do, and who

do not acquire the host country language, like education, ability tests, and partner information. However, if these variables do not fully account for unobserved factors that select individuals into the group of those who are proficient and non-proficient in the English language, language effects may still be upward biased.

An alternative way to address the problem of endogenous language choice would be data produced by an exogenous mechanism, providing immigrants with different opportunities to acquire language proficiency. This could be, for instance, a situation where immigrants have had different access to language facilities, and where the assignment to facilities is exogenous. One mechanism that could generate this are settlement policies that allocate immigrants to different communities upon arrival. Schemes like this were in place in different countries. Future research could use these assignment mechanism to address the problem of endogenous language choice.

Finally, we would like to stress again that the data we use in this analysis does not cover the entire immigrant population in the UK, but only those immigrants who belong to ethnic minority communities. According to the Labour Force Survey (2000), immigrants from ethnic minority groups constitute only 49 percent of the total immigrant population in the UK. Hence, our analysis covers only half of the immigrant population. Other research on UK immigrants (see e.g. Chiswick [1980], Bell [1997], Wheatley Price [2001], and Dustmann et al. [2002]) shows that the assimilation patterns of ethnic minority immigrants and white immigrants differ quite substantially. It is not unlikely that effects of language proficiency on economic outcomes are also different for these groups. More comprehensive surveys are needed to allow investigating language effect for the entire immigrant population in the UK.

2.7 Appendix

Table 2.9: Language Information

	All groups	Caribbean	Indian	Afroasian	Pakistani	Bangladeshi	Chinese
	Speaking, FWLS						
Very well	37.81	54.55	50.44	64.77	38.16	25.93	-
Quite well	23.12	13.64	27.43	27.27	26.05	18.46	-
Not well	20.12	18.18	18.14	5.68	21.32	22.82	-
Hardly	11.69	13.64	3.54	2.27	10	18.46	-
Not at all	7.26	–	0.44	–	4.47	14.32	-
	Reading, FWLS						
Very well	34.64	40.91	48.67	61.36	33.16	24.07	-
Quite well	21.12	18.18	23.89	26.14	21.58	18.67	-
Not well	15.86	22.73	14.16	7.95	17.11	16.8	-
Hardly	13.19	9.09	7.96	1.14	14.47	17.01	-
Not at all	15.19	9.09	5.31	3.41	13.68	23.44	-
	Writing, FWLS						
Very well	32.39	40.91	45.13	56.82	29.47	23.86	-
Quite well	19.2	18.18	21.68	23.86	20.79	15.98	-
Not well	16.61	22.73	15.49	13.64	18.16	16.18	-
Hardly	12.77	4.55	11.06	2.27	13.68	15.15	-
Not at all	19.03	13.64	6.64	3.41	17.89	28.84	-
	Speaking, FNSEM						
Fluent	48.73	89.65	39.98	65.63	25.56	25.97	56.59
Fairly	20.4	9.62	24.37	19.2	25.56	23.02	12.64
Slightly	21.2	–	25.84	11.76	32	34.25	18.13
Not at all	9.67	–	9.81	3.41	16.88	16.76	12.64

Table 2.10: Age and Time Patterns

Ethnicity	Age Migration		Year		Age	
	FWLS	FNSEM	FWLS	FNSEM	FWLS	FNSEM
Caribbean	19.460	20.379	1964.2	1963.435	49.755	50.927
S.D.	8.840	10.186	7.968	7.772	12.199	13.933
Indian	18.971	23.892	1973.2	1972.431	40.299	45.145
S.D.	10.235	13.535	9.750	9.515	11.004	14.384
Afro-Asian	17.813	20.662	1974.1	1973.828	37.976	40.735
S.D.	10.320	12.702	7.035	7.221	10.393	13.026
Pakistani	18.167	20.424	1976.4	1974.207	35.870	39.672
S.D.	9.374	11.275	9.634	9.677	11.953	13.738
Bangladeshi	18.676	20.579	1979.6	1977.695	33.266	36.645
S.D.	9.374	10.545	9.046	9.637	13.985	14.156
Chinese	-	22.088	-	1976.35	-	39.641
S.D.	-	11.860	-	8.670	-	12.532
All	18.663	21.418	1974.7	72.414	38.308	42.707
S.D.	10.084	11.918	1.402	9.944	13.587	14.572

Table 2.11: Attrition

	FWLS			FNSEM		
Variable	Report	Missing	Diff.	Report	Missing	Diff.
	Mean	Mean	t-value	Mean	Mean	t-value
male	0.618	0.730	2.59	0.695	0.682	0.46
married	0.767	0.802	0.91	0.840	0.861	1.04
nchild	1.480	1.995	3.53	1.484	1.503	0.20
degree	0.153	0.110	1.36	0.2	0.230	1.22
Alevtea	0.212	0.144	1.92	0.163	0.167	0.18
OlevCSE	0.208	0.278	1.74	0.269	0.242	1.06
age	37.704	38.274	0.59	39.358	40.726	2.32
YSM	22.303	22.783	0.54	22.021	22.542	1.00
Caribbean	0.204	0.182	0.59	0.229	0.101	6.27
Afro-asian	0.145	0.129	0.49	0.244	0.227	0.67
Indian	0.338	0.264	1.73	0.225	0.382	5.70
Pakistani	0.173	0.278	2.69	0.140	0.181	1.86
Chinese	–	–	–	0.078	0.029	4.05
speak	0.877	0.865	0.40	0.877	0.876	0.03
write	0.850	0.793	1.58	–	–	–
No. of Obs.	254	208		413	920	

Note: t-statistics computed as $(m_1 - m_2)/\sqrt{se_1^2 + se_2^2}$, where m_i, se_i are means and standard errors of the two sample values, respectively.

Chapter 3

Ethnic Enclaves and Language Proficiency of Immigrants

Abstract:

In this chapter we investigate the association between immigrant concentration and language proficiency of immigrants in the UK. We provide a detailed analysis of the geographic distribution of immigrants and ethnic minorities in the UK. We present a simple human capital model where investment in language capital depends on the level of ethnic concentration at immigrants' arrival and following changes. The model also allows for potential initial selection based on the level of language fluency at arrival. Differently from previous literature, we can observe ethnic concentration at arrival. Results show that the effect of ethnic concentration is only significant for the fluency of the female immigrants' sample. Furthermore, there is no significant evidence of initial sorting.

3.1 Introduction

Recent analyses for the US, Canada, Australia, the UK, Israel, and Germany show that fluency and literacy in the dominant host country language are important components for explaining immigrants' labour market success (for all, Rivera-Batiz [1990], Chiswick and Miller [1995], Dustmann [1994]). Chapter 2 provides this analysis for the UK. Language fluency and literacy are, furthermore, factors which are potentially influenceable by active policies.

A most important factor which is associated with language performance is the ethnic composition of the immigrant's immediate neighbourhood. Ethnic enclave economies are characterised by locational clustering of firms and co-ethnic employees. Therefore, language and cultural barriers, which may hinder immigrants' success in the general labour market, are not encountered in the ethnic enclave economy. As well as labour market opportunities, in an ethnic enclave, minority individuals have access to ethnic goods, trade opportunities with "co-ethnics", information networks and communication in the language of origin.

A number of papers for the US have established a negative association between minority concentration, and language. McManus [1990] shows that Hispanic immigrants with low proficiency in English tend to live in enclaves. He suggests that enclaves may represent a convenient location for low proficient immigrants. McManus also finds that earnings returns to English proficiency decrease with the size of the ethnic enclave. However, earnings losses deriving from low English proficiency are lower in areas at high ethnic concentration.

Chiswick and Miller [2002] analyse the association between language and ethnic and immigrant concentration for male immigrants in the US. They investigate what the poles of attraction which lead to the clustering of immigrants are. These can be summarised in the proximity to "ports" of entry, to early immigrant settlements, and to job opportunities. Their findings show that low immigrant concentration areas are associated with higher proficiency in English. In addition, immigrants' earnings are lower in higher minority concentration areas.

In an influential recent study, Lazear [1999] has developed a model where trade between different groups requires the ability to communicate with each other. To enhance trading possibilities, minority individuals may learn the language of the majority

group. The incentive of learning the language is larger the smaller the relative size of the minority group. Moreover, minority individuals with low proficiency in the majority language may sort themselves into communities where individuals speaking their own minority language are concentrated. As Lazear points out, the two processes both lead to a negative association between minority concentration and fluency in the majority language.

In this chapter, we analyse the association between ethnic/immigrant concentration and language proficiency in the UK. To this purpose, we look at a sample of ethnic minority immigrants. To our knowledge, data on language proficiency is not available for other immigrants.

One contribution to the existing literature is to provide a detailed analysis of the geographic distribution of immigrants and ethnic minorities in the UK. We explore how immigrant and ethnic minority settlements have developed over time. Our data sources enable us to observe such developments at different spatial levels. In particular, we can look at the distribution of immigrants in relatively small geographical areas. It is very important to have access to detailed demographic characteristics because, as will be shown in the Sections below, unlike in the US, in the UK immigrants are concentrated in a few regions. This implies a low variation in their distribution at regional level, which may render this kind of analysis difficult[1]. To our knowledge, no other data set on immigrants allows an analysis at such detailed geographic level.

In previous literature, estimates are performed on ethnic concentration measured at the same time as the survey. The data sets used are cross-section, for which information on past levels of ethnic concentration is usually not available. Language fluency, though, is likely to be affected by the level of ethnic concentration when the immigrant arrived, rather than after several years of stay in the host country. In this chapter, we develop a simple model of human capital where investment in language capital depends on ethnic concentration at the time of arrival and subsequent changes. Previous specifications are enclosed in our model. We use information on past immigrant distribution to derive initial immigrant concentration.

Finally, the last contribution of the chapter is to investigate the correlation for male and female immigrants separately. Evidence from Chapter 2 suggests that female immigrants' participation to the labour market is substantially lower than for male im-

[1]Chapter 5 discusses this issue further.

migrants. Differences in the social and economic situation between male and female immigrants (also due to cultural and traditional reasons) may account for differences in the effect of ethnic concentration on language proficiency.

We base our analysis on data from a survey on ethnic minorities in the UK: the Fourth National Survey on Ethnic Minorities (FNSEM), which has been collected between 1993 and 1994.[2] The data set consists of two sub samples. It contains a main sample of respondents belonging to ethnic minorities, and a reference sample of individuals belonging to the white majority population. It includes questions on the social and economic conditions of the interviewees, and measures on language proficiency.

The data set also provides us with geographical information at ward level[3]. We are therefore in the position to combine this survey with information from the UK Population Census for the years 1991, 1981, and 1971 to derive immigrant concentration at the time of immigrants' arrival and at the time of the survey.

Our results suggest that ethnic concentration is associated to the language proficiency of female immigrants, but not of male immigrants. This is in clear contrast with results from the US (Chiswick and Miller [2002]). For female immigrants, we find that investment in language capital is affected by ethnic concentration both in levels and in changes. Higher initial concentration and larger changes in concentration imply slower learning rate. Finally, there is no evidence of initial sorting depending on the level of language fluency for the whole sample. This last result also is in contrast with evidence from the US (Lazear [1999]).

The chapter is structured as follows. Section 3.2 discusses the data sets and gives some descriptive statistics. Section 3.3 provides an outline of the historical development of ethnic and immigrant settlements in the UK from 1971 to 1991. In Section 3.4 we explain the economic rationale behind the association between language and ethnic concentration. Results are presented in Section 3.5, and Section 3.6 concludes.

[2]Initially, as in Chapter 2, the analysis was conducted on another data set which also "boosts" ethnic minorities and provides information on language proficiency: the Family and Working Lives Survey (FWLS). However, in this case, the sample size is insufficient for our purposes, and results from this data set are not informative.

[3]In the UK, a ward is the smallest geographical area identified in the Population Census. According to the 1991 Census, the mean population within ad ward is 5459 individuals, and the median is 4518. The number of wards in the 1991 Census is 8572.

3.2 Data

The Fourth National Survey on Ethnic Minorities (FNSEM) is a cross-sectional survey, which has been carried out between 1993 and 1994. Individuals included are aged 16 or more, and of Caribbean, Indian, Pakistani, Bangladeshi, or Chinese origin. There are 5196 observations in the minority sample, and 2867 observations in the independent comparison sample of white individuals. More than 77% (4019) of the individuals in the ethnic minority sample are foreign born, of which half are women.

The FNSEM contains information on language. The language variable is based on the interviewer's evaluation on the individual's spoken language ability, on a 4 point scale. For the empirical analysis, we re-define the language indicator as a dichotomous variable which is equal to 1 if individuals fall in the categories "fairly" or "fluently".

The sample design of the FNSEM was complex, considering wards with any percentage of ethnic minorities on the population and oversampling Bangladeshis to obtain a sufficient sample size. For more details, see Appendix 1 in Modood and Berthoud [1997].

Table 3.1: Description of Variables and Sample Characteristics

Variable	FNSEM		Description
	Mean	S.D.	
Speak	0.691	0.462	Dummy=1 if spoken English is good or very good
Male	0.505	0.500	Dummy=1 if male
Age	42.604	14.407	Age
YSM	21.367	10.001	Years of stay in the UK
Degree	0.127	0.333	Dummy=1 if university degree
Alevels	0.109	0.312	Dummy=1 if high vocational
WIC at entry	0.158	0.114	Ward immigrant concentration at arrival
Oethr91	0.141	0.141	Ward own ethnic concentration in 1991

Table 3.1 summarises information available in the two data sets on variables used for the analysis. Chapter 2 provides a more extensive discussion on language determinants.

About 70% of the sample population speak English fluently. Average age is 42 and the average period lived in the UK is 21 years. In the sample, 12.7% individuals have a degree and 10.9 have A level or equivalent qualification. The average ward immigrant/population ratio at arrival is 15.8% and average concentration of the ethnic minority to which the respondent belongs is 14.1%.

3.3 Immigrant and Ethnic Concentration in the UK

Information on the characteristics of the area where the respondent lives, at the time of migration and at the time of the surveys, is derived merging the FNSEM with the 1971, 1981, and 1991 Population Censuses. The information provided by the three Census data sets, however, is not entirely consistent across decades. The 1991 Population Census allows to separately identify ethnic minority, immigrant and native individuals. It also allows to infer information on the density of the specific ethnic minorities to which the respondents belong. Data from the Census is available in selected cross-tabulations, and information on the immigrant status of ethnic minorities is not accessible. However, evidence shows that ethnic minorities born in the UK are mostly second generation immigrants, i.e. children of ethnic minority individuals born abroad (see Introduction). Extensive sociological literature shows that ethnic minorities in the UK tend to maintain their original culture and to hand it on to the next generations (Modood and Berthoud [1997]). This is likely to imply that the language of origin is spoken inside ethnic minority communities, regardless of whether their members are foreign or UK-born.

We then define the following measures of concentration (at ward and district levels) from the 1991 Population Census. Let ethnic density be the ratio between the number of individuals in the area who report to belong to some ethnic minority[4] and the population in the same area. In addition, let "own" ethnic density be the ratio between the number of persons in the area who are of the same ethnic origin as the respondent and the total population in the same area.

The 1981 and 1971 Censuses only provide data on immigrant status. To obtain consistent time trends on ethnic density, we have to use an ethnic density variable based on country of origin, rather than ethnicity. Therefore, from the 1991, 1981 and 1971 Censuses, we define as immigrant concentration the ratio between the number of persons in the area who were born outside the UK and the total population in the same area.[5]

A further problem stems from the fact that, whereas the smallest spatial unit avail-

[4]In the Census nine ethnic groups are defined: White, Black Caribbean, Black African, Black other, Indian, Pakistani, Bangladeshi, Chinese, and "other". As in the UK whites are the ethnic majority, all other groups are considered ethnic minorities.

[5]As in Chapter 6, we have considered using the percentage of immigrants from South Asia and the West Indies only, to have a variable closer in definition to the one of "own" ethnic density. However, we have found that results do not change significantly.

able for the 1991 and 1981 Censuses is the electoral ward, for the 1971 Census the smallest spatial area is a district[6]. Therefore, we use district concentration values as proxies for ward level for 1971. In this way, the information on smaller spatial areas is preserved from 1981 onwards, at a loss, however, of information on the variation in concentration at ward level for 1971[7].

To derive immigrant concentration at the time of migration, we use linear interpolation across the years. Due to the time limitation of the data, we cannot infer immigrant concentration at arrival for immigrants settled before 1971.

Table 3.2 reports figures on the distribution of immigrants and ethnic minorities in the UK. In the first three rows of Table 3.2 we describe figures for immigrants derived from the 1971, 1981, and 1991 Censuses. In the next rows we describe analogous figures for ethnic minorities from the 1991 Census. According to the Population Censuses, the immigrant-population ratio steadily increased from 6.5 percent in 1971 to 7.5 percent in 1991. Immigrants are strongly concentrated in Greater London, where they represent 15.2% of the overall population already in 1971 and 21.7% in 1991. Furthermore, about 41% of all immigrants in the UK live in Greater London. Similarly, ethnic minorities represent 5.2 percent of UK population, but 16.6% of Greater London population. In addition, 45.5% of ethnic minorities live in Greater London.[8] The distribution among ethnic minorities is, however, diverse. Caribbeans and Bangladeshis are more strongly concentrated in London than other groups, with 60% and 54% of their population inhabiting the capital, respectively. Indians and Chinese follow, with about 40% of their community living in London. Finally, only 20% of Pakistanis live in London. Table 3.2 clearly illustrates how London is increasingly a pole of attraction for immigrants and ethnic minorities[9]. The remaining immigrants are distributed in a few more areas, mostly in the South East and West Midlands[10].

[6]A district is the second larger spatial area after the ward. According to the 1991 Census, the mean population within a district is 128,566 individuals, and the median is 102,939.

[7]In Section 3.7 we discuss differences in concentration between ward and district.

[8]Figure 5.4 illustrates how the stock of immigrants in 1971 is positively associated to the change in the immigrant population between 1981 and 1991 at county level. These figures are consistent with evidence from the US. Bartel [1989] argue that immigrants in the US tend to settle in areas where immigrant settlement is already strong.

[9]More detailed information on the composition of immigrants in London can be found in the Introduction to the thesis.

[10]This high level of concentration arose partly because immigrants followed the pattern of demand

Table 3.2: Immigrants and Ethnic Minorities

	Year	% of pop.	% of GL pop.	% living in GL
Immigrants	1971	6.5	15.2	38.4
	1981	6.9	18.1	38.2
	1991	7.5	21.7	41.2
All Ethnic Minorities	1991	5.2	16.6	45.5
Caribbean	1991	1.0	4.3	58.7
Indian	1991	1.7	5.2	42.2
Pakistani	1991	0.9	1.3	19.5
Bangladeshi	1991	0.3	1.3	54.3
Chinese	1991	0.3	0.8	40.0

Notes:

GL: Greater London.

Source: Population Census 1971, 1981, 1991.

More information on the distribution of immigrants in the UK can be derived from the dissimilarity index. The segregation index is defined as:

$$ID = \frac{1}{2}\sum_{i=1}^{N}\left|\frac{w_i}{W} - \frac{m_i}{M}\right| \tag{3.1}$$

where w_i and m_i are the numbers of majority and minority individuals living in area i, respectively. W is the total number of majority individuals, M is the total number of minority individuals, and N is the total number of areas considered. By definition, the index ranges between 0 (the minority group are evenly distributed with respect to the majority group - no segregation) and 1 (the minority group and the majority group live in separate areas - total segregation). The index of dissimilarity can be interpreted as the percentage of either of the two groups that would have to move in order to be evenly dispersed with respect to the other group (see Duncan and Duncan [1955]). Table 3.3 reports dissimilarity (or "segregation") indexes at different spatial levels for immigrants and ethnic minorities.

Table 3.3 shows that at spatial level immigrants and, in particular, ethnic minorities appear to be increasingly clustered, the smaller the geographical area considered. In 1991, ID at ward level is 0.412 for immigrants and 0.632 for ethnic minorities, whereas

in the labour market. They settled in the area of Greater London to work in service industries such as transport, in the West Midlands to work in the metal manufacturing industries, and in the North West to work in the textile industry (see Duffield [1985] and Ratcliffe [1981]).

at county level[11] it is 0.323 for immigrants and 0.474 for ethnic minorities. At time level, the index of dissimilarity for immigrants has increased by about 5% between 1971 and 1991 at both county and district levels. This is consistent with the increasing representation of immigrants in the London population in Table 3.2. Furthermore, there are differences in the index of dissimilarity across minority groups. In 1991, Black Caribbeans, Pakistanis and Bangladeshis, each with an ID of more than 0.7, appear to be more clustered than the Chinese and Indians, with ID equal to 0.439 and 0.663, respectively.

Table 3.3: Segregation Indexes

	Year	Ward	District	County
Immigrants	1971	-	0.319	0.276
	1981	0.402	0.331	0.293
	1991	0.412	0.362	0.323
All Ethnic Minorities	1991	0.632	0.546	0.474
Caribbean	1991	0.704	0.639	0.566
Indian	1991	0.663	0.595	0.506
Pakistani	1991	0.766	0.627	0.481
Bangladeshi	1991	0.760	0.636	0.528
Chinese	1991	0.439	0.345	0.291

Source: Population Census 1971, 1981, 1991.

Tables 3.2 and 3.3 display the general picture of the degree of clustering of immigrants in the UK. In Table 3.4 we describe the distribution of the immigrants included in the sample. We report ethnic concentration of the individual's own group in 1991, immigrant concentration at arrival, and change in immigrant concentration. The complex sampling scheme of the FNSEM may account for differences between the distribution of respondents and the overall distribution derived from the whole minority population in the Census. According to Table 3.4, Caribbean respondents live in neighbourhoods where their own group represents 8.6% of the population, but, at arrival, immigrants represent 17% of the population and the average increase, after that, is 7.6%. Indians, African Asians[12], Pakistanis, and Bangladeshis live in areas where their own ethnic group

[11]Note that the Greater London region discussed in Table 3.2 consists of two counties, Inner and Outer London.

[12]African Asians are individuals who consider to be of Indian ethnicity, but were born in Africa. Such a distinction is not available in the Census.

accounts for about 16% (17% for Indians and Bangladeshis) of the population. These groups also settled in areas at relatively high concentration of immigrants, from 14% for the Pakistanis, to 20% for the Bangladeshis. Also, for these groups, the average increase in immigrant concentration since arrival was around 12-13%. Finally, the Chinese live in areas where they represent 0.9% of the population. At their arrival, immigrant concentration was on average 10.9% and it increased by 5.6%.

Table 3.4: FNSEM - Mean Levels and Changes in Concentration by Ethnic Group

	1991 Ethn.Conc.		Imm.Conc.		ΔImm.Conc.		
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Obs.
Caribbean	0.086	0.068	0.167	0.106	0.076	0.102	84
Indian	0.175	0.151	0.158	0.123	0.124	0.129	478
African Asian	0.164	0.140	0.150	0.101	0.138	0.128	428
Pakistani	0.164	0.137	0.139	0.103	0.132	0.114	472
Bangladeshi	0.177	0.170	0.203	0.126	0.112	0.102	384
Chinese	0.009	0.008	0.109	0.082	0.056	0.086	132

Note: Ethnic concentration of own group. Immigrant concentration at arrival. Change in immigrant concentration between time at arrival and time of the survey.

3.4 Language and Ethnic Concentration

3.4.1 Some Evidence

The value of language capital differs across locations in the host country, according to the relative size of the ethnic minority population the immigrant belongs to. To see the raw correlation between language proficiency and concentration, Table 3.5 presents the mean levels of ethnic and immigrant concentration among the fluent and the non-fluent. As the empirical analysis is performed separately by gender as well, the Table reports the same information for males and females. Immigrants who are not fluent in English tend to live in areas at higher ethnic or immigrant concentration than those who are fluent.

On average proficient individuals live in a community where, in 1991, 12% of the population are from the same ethnic minority, and at arrival 15% of it were born outside

the UK. The same individuals have seen an increase of 12% in the number of immigrants living in their same ward. In contrast, on average non proficient individuals live in areas characterised by higher minority concentration (18.6%) and which, at their arrival, contained more immigrants (17%). The change in immigrant concentration is not significantly different from that of fluent individuals. Values for the male and female sub-samples are similar.

Table 3.5: Ward Concentrations by Proficiency Level - All

	All				Males				Females			
Proficiency	Yes		No		Yes		No		Yes		No	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Ethn.Conc.1991	0.120	0.126	0.186	0.161	0.126	0.132	0.197	0.167	0.113	0.118	0.180	0.158
Obs.	2546		1125		1428		418		1118		707	
Imm.Conc.	0.149	0.108	0.171	0.123	0.151	0.107	0.169	0.123	0.147	0.109	0.172	0.123
ΔImm.Conc.	0.119	0.121	0.121	0.115	0.131	0.127	0.116	0.112	0.106	0.112	0.123	0.116
Obs.	1241		737		661		179		580		558	

Notes:
Ethnic concentration of own group. Immigrant concentration at arrival. Change in immigrant concentration between time at arrival and time of the survey.

3.4.2 Derivation of the Language Equation

So far, we have provided evidence that immigrants tend to settle in areas where there are already many of their kind. Furthermore, non-fluent immigrants tend to live in areas at higher immigrant concentration than fluent individuals. We have seen that the literature has often established a negative link between ethnic concentration of the region where the immigrant settles and his/her language skills. However, this analysis is incomplete in many ways.

Before we set up a simple model which puts some structure on our empirical analysis below, we will start with some considerations about why we should expect any relationship between language capital, and ethnic concentration of minorities in the neighbourhood where the individual lives.

Language is a means of decreasing transaction costs if individuals trade with each other. This idea has been developed formally by Lazear [1999]. Accordingly, the higher

the percentage of individuals who speak a minority language, the lower will be the incentive to acquire the majority language. This simple idea has consequences for the intensity with which immigrants seek to acquire proficiency in the majority language, as well as the initial choice of settlement. In particular, the initial choice of settlement will depend on the level of fluency in the language of the host country at arrival: those who have no knowledge of the language spoken by the majority community may find it beneficial to settle in areas with high minority concentrations. In addition, the minority concentration in the area where the individual settles affects the rate of subsequent investments. Finally, changes in minority concentration affect these same investments.

To model the initial selection process, and subsequent investment decisions, and the way these are related to ethnic concentrations is difficult. It requires a structural theoretical approach, and data of complete migration histories.

Previous literature has estimated the association of language and ethnic concentration either by including a concentration measure to a language equation (Chiswick and Miller [1996] and Chiswick and Miller [2002]), or by adding an interaction term between ethnic concentration, and years of residence (Lazear [1999]). In all these studies, ethnic concentration is measured at the time of the interview, which could mean many years after the immigrant's arrival. In what follows, we set up a very simple human capital model, where the rate of investment in language capital depends on initial ethnic concentration, and changes thereof. The consequent empirical specification requires less assumptions than previous work. We carefully state the assumptions we make, and the evolving interpretation of our parameter estimates. We also explain what the additional assumptions implied by previous work are.

Assume for the moment that immigrants do not select into areas according to their initial language skills. We will relax this assumption below. Assume also that immigrants, after having settled, do not move[13]. Then a higher initial concentration should reduce the benefits of learning the foreign language. A simple matching model where productive matches take place only if both partners communicate in the same language would deliver that result (Lazear [1999]). However, we have seen that ethnic concentra-

[13]This assumption has to be maintained in any analysis that uses cross-section data only to estimate the relationship between language and ethnic concentration. It may be restrictive. However, evidence from the 1991 Census shows that the internal mobility of ethnic minorities is lower than that of whites. On the other hand, Bartel [1989] found the opposite trend for immigrants and natives in the US.

tion may change over time. This again affects language acquisition, and the immigrant adjusts behaviour accordingly.

If these changes are unforeseen, modelling these adjustments requires a dynamic model. We assume here that these changes are deterministic. In addition, we assume that ethnic concentration changes at a constant rate. In a deterministic setting, an extension of the basic Mincer model gives us some structure on how language proficiency should relate to ethnic concentration. Denote the language proficiency of individual i in period t (where t is time after arrival), measured on some continuous scale, as L_{it}. Denote language proficiency at arrival as L_{i0}. The immigrant invests into language capital after arrival, with the rate of investment at time τ being equal to $k_{i\tau}$.

In a discrete time framework, language capital learned by individual i will follow the path

$$L_{i1} = L_{i0} + \rho k_{i0} L_{i0} = L_{i0}(1 + \rho k_{i0}) \tag{3.2}$$

$$L_{i2} = L_{i1} + \rho k_{i1} L_{i1} = L_{i0}(1 + \rho k_{i0})(1 + \rho k_{i1}) \tag{3.3}$$

$$\dots \tag{3.4}$$

$$L_{it} = L_{i0} \prod_{s=0}^{t-1} (1 + \rho k_{is}) \tag{3.5}$$

where ρ is the rate of return to the investment on language. Taking the logarithm of both sides and remembering that $\ln(1 + x) \cong x$ when x is small yields

$$\ln L_{it} = \ln L_{i0} + \rho \sum_{s=0}^{t-1} k_{is}. \tag{3.6}$$

In a continuous time framework, language capital at time t can be written as

$$\ln L_{it} = \ln L_{i0} + \rho \int_0^t k_{i\tau}\, d\tau\,. \tag{3.7}$$

We further assume that the rate of investment $k_{i\tau}$ depends on both the ethnic concentration of the area to which the immigrant is allocated, and on the change in ethnic concentration. Denote the initial concentration for some spatial measure j to which immigrant i has been allocated as $c_{j(i)0}$. Furthermore, assume that the rate of change in ethnic concentration is constant, and denote this change as $\Delta\, c_{j(i)}$. Finally, assume that investment declines linearly, and approaches zero at retirement (after which language

capital has no value any more), which occurs T periods after arrival. Then the rate of investment at any time τ is given as

$$k_{i\tau} = (1 - \frac{\tau}{T})[\kappa_{i0} - \gamma\, c_{j(i)0} - \delta\, \Delta\, c_{j(i)}], \tag{3.8}$$

where κ_{i0} is the initial investment in language for individuals moving into areas at zero concentration, and γ and δ are positive coefficients. Combining (3.7) and (3.8) yields

$$\ln L_{it} = \ln L_{i0} + \rho[\kappa_{i0} - \gamma\, c_{j(i)0} - \delta\, \Delta\, c_{j(i)}] \int_0^t (1 - \frac{\tau}{T} d\tau). \tag{3.9}$$

Solving the integral and simplifying the coefficients results in the following estimable expression:

$$l_{it} = \alpha_0 + \alpha_1\, t + \alpha_2\, c_{j(i)0}\, t + \alpha_3\, t^2 + \alpha_4\, c_{j(i)0}\, t^2 + \alpha_5\, \Delta\, c_{j(i)}\, t + \alpha_6\, \Delta\, c_{j(i)}\, t^2, \tag{3.10}$$

where $l_{it} = \ln\, L_{it}$. By construction, the parameters $\alpha_1, \alpha_4, \alpha_6$ are positive, and $\alpha_2, \alpha_3, \alpha_5$ are negative. The initial level of language capital is absorbed by the intercept term. Equation (3.10) says that language capital increases with time in the country ($\alpha_1 > 0$), but language capital accumulation slows down with more years in the host country ($\alpha_3 < 0$). Furthermore, at higher initial concentration correspond smaller levels of investment ($\alpha_2 < 0$), but at a higher rate of accumulation ($\alpha_4 > 0$). Similarly, initial level of investment is lower the larger the change in immigrant concentration ($\alpha_5 < 0$), but with a lower rate of decrease ($\alpha_6 > 0$).

Estimation of equation (3.10) gives us estimates of the effect of initial ethnic concentration, and its changes, on language capital, under the assumptions we have stated above. Notice also that (3.10) corresponds to a first order approximation if we allow ethnic concentration to increase or decrease with a varying rate. Finally, equation (3.10) shows that the standard specification of regressing language proficiency on years since migration and its square identifies composite parameters that vary across individuals according to the level and the change in ethnic minority concentration.

One key assumption that we have so far maintained is that individuals are randomly allocated to different areas. This is unlikely to be the case. Those who have no command of English language upon entry may find it more appropriate to choose areas with higher concentrations. Accordingly, the initial level of language capacity (absorbed in the intercept α_0) should be correlated with the initial concentration $c_{j(i)0}$. Suppose we can

write the log of initial language capital as a linear projection on the ethnic concentration in the area where the immigrant settles[14]. Then we obtain

$$l_{it} = \alpha_0 + \phi\, c_{j(i)0} + \alpha_1\, t + \alpha_2\, c_{j(i)0}\, t + \alpha_3\, t^2 + \alpha_4\, c_{j(i)0}\, t^2 + \alpha_5\, \Delta\, c_{j(i)}\, t + \alpha_6\, \Delta\, c_{j(i)}\, t^2, \quad (3.11)$$

According to this specification, a negative estimate of the parameter ϕ indicates that those with higher language capital upon arrival select into areas with lower ethnic concentrations. Furthermore, parameters on the interaction terms allow testing the hypothesis whether subsequent investments into language capital are in fact influenced by initial ethnic concentration, or changes in concentration.

To summarise, the assumptions we have imposed here are: (i) immigrants do not move after first settlement, (ii) perfect foresight, and (iii) the change in ethnic concentration is constant.

Chiswick and Miller [2002] introduce concentration measures only in levels, where measurement is at the time of interview. They interpret the coefficient estimate as an effect of ethnic concentration on language acquisition. In addition to (i)-(iii), they assume that (iv) the change in ethnic concentration over time is equal to zero, (v) the rate of investment into language capital is not affected by ethnic concentration, and (vi) immigrants do not select into areas according to their initial language potential. The estimate on the level variable has no clear interpretation, as it reflect the impact of ethnic concentration on language acquisition, independent of the number of years of residence.

McManus [1990] also uses concentration in levels, measured at the time of the interview. However, he does not attributes causality to its coefficient and acknowledges that immigrants may settle according to their initial proficiency.

The formulation in Lazear [1999] relaxes assumptions (v) and (vi) by allowing for an interaction term of ethnic concentration and years of residence, and interpreting the level estimate as being due to selection. However, he maintains assumption (iv) and also assumes that the effect of ethnic concentration on investment in language capital is constant.

[14]The linearity assumption can be easily relaxed. A quadratic relationship (such as in Lazear [1999]) could be an alternative. However, our estimates do not support it.

3.5 Results

In this Section we estimate equation (3.11) using a linear probability model. We discuss results for the whole sample, and separately for male and female immigrants.

Table 3.6 reports estimation results. In the first three columns we estimate the standard specification, as chosen by Chiswick and Miller [2002], with a second order polynomial in years since migration, but with ward immigrant concentration at the immigrant's arrival (rather than interview year). In the second set of columns, we add interaction variables as in specification 3.11. We report F-statistics for the joint significance of the years since migration polynomial, and the joint significance of the interactions with ward ethnic concentration at entry and its change. Our specifications are conditional on age at entry, and educational qualification dummies[15].

According to the first specification, the years since migration polynomial is significant for the pooled sample as well as for males and females, with a p-value of 0.000, indicating an increase in language capital with time of residence at a decreasing rate. Ethnic concentration at ward level upon immigrants arrival is however insignificant. The second set of columns adds the various interaction terms. Most interesting here are the differences between males and females. For males, all ethnic concentration variables are insignificant. This evidence suggests that male immigrants do not choose where to settle according to their initial language fluency. Furthermore, ethnic concentration (both in level and in change) has no effect on the accumulation of language capital.

For females, adding the two sets of interaction terms leads to an increased effect of years since migration. Furthermore, the interaction terms for immigrant concentration at arrival are jointly significant (p=value = 0.023). This suggests that the investment in language capital is lower the higher initial immigrant concentration and it increases faster. Also, the interaction terms for the change in immigrant concentration are jointly significant (p=value = 0.001) and indicate a strong dependence of investment intensity on the change in concentration. The level term of ethnic concentration at arrival is insignificant, indicating that initial selection on language capacity is weak. The effect of an additional year in the UK for a female who lives in a ward with average immigrant concentration (0.159) and average change in concentration (0.114) is

[15]Refer to Chapter 2 for a discussion of these coefficients.

Table 3.6: Language Equations

	All		Males		Females		All		Males		Females	
	Coeff	t-ratio	Coeff	t-ratio	Coef	t-ratio	Coeff	t-ratio	Coeff	t-ratio	Coef	t-ratio
Male	0.187	8.84	–	–	–	–	0.192	9.28	–	–	–	–
Age at Entry	-0.011	-15.37	-0.010	-9.08	-0.013	-12.68	-0.012	-16.16	-0.010	-9.26	-0.013	-13.03
Alevels	0.264	11.14	0.176	6.69	0.348	9.91	0.259	10.83	0.176	6.60	0.326	9.26
Degree	0.387	17.55	0.327	11.93	0.447	12.15	0.376	15.93	0.328	11.50	0.428	11.39
YSM	0.022	2.90	0.036	3.58	0.015	1.46	0.037	2.45	0.012	0.59	0.051	2.86
$YSM^2/100$	-0.011	-0.43	-0.074	-2.02	0.022	0.63	-0.078	-1.38	-0.006	-0.01	-0.115	-1.63
WIC at entry	0.002	0.02	0.182	1.44	-0.103	-0.54	0.145	0.40	-0.462	-1.08	0.512	1.11
YSM*WIC at entry	–	–	–	–	–	–	-0.086	-1.28	0.094	1.22	-0.181	-2.11
YSM^2*WIC at entry/100	–	–	–	–	–	–	0.488	1.84	-0.254	-0.84	0.857	2.47
YSM*ΔWIC	–	–	–	–	–	–	-0.030	-0.98	0.013	0.35	-0.053	-1.33
YSM^2*ΔWIC/100	–	–	–	–	–	–	-0.059	0.41	-0.067	-0.35	0.099	0.51
Constant	0.606	8.97	0.686	8.17	0.654	7.12	0.598	6.16	0.850	6.59	0.545	4.54
F-test: YSM=0, YSM^2=0	p=0.000		p=0.000		p=0.000		p=0.000		p=0.037		p=0.000	
F-test: YSM*WIC=0, YSM^2WIC=0	–		–		–		p=0.026		p=0.279		p=0.023	
F-test: YSM*ΔWIC=0, $YSM^2\Delta$WIC=0	–		–		–		p=0.031		p=0.939		p=0.001	
N.Obs.	1978		840		1138		1978		840		1138	

Notes:

Dependent variable = 1 if fluency is good or very good.

Standard errors are adjusted for clustering on wards.

$$0.051 - (0.181)(0.159) - 2(0.001) + 2(0.009)(0.159) - (0.053)(0.114) + 2(0.001)(0.114) - 0.013 = 0.005.$$

This effect is significant, with a standard error of 0.002[16]. In contrast, if the individual settles in an area initially with no other immigrant and with average change in concentration, the effect is 0.031. Thus, at average change in concentration, female immigrants who settle in the majority community are about six times more likely to learn English in any given year than those who settle in a ward with mean immigrant concentration. Females' low participation rate to the labour market[17] may be a possible explanation for the different effects of immigrant concentration on language fluency with respect to males. The effect of ethnic density in the neighbourhood may be lower for males as a consequence of differences in the occupational situation: higher participation may mean more opportunities to interact with the majority population outside the enclave. Furthermore, cultural and traditional habits may restrict women's contact to their minority community more than for men.

3.6 Conclusion

In this chapter we analyse the association between ethnic concentration and the language fluency of ethnic minority immigrants in the UK. At first, we accurately describe the changes in ethnic concentration in the past thirty years and the settlement pattern of immigrants. We show that non-fluent immigrants tend to live in higher concentration areas than fluent individuals.

We develop a simple human capital model to formalise the association between language fluency and ethnic concentration. The model takes account of the self-selection of immigrants into areas according to their initial language skills. The model is also a generalization of the specifications developed in previous literature. In particular, we relax the implicit assumption adopted by previous studies that ethnic concentration is constant over time and we investigate how the level and the change in ethnic concentration affect language fluency.

[16]For the computation of this standard error, see Wooldridge [2003], page 139.

[17]See Chapter 2.

Estimates are for the whole sample and separately for males and females. Overall, there is weak evidence of initial selection based on language fluency. Furthermore, immigrant concentration does not seem to have any significant effect on the language proficiency of male immigrants. For females, the investment in language capital is lower in higher concentration areas and increases faster. Similarly, higher changes in concentration lead to lower accumulation of language skills. Differences in the participation rate and the kind of socio-economic environment where male and female immigrants usually interact may account for the different results. The evidence suggests that it is crucial to analyse the association between language and immigrant concentration separately for men and women.

3.7 Appendix

Table 3.7 shows that both in 1981 and 1991, average concentrations at district and ward level are almost equivalent, whereas the standard deviations are higher at ward level. In our view, this procedure is preferable to using only district concentration along the whole time interval considered in the analysis. This would lead to further loss of information at local level.

Table 3.7: Immigrant Concentration in the UK: Means and Standard Deviations

	1971		1981		1991	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Ward	-	-	0.056	0.066	0.060	0.073
District	0.053	0.048	0.057	0.055	0.061	0.064

Source: Population Census 1971, 1981, 1991.

Part II

Health

Chapter 4

Are Immigrants Healthier than Natives?

Abstract:

This chapter provides an analysis of health inequalities between ethnic minority immigrants and the native population in England. A model of migration decisions is developed that includes health as a determinant of migration. According to the model, immigrants are likely to be positively selected in terms of health if health increases the gain from migration. Two different measures of health are used in the empirical analysis, incidence of chronic disease and self-assessed health. Results show that health differentials clearly depend on the type of health proxy used. Estimates from the chronic disease measure do not reject the hypothesis that immigrants are positively selected. The other health measures provide less conclusive evidence.

4.1 Introduction

A large number of epidemiological and sociological studies in the UK claim that ethnic minorities are less healthy than British born whites. In particular, such findings show that the incidence of some diseases is higher among ethnic minorities than in the white majority population (Lavender [1997], Hawker et al. [1999], Hargreaves [2000], and Raleigh [1997]). Issues of ethnic health inequalities appear to be of great public interest, as several government agencies, such as the London Ethnic Health Network, deal with them. Furthermore, they have great resonance in the media (see, for instance, James [2000], Smith [2001], and Murray [2000]).

Health can be considered an important measure of social integration and well-being of individuals in a society. Therefore, it seems important to accurately evaluate possible health inequalities between the white British majority and other ethnic groups. Previous research on this topic is mainly descriptive and provides analyses of health inequalities where individual factors, such as socio-economic status and standard of living, are often not taken into consideration. Some research is more thorough, but is based mainly on self-assessed measures of health and produces conflicting evidence on how social and economic characteristics affect health. Other research focuses on very specific diseases and is run on very small samples (mostly not representative of the population). In addition, the distinction is rarely made between British and foreign born ethnic minorities. In the Introduction to the thesis, we have shown that it is important to analyse the two groups separately because of the potential cultural and socio-economic differences characterising them.

A few recent studies from the US and Canada analyse the health status of immigrants compared to that of the native population. For Canada, Chen et al. [1996] find that immigrants have a significantly longer life expectancy than Canadian born individuals, other things being equal. For the US, Rumbaut and Weeks [1996] find that Hispanic immigrant women have higher antenatal health outcomes than US born women. Furthermore, Jasso et al. [2002] find that immigrants in the US appear to be significantly less susceptible to chronic illness than US born individuals.

The aim of this study is to provide an economic analysis of the health differentials across ethnic groups, focusing on immigrants. In particular, we want to understand whether migration is positively selective in health. A simple model of migration de-

cision is developed to include health status as a determinant of migration. In this context, health is considered as a component of human capital Grossman [1972]. As such, health can affect earnings capacity and therefore the migration decision of the individual. According to the model, if better health implies a higher gain from migration, then migration is selective in health. That is, immigrants are healthier than the average population of the country of origin. Under the assumptions underlined below, this may also imply that immigrants are healthier than the average population of the destination country.

Empirically, available data allows to test such a claim only indirectly. As it is common to most studies on migration decisions, no or very little information is available on individuals who remain in the source country, which would be adequately comparable to data on immigrants. In this study, the only possible comparison is between immigrants and those born in the destination country. In the case of health, UK-born individuals are not necessarily an appropriate baseline, as there are large differences in the distribution of health between the UK and many sending countries. However, under certain conditions, the health status of UK-born individuals may be used indirectly to measure the health selection of immigrants. We will see that such conditions depend on how the health distributions of the UK and the source countries relate to each other. In particular, if average health in the UK is higher than average health in the source country, UK-born individuals may provide a meaningful comparison group.

The model allows different health paths according to whether the individual stays or migrates. It shows that migration can still be profitable even when the immigrant expects a lower health profile in the host country than in the home country. Furthermore, the model can be extended to groups of individuals, such as families, who jointly decide to migrate. It allows for positive selection to be not necessarily limited to economic migrants, but to family members who can well be tied-movers (such as immigrant women and children). Accordingly, the aim of the empirical section is to determine whether immigrants are positively selected on unobserved health factors. Furthermore, we investigate whether selection applies differently depending on the migration reason. Males and females are analysed separately, as well as individuals who migrated as adults or as children, to allow to distinguish between economic migrants and tied-movers.

Unlike previous studies, this chapter also investigates UK-born ethnic minorities. They are mainly second generation immigrants, i.e. children of ethnic minorities born

abroad[1]. Extensive sociological literature shows how UK-born ethnic minorities retain social and cultural habits and traditions inherited from their countries of origin, as well as adopt British culture (Modood and Berthoud [1997]). Furthermore, some inter-generational correlation in health may persist. All these factors may determine health differences between ethnic minority natives on one side and ethnic minority immigrants and white natives on the other. Therefore, it seems important to analyse UK-born minorities separately from both first generation immigrants and UK-born whites.

The analysis is based on two data sets, which provide unique information on ethnic minorities in England: the Fourth National Survey on Ethnic Minorities (FNSEM) and the 1999 Health Survey for England (HSE). Both contain a "boost" sample of ethnic minorities and a random sample of white natives, and provide, in addition to social and economic information on respondents, information specific to their ethnic and migration status as well as to their health condition. It is worth stressing that the HSE is a survey which aims to determine the health of the English population. In 1999 it focused on ethnic minorities for the first time. The two surveys were conducted five years apart and some researchers contributed to both. The resulting similarities in categorising ethnic minorities and defining health are useful for the research.

Health, like ability, is a difficult concept to measure. Available indices of health are usually discrete and imperfect proxies, such as mortality rates, self-assessed health, and incidence of chronic disease. There are three measures of health available in the FNSEM and HSE. To render the analysis more complete, we use all three proxies to measure health status, as we regard them as complementary and useful to provide evidence of the potential differences in the more general wellbeing of ethnic groups.

The first measure, available in the HSE, is self-reported assessment of global health, or self-assessed health status (SAHS). Bound and Burkhauser [1999] argue that the measure is a subjective one. It does not necessarily reflect the actual physical health status of the individual, but may uncover other personal conditions. However, it is largely used in the literature of health economics (e.g. Smith [1999], Kennedy et al. [1998]), as it is believed to be a good predictor of mortality and morbidity (e.g. McCallum et al. [1994]). The second measure, available in the FNSEM, is a less general self-assessment of health. Respondents are asked to compare their recent health status with that of people of the same age. For brevity, we define it as self-assessed relative health status

[1]See the Introduction to the thesis for a survey on migration trends in the UK.

(SARHS). This is also a subjective measure, although it is more specific than the SAHS and individuals may provide more objective judgement. For the last measure, available in both surveys, individuals self-report the presence of chronic disease (CD). This proxy is usually recognised as a more objective proxy of health (Bound [1991])[2]. However, we worry that this variable would not capture important aspects of the more general well-bing of individuals.

Results indicate that, according to the CD health measure, all immigrant groups enjoy better health than UK born individuals. According to this measure, findings support the hypothesis that immigrants are positively selected in terms of health. Furthermore, in the case of migration of entire families, good health of other family members appears to be also a determinant in the migration decision. Evidence from the SARHS and SAHS measures is less clear-cut, as it seems to depend on the specific measure used. According to the SARHS, at arrival immigrants seem to be healthier than natives. On the other hand, estimates from the SAHS measure do not suggest any significant health advantage of immigrants with respect to natives, except for the white group. Therefore, there is evidence of migration being selective mainly in terms of prevalence rates of chronic conditions.

When comparing health conditions at average years since migration, it is apparent that, according to all health measures, health differentials between immigrants and natives decrease significantly. In other words, duration of stay in the UK seems to have a negative effect on the health of immigrants. This may be due to faster depreciation of health for immigrants than for native, as well as changes in cohort quality. Given the cross-sectional nature of the data sets, we cannot distinguish the two factors. In particular, at average years since migration, the health advantage of immigrants is lower, and, according to the CD measure, it still persists and is significant. In contrast, the situation depicted by both the SARHS and SAHS measures has reversed, with immigrants now appearing less healthy than natives. The UK is likely to provide better health care and higher living standards than the country of origin. However, migration may also have a negative effect on aspects of health only captured by self-assessed health, and possibly

[2]Chronic illnesses are often diagnosed by a doctor. It is sometimes argued that immigrants tend to make use of health services less than natives (e.g., due to cultural differences or poor proficiency in English), therefore the incidence of chronic diseases may be systematically underestimated for these individuals. Some evidence will be provided in Section 4.3.1, which suggests this not to be the case.

partly on the predisposition to longstanding illness. Migrants may suffer from the loss of family networks, or from a new potentially hostile environment and this may have repercussions mostly on those aspects of health which are not reflected by any chronic condition.

To conclude, the chapter provides a theoretical background to understand whether immigrants may be positively selected among the population of the source country. Immigrants' health is then compared to the health of the population of the destination country, as conditions are provided under which such a comparison is meaningful. Results show that at arrival, immigrants appear to be healthier than natives. However, the results show some degree of sensitivity to the health measure used. This issue is common in the literature of health economics, as health is a difficult concept to measure and only imperfect proxies are usually available.

The remainder of the chapter is organised as follows. Section 4.2 proposes a simple model of selection. Section 4.3 discusses health measures, health distributions in sending and receiving countries, and introduces the empirical framework adopted for the estimation. Section 4.4 describes the data sets used and outlines some summary statistics of interest for the analysis. Sections 4.5 presents the empirical results and Section 4.6 concludes.

4.2 Theoretical Framework

The following approach is motivated by standard models of migration decisions for the individual and for the family, as suggested by Borjas [1987] and Mincer [1978]. We extend this type of model, incorporating health as a determinant of migration and analyse how health status can affect the migration decisions of individuals and families. The behavioural assumption underlying the family decision process is that migration decisions are motivated by the maximisation of family utility.

4.2.1 The Individual's Decision

Consider a two-period setting. In period 1 the individual decides whether to migrate to country D or to stay in the origin country O and his health status is measured by the health index h_1. In period 2, the individual's income in country I, where $I = D, O$ is $y_2^I(h_1)$. Health is a component of human capital, as it can affect wages in various

ways. Poor health may reduce productivity, resulting in lower wages. Furthermore, the employer's cost of employing a worker in poor health may be reflected in lower wages, and unhealthy individuals may be subject to discrimination (Currie and Madrian [1999]). Finally, poor health may also reduce abilities which are crucial for newly arrived economic migrants, such as the ability to learn the language and to search for jobs. Accordingly, we assume that income in period 2 is increasing in h_1.

The individual's health status in period 2 is h_2^D if he migrates to country D, and h_2^O if he stays in the origin country O. Therefore, we allow two potentially different health trajectories for the origin and destination countries. In particular, $h_2^I = h_2(h_1, y_2^I(h_1), X^I)$, where X^I reflect the individual's (often unobserved) quality of life in country I. We assume that $\frac{\partial h_2}{\partial h_1} > 0$ and $\frac{\partial h_2}{\partial y_2} > 0$. However, we make no assumption on the sign of $\frac{\partial h_2}{\partial X}$, as X comprises factors which may affect health both positively and negatively[3].

The expected utility from living in country I is then defined as

$$U^I = U(h_1, h_2^I(h_1, y_2^I(h_1), X^I), y_2^I(h_1)) \tag{4.1}$$

where $\frac{\partial U}{\partial h_1} > 0$, $\frac{\partial U}{\partial h_2} > 0$, and $\frac{\partial U}{\partial y_2} > 0$.

Given fixed migration costs C, the individual migrates if

$$U^D - U^O - C > 0 \tag{4.2}$$

In this framework, we can distinguish three cases. In the first and simplest case, $h_2^D = h_2^O$. Then, the utility gain derives only from the gain in income. Necessary condition for (4.2) to hold is that $y_2^D(h_1) > y_2^O(h_1)$. In other words, the returns to health are higher in the destination country than in the origin country. In addition, if the rate of return to health in D is higher than in O (i.e. $\frac{\partial \Delta y_2}{\partial h_1} > 0$, where $\Delta y_2 = y_2^D - y_2^O$), migration is selective in terms of health. Accordingly, healthier individuals would have higher incentive to migrate.

[3]Positive determinants of health may be access to better health care and higher living standards in the host country. On the other hand, immigrants may face uprooting from family networks, and potential hostile behaviour of residents leading to social exclusion. These problems may negatively affect immigrants' health (see, for example, evidence from Benzeval et al. [1992], and Karlsen and Nazroo [2002]).

In the second case, $h_2^D < h_2^O$. Individuals migrate only if the utility gain from income compensates the utility loss deriving from lower health in the host country. As in the first case, migration is selective if the difference in rate of returns to health is increasing in health. As health in the host country deteriorates, incentives to migrate will be lower and the average migrant will be healthier than in the first case.

Finally, if $h_2^D > h_2^O$, the incentive to migrate will be higher. Therefore, more individuals will migrate, but the average migrant will be less healthy that in the previous cases.

4.2.2 The Family Decision

In the individual's decision model, it is implicitly assumed that the subjects of the analysis are economic migrants. However, when considering migration in the family context, it is not obvious that all components of the family are economic migrants. In particular, children under working age and non-working spouses are "tied-movers". Since they do not work when they arrive in the UK, their health condition may not determine their migration decision. However, even in a family context, it is likely that all members' health status affects the migration decision of the household as a whole. The ill-health of one individual, such as a child or the spouse, may decrease the net gain of migration for the family.

In the case of the UK, in Chapter 2 of the thesis, we provide evidence that a large part of female ethnic minority immigrants do not participate in the labour market. If the migration decision is taken considering the aggregate interest of all family members, incentives related to health may differ from the individual case. Therefore, it seems appropriate to extend the previous analysis to the family context.

Consider a family of potential immigrants consisting of two individuals, a husband (H) and a wife (W)[4]. Under the assumption that in the household decisions are taken to maximise the joint utility of the family, a family migrates if the sum of the individual gains is positive, or $U^H + U^W > 0$.

In this context, migration is selective for both members when both U^H and U^W are higher for healthier individuals. However, it may still be optimal for a family to migrate even when the gain of one of the members is zero or negative, provided that the other person's gain compensates this loss. For instance, the gain for one member is zero when

[4]This framework can be extended to include more family components.

the husband works and the wife does not. Furthermore, if the wife is ill, her own gain may be negative. This may happen, for example, if the health is bad enough to require costly care. Although in the UK about 85% of health care is publicly financed (Propper [2000]), there may be other indirect costs related to ill-health. These costs may be higher than in the home country, due to the lack of family networks and of proper knowledge of the health and welfare system.

It is not straightforward to determine how the health of non-working members affects the migration decision of the family. It seems plausible to assume that, on average, the costs of ill-health of any family member would reduce the net gain of migration and therefore that, on average, husband and wife will have similar health conditions. However, selection in terms of health may be attenuated when migration decision is taken at family level[5], as the costs or missed earnings for the non-working individual may be compensated by the gain of the working member. The empirical section aims to clarify this issue and to determine whether there are different degrees of selection between economic migrants and tied-movers.

4.3 Empirical Framework

4.3.1 Health Measures

Measuring health differentials is difficult, because people's health status is most often not directly observable. Self-assessments of general health or self-assessment of specific illnesses are among common measures of health provided by microeconomic surveys. As such, they are widely used in the literature of health economics as proxies for health (eg, Smith [1999], Deaton [1999]), although they are imperfect and discrete indices of health. Furthermore, some researchers argue that self-assessment of general health is a very subjective measure (Bound and Burkhauser [1999]), as it potentially reflects the degree of satisfaction enjoyed by the respondent at work or generally in life. In contrast, the specificity of the question on the presence of chronic illnesses is believed to make this measure a more objective proxy.

The importance of the issue of subjectivity in this context, where health is the dependent variable, arises because health differentials are analysed between groups whose

[5]Borjas and Bronars [1991] arrive to a similar conclusion in terms of selection in ability.

health evaluation may differ systematically. In other chapters of the thesis, we show that ethnic minority individuals are more likely to be exposed to socio-economic alienation than white natives, in the form of, for example, employment and earnings disadvantages, and racial harassment. This may be correlated to their health evaluation, leading to systematic differences in self-assessed health among groups. Furthermore, health differences may appear in form of higher disposition to more common diseases not necessarily classifiable as chronic conditions.

When analysing the incidence of chronic illness among different groups in the population, a further issue may arise. The degree of contact with the health service may vary among ethnic groups, if it is influenced by cultural differences (Jasso et al. [2002]). For instance, immigrants who are not fluent in English may find it difficult to communicate with medical staff and therefore refrain from doing so. This may induce immigrants to under-report chronic conditions. With regard to communication issues due to language differences, it must be observed that the British National Health Service has been offering special support to individuals of different ethnic or religious background or not fluent in English[6]. Furthermore, a large number of physicians and medical staff are of ethnic minority origin[7]. Unconditional estimates provided in Section 4.5 show that immigrants tend to self-assess their health worse than white natives, in spite of reporting analogous or lower incidence of longstanding illness. To shed more light on the issue, we have used the information on the frequency of recent visits to the GP (General Practitioner) for personal reasons[8]. According to simple regressions, in both surveys, ethnic minority immigrants and natives seem to visit their GP significantly more frequently than white natives, and this differential increases further when controlling for age. Although not necessarily conclusive, this evidence suggests that immigrants do not under-report the incidence of chronic disease[9].

In addition to responding to questions regarding specific longstanding illnesses, respondents in the FNSEM and HSE are asked how they evaluate their own health via

[6]The surveys themselves were run in the respondent's own language, if problems with the English language were identified by the interviewers.

[7]For instance, in 1996, 28% of doctors in the UK and 16% of consultants are from ethnic minority groups. In addition, 20% of Indians, 50% of Pakistanis, and 59% of Bangladeshis communicate with their GP in one or other of the Asian languages (Alexander [1999]).

[8]The FNSEM and the HSE contain information on visits to the GP (for personal issues) in the month and the two weeks previous the interview, respectively.

[9]An analysis of the immigrants' demand for health may be an interesting topic for future research.

two different questions. More specifically, the question on general health in the FNSEM is (what we refer to as self-assessed relative health status, or SARHS) : "Please think back over the last 12 months about how your own health has been. Compared to people of your own age, would you say that your health on the whole has been excellent, good, fair, poor, very poor?". Whereas the question in the HSE is (what we refer to as self-assessed health status, or SAHS): "How is your health in general? Would you say it is: very good, good, fair, bad, very bad?". The two questions have a different degree of specificity, and they may not lead to equivalent outcomes.

Health differentials are estimated on all three available measures. Different results may arise, depending on the proxy used, as they are not equivalent. The debate is still open on which measure may be best. Rather, it seems appropriate to us to use all information available to provide a wider picture. Potential differences in the results may provide a better insight on the actual health status and the degree of social integration of immigrants.

4.3.2 Health Distributions

This study compares the health status of immigrants to that of UK natives to determine whether immigrants are positively selected in terms of health. This is not the ideal way to proceed, as the distribution of health in the UK is likely to differ substantially from the distribution in many of the source countries considered in our sample. As the focus of the HSE and the FNSEM is on ethnic minorities, the source countries of interest are the Caribbean Commonwealth, Bangladesh, India, Pakistan, and China. In this section we show that it is possible to compare the health of immigrants to the health of UK-born individuals if certain conditions are satisfied.

Consider average health in the home country and average health in the host country. Suppose initially that average health in the UK is lower than average health in the home country. In this case, selection would not be testable, because any individual from the home country is likely to be healthier than the average British, regardless of whether s/he is a migrant.

Suppose now that average health in the UK is higher than average health in the home country. Then we can distinguish two cases: a) the entire distribution of health in the source country is below the average health in the host country (the UK); b) part of the health distribution in the source country lies above UK average health, as depicted in

Figure 4.1. In case a), even if there were positive selection of immigrants, it would not be possible to test it, as all immigrants would be less healthy than UK-born individuals. In contrast, in case b), selection would be testable through a comparison between UK-born and immigrants. Under the hypothesis of positive selection in health, immigrants will belong to the highest percentile(s) of the sending country distribution. If the highest percentiles of the distribution in the sending country correspond to percentiles which are higher than the average of the health distribution in the UK, immigrants will be on average healthier than UK-born individuals.

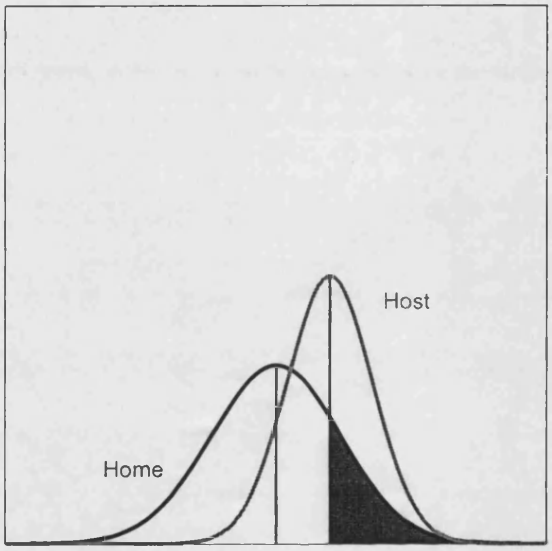


Figure 4.1: Hypothetical Health Distributions in Home and Host Countries

Therefore, we need to know what the relative position of the health distributions of the source countries is with respect to the health distribution in the UK. To derive this information, we refer to data available from the World Health Organization (WHO) on their basic indicators of health, mortality and life expectancy. Table 4.1 reports the probability of dying for males aged between 15 and 59 [10], and life expectancy at birth for males in countries of interest in 1999.

Table 4.1 gives a rough indication about differences in average living standards in most countries where ethnic minorities come from, with respect to the UK nowadays. According to these measures, it is apparent that average health in the UK is higher than

[10]Information on mortality rate in the WHO Report is available either for the whole population or limited to the age interval 15-59. To focus on individuals of working age, the second measure is used here.

Table 4.1: Male Probability of Dying and Life Expectancy in Sending Countries and the UK

Country	Prob. of Dying (per 1000) age 15-59	Life Expectancy
Bangladesh	300	57.5
China	170	68.1
India	275	59.6
Jamaica	135	75.2
Pakistan	194	62.6
UK	111	74.7

Notes:
Main countries of origin for immigrants in the UK.
Source: World Health Report 2000, World Health Organization.

in source countries. Although in our sample years of arrival are mainly prior to 1999, most probably such differences were even larger two or three decades ago, when the large majority of ethnic minority immigrants arrived.

Hence, evidence supports the assumption that average health in the UK is higher than in the countries of origin of ethnic minority immigrants. This evidence, however, does not indicate how distant the health distribution in the UK is with respect to the health distribution in the other countries. Marmot et al. [1984] provide some evidence about differences in health between immigrants and natives in the UK. Their estimates show that the mortality rates of most immigrant groups were lower than for UK born, notwithstanding the higher corresponding rates in their home countries. This, together with the data analysis below, suggests that selection may be testable.

4.3.3 Empirical Method

In order to analyse the health differentials between immigrants and natives, and to understand the factors that determine such differentials, an empirical model must be formulated. The available health measures (SAHS, SARHS, or CD) are regressed on a set of observable characteristics: age, years of residence, education and several immigrant status dummies. As white natives are the base category, the coefficients of the immigrant dummies reflect the health differentials between the corresponding immigrant group and UK born whites unexplained by the observable characteristics.

As previously mentioned, there is evidence that, among ethnic minorities, men are

economic migrants, whereas women are mainly tied-movers. Furthermore, in all immigrant groups, young children are necessarily tied-movers, as they cannot work at the time of their arrival[11]. For this reason, men may be expected to be positively selected in terms of health, whereas women and young children are not necessarily so. Their migration decisions are bound to their husbands' or fathers' migration decisions, and their potential loss (in case of bad health) deriving from migration may be smaller than their husbands' or fathers' net gains. However, according to the model in Section 4.2, the possibility that the health of each family member is a determinant in the decision of a whole family to migrate cannot be ruled out and needs to be tested. Therefore, it is crucial to make a distinction between individuals who migrate as children and individuals who migrate as adults. Define as "adult" those immigrants who arrived in the UK after the age of 16 and as "child" those immigrants who arrived before the age of 16. Health differentials should be indicative of whether individuals who migrated as children are also positively selected.

The health of ethnic minority natives is also analysed, to be compared to first generation immigrants and UK born whites.

Furthermore, where possible, we analyse white immigrants separately from ethnic minority immigrants. As we have seen in the Introduction, the two groups perform differently in the labour market, and they come from countries characterised by different living standards and cultural backgrounds[12]. Both facts may be correlated to health status. Finally, white female immigrants are less likely to be tied-movers than ethnic minority immigrants, as they have similar participation rates to the labour market as white female natives (see Dustmann et al. [2002]).[13]

As mentioned above, education is included in the general specification. In this study, no attempt is made to address the endogeneity of the education variable, as the information available in the data sets is not adequate for this type of analysis. Hence, the education coefficients should be interpreted in terms of association rather than causal effects.

Age and years since migration are further explanatory variables. Besides being an

[11] In the Introduction, we have seen that 34% of immigrants arrived in the UK before the age of 16.

[12] In the HSE, more than 60% of white immigrants were born in Europe, whereas most ethnic minority immigrants were born in the West Indies or in Asia.

[13] The distinction between "adult" and "child" immigrant is omitted for the white group to maintain reliable sample size.

obvious determinant of health, age plays a crucial role in the determination of health differentials, since the age structure varies widely between the groups considered (see Section 4.4). When both age and years since migration are included in the regression, the coefficient of years since migration can be interpreted as the combined effect of the difference in the time rate of health deterioration between immigrants and natives and potential cohort effects. As both data sets are cross-sections, it is not possible to identify these effects separately. In addition, the coefficients on the immigrant dummies should be interpreted as the health differential between white natives and the relative immigrant group at their time of arrival. In contrast, when only age is included in the regression, the health differentials are evaluated at average time of residence in the UK.

Although health status has a continuous distribution, its available measures are discrete. The presence of chronic illness is a binary variable, with 0 representing the presence of an illness, and 1 its absence. Both measures of self-assessed global health are on a scale from one (excellent or vary good) to five (very bad). As there seems to be little loss of information in simplifying the analysis, we transform the index in a dichotomous variable, taking value 1 for good or very good reported health and 0 otherwise[14]. Discrete response analysis, such as probit, seems appropriate.

Consider a generic health outcome function, separately for males (M) and females (F), where h_i^* is the underlying latent index of health for individual i. We use the general specification

$$\begin{aligned} h_{iM}^* &= \mathbf{X}_i'\alpha_M + \beta_M YSM_i + \gamma_{1M} D_i^A + \gamma_{2M} D_i^C + \gamma_{3M} D_i^N + \gamma_{4M} D_i^W + u_i \\ h_{iF}^* &= \mathbf{X}_i'\alpha_F + \beta_F YSM_i + \gamma_{1F} D_i^A + \gamma_{2F} D_i^C + \gamma_{3F} D_i^N + \gamma_{4F} D_i^W + v_i \end{aligned} \tag{4.3}$$

The observed binary indexes are $h_i = 1[h_i^* > 0]$, where $1[\cdot]$ is the indicator function[15].

Equations are estimated separately by gender to allow for possible systematic differences between males and females and in the effects of regressors. In both specifications, $\mathbf{X}_i$ is a vector of observable factors associated with health, such as age, education, and household income, and YSM_i is years since migration. D_i^A, D_i^C and D_i^N are dummies indicating whether the individual is an "adult" minority immigrant, a "child" minority

[14]Results from a binary probit do not substantially differ from those of an ordered probit. In particular, the results from the "fair" category are most often not significantly different from those relevant to the "bad" or "very bad" categories.

[15]Subscripts M and F are neglected for simplicity of notation.

immigrant, or a minority native, respectively; and D_i^W is a dummy denoting white immigrant individuals. Hence, white natives are the reference group. Finally, u_i and v_i are idiosyncratic error terms.

For both genders, the coefficients on the D_i^J dummies (J={A, C, N, W}) each reflects the health differential between one of the four types of immigrants considered, respectively, and UK born white individuals. As discussed above, individuals are characterised by differences in education, age and years of residence. When controlling for observable characteristics, any remaining differential is due to unobservable characteristics, which is captured by the immigrant dummies.

If immigrants are positively self-selected on the basis of unobservables, and health is correctly measured, estimates will be expected to provide significant (positive) coefficients of the immigrants' dummies. Further discussion follows regarding health measures and their reliability.

4.4 The Data

The Fourth National Survey for Ethnic Minorities (FNSEM) is a cross-section survey collected between 1993 and 1994. It consists of a main sample of respondents belonging to ethnic minorities, and a reference sample of individuals belonging to the white majority population. The FNSEM contains information on individual characteristics of the interviewees, information specific to their ethnic and migration status and on their health and health habits.

The Health Survey for England (HSE) is an annual cross-section survey which aims to monitor the health trends of England. For the first time in 1999, separate attention was given to the health of ethnic minorities. For this purpose, information was collected from a general population sample and a ethnic minority "boost" sample. Its information structure on individual characteristics is similar to the FNSEM, and it contains a wide variety of information on health. Furthermore, the HSE contains a small sub-sample of white immigrants.

The two surveys have different sample designs. In the FNSEM, 59% of the ethnic minority sample was selected from wards were, according to the 1991 Census data, ethnic minorities represent at least 10 per cent of the whole population. About 38% were selected from areas with ethnic concentration between 1 and 5 per cent and the

rest in areas with concentration of less than 1 per cent. In contrast, the ethnic "boost" sample of the HSE was selected choosing addresses randomly from the Postcode Address File (PAF) and selecting only respondents from ethnic minority groups.

As mentioned above, both data sets contain information on gender, age, marital status, number of children, years since migration, country of origin and ethnicity. Both also contain information on education and household income. In the FNSEM the questions on health are based on the personal assessment of the interviewee. The HSE, on the other hand, contains self-evaluated information on health, as well as more detailed information collected by a nurse via thorough medical exams. Such technical information, however, was only collected for individuals who, in the main survey, declared to suffer from some heart or coronary disease. Therefore, the data would add little support to obtain truly objective measures of health.

To analyse health outcomes we use two sets of questions, which are contained in both surveys and are widely used in the evaluation of health in health economics literature (see Goldman and Lakdawalla [2001], and Bound and Burkhauser [1999]). The first question regards the incidence of chronic or long-standing illness. Both in the FNSEM and the HSE, individuals report whether they "have any long-standing illness, disability or infirmity". The second question concerns how individuals evaluate their health. In the FNSEM respondents report how they see their health, in the 12 months previous the interview, compared to people of their own age, on a scale from 1 (excellent) to 5 (very poor). In the HSE individuals simply report how they evaluate their own health, again on a scale from 1 (very good) to 5 (very bad).

Table 4.2 provides means and standard deviations for the variables of interest relative to the sub-samples under analysis. To begin, the gender distribution is similar in both surveys, with females being slightly over-represented in most sub-samples. The average age of "adult" ethnic minority immigrants is about 47 and corresponds to white natives' average age. In contrast, "child" immigrants are about 15 years younger and UK born minorities are 22 years younger than white natives. Finally, white immigrants are slightly older (3 years).[16]

The average duration of stay in the UK is about 21 and 23 years for immigrants

[16]The FNSEM interviewed individuals aged 16 or more, whereas the HSE contains information on respondents of any age. To render the two data sets compatible, the analysis is restricted to individuals aged over 16.

Table 4.2: FNSEM and HSE - Sample Characteristics

Variables	FNSEM								HSE									
	Imm.(≥16)		Imm.(<16)		Min. Nat.		White Natives		Imm.(≥16)		Imm.(<16)		Min. Nat.		White Imm.		White Natives	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Female	0.52	0.50	0.44	0.50	0.54	0.50	0.59	0.50	0.53	0.50	0.50	0.50	0.55	0.50	0.59	0.49	0.54	0.50
Age	47.19	14.44	32.18	8.02	24.70	6.35	47.60	19.08	47.92	14.73	32.51	10.15	26.09	8.37	50.56	18.25	47.23	18.35
Residence Years	20.64	10.91	23.03	7.32	–	–	–	–	21.79	13.09	24.41	9.19	–	–	32.84	18.77	–	
Married	0.81	0.39	0.68	0.46	0.35	0.47	0.58	0.49	0.78	0.41	0.63	0.48	0.34	0.47	0.66	0.47	0.65	0.47
Degree	0.13	0.34	0.09	0.29	0.09	0.28	0.08	0.27	0.14	0.34	0.14	0.34	0.14	0.34	0.20	0.40	0.13	0.33
A-level	0.09	0.28	0.14	0.35	0.17	0.38	0.16	0.36	0.11	0.31	0.23	0.42	0.30	0.46	0.19	0.39	0.22	0.41
O-level	0.17	0.37	0.35	0.47	0.50	0.50	0.29	0.45	0.18	0.38	0.32	0.46	0.39	0.48	0.26	0.44	0.34	0.47
Indian	0.27	0.44	0.18	0.38	0.25	0.43	–	–	0.17	0.37	0.13	0.33	0.40	0.39	–	–	–	–
Black Caribbean	0.17	0.37	0.18	0.38	0.43	0.49	–	–	0.16	0.36	0.14	0.34	0.19	0.48	–	–	–	–
African Asian	0.15	0.35	0.19	0.39	0.06	0.23	–	–	0.06	0.24	0.10	0.30	0.005	0.07	–	–	–	–
Pakistani	0.23	0.42	0.25	0.43	0.19	0.39	–	–	0.21	0.40	0.22	0.41	0.25	0.43	–	–	–	–
Bangladeshi	0.14	0.33	0.15	0.35	0.03	0.18	–	–	0.22	0.41	0.30	0.46	0.09	0.27	–	–	–	–
Chinese	0.05	0.21	0.04	0.19	0.02	0.15	–	–	0.16	0.36	0.10	0.29	0.07	0.25	–	–	–	–
No. Obs	2854		1210		1161		2978		2915		1110		1669		633		7522	

Note: Imm.(≥16) are immigrants whose age at migration was years or older. Imm.(<16) are immigrants who migrated before turning 16.

arrived after and before the age of 16, respectively, whereas white immigrants have a much longer stay of 33 years.

In both surveys, a higher percentage of "adult" immigrants are married (81% for the FNSEM and 78% for the HSE) than any other group; and a smaller percentage of minority natives are married. Such evidence is certainly due to the age differences between minority natives and the other groups.

With regard to the education variables, the two surveys differ in the distribution of education qualifications among the groups considered. About the same number of "adult" ethnic minority immigrants have a degree in both surveys. However, the number of "child" immigrants, minority natives and British born whites in the FNSEM having a degree (9%, 9% and 8%, respectively) is lower than in the HSE (14%, 14% and 13%, respectively). Furthermore, in the HSE, there are 20% white immigrants with a degree. Similar differences exist for the other two education variables, measuring A-levels and O-levels or vocational.

Finally, the different ethnic composition within both the immigrant and native groups may well depend on the different sampling procedure that characterised the two survey outlined at the beginning of this Section. The immigrant samples contain about the same percentage of Black Caribbeans and Pakistanis (about 17% and 24% respectively), but different distributions of the other groups. With regard to the non-white British born samples, the ethnic distributions are dissimilar for all groups. The information available in the HSE on the composition of the white immigrants' sample is not very detailed: about 51% come from the republic of Ireland; some 18% from other European countries; some 6% from the African continent; about 2% from India. The origin of the remaining respondents is not specified.

4.5 Results

In the analysis that follows, we report the health differentials for each health variable. We use three specifications. In the first specification, such differentials are estimated without controlling for observable characteristics and to understand the extent to which observable characteristics account for raw differentials. In the second specification, observable characteristics such as education, age, age square, years since migration (YSM), and YSM square are included. Therefore, differentials are estimated for individuals with

the same characteristics and, in the case of immigrants, at time of arrival. In the third specification, estimations are run excluding years since migration, to obtain health differentials at average time of residence in the UK. Reported results are marginal effects evaluated at sample means.

Immigrants are divided into several categories, to allow for possible differences in the cultural background and living standards enjoyed in the countries of origin. Such categories are: ethnic minority immigrants who migrated at or after the age of 16 and ethnic minority immigrants who migrated before the age of 16; second generation ethnic minority immigrants (or minority natives); and white immigrants (this category being only available from the HSE)[17]. This categorisation is possibly imperfect, as the definition of ethnic minority applies to individuals with very different cultures and backgrounds. However, due to limited sample size, a further break down of the samples would seriously reduce the robustness of the estimates.

Table 4.3 and Table 4.4 report health differentials according to the chronic disease measure and the SARHS (available from the FNSEM), and the SAHS measures (available from the HSE), respectively. Results for different specifications are shown in both tables. Finally, results from the FNSEM are on the left side and results from the HSE are on the right side of the tables.[18]

Results in the upper panel of Table 4.3 show that the incidence of chronic disease appears to be lower for "child" immigrants and minority natives in both survey. Among "adult" immigrants, only females in the FNSEM show a significantly lower probability of having a disease. The health status of white immigrants does not seem to differ from the one of their UK born counterpart.

When controlling for observables, the picture changes substantially. Part of the above health differentials are explained by differences in age and education characterising the groups of interest, underlined in Section 4.4. Minority natives appear now about as healthy as white natives, but for a slightly significant advantage among men. "Adult" immigrants and "child" immigrants appear to be healthier than white natives to a similar extent. It is interesting to see how the results from both surveys are similar in magnitude and significance. White immigrants also appear to be healthier than natives and as healthy as minority immigrants.

[17] As noted before, no distinction is made between "adult" and "child" immigrant inside this category.

[18] The full results from these specifications can be found in the Appendix.

Table 4.6 in the Appendix reports full estimates. The table shows that the regressors YSM and YSM square play a crucial role in the determination of health differentials. In particular, YSM has a significantly positive effect on the probability of having a longstanding illness. One possible interpretation is that immigrants' health deteriorates at a faster rate than natives' health. However, as the data are both cross-section, we cannot exclude the possibility of cohort effects.

When controlling for household income, health differentials between male ethnic minority immigrants and white natives appear to be even larger. However, the same increase in the health differential coefficients is not as noticeable for female immigrants. This evidence seems to suggest that disadvantaged socioeconomic conditions significantly affect the health of male immigrants[19], but has no significant effect on the health of female immigrants.

It is interesting to look at such differentials at average duration of migration, as reported in the lower panel of Table 4.3. At average years since migration, the health advantage of all immigrant groups is lower, according to both surveys. According to the FNSEM, men and women belonging to the "child" and "adult" immigrant groups are still more likely to be healthier than natives, although this probability is smaller than when controlling for YSM. According to the HSE, a similar advantage remains, for minority immigrant males. Although it is a possibility accounted for in the theoretical part, the effect of different health trajectories on health selection is not measurable, as we cannot disentangle cohort effects from rate of deterioration of health specific to the immigrant population. Interestingly, the health differentials at average years since migration do not change significantly when controlling for household income.

The evidence provided by these results supports the hypothesis we wanted to test: migration is selective in terms of health. Immigrants appear to be healthier than natives. These results are consistent with evidence from the US (Jasso et al. [2002]), Canada (Chen et al. [1996]), and Germany (Razum et al. [2000]), which are based on the same health measure.

In addition, positive selection seems to apply equally to individuals who are not necessarily economic migrants, such as women and individuals who migrated as children. In terms of the theoretical model, this suggests that the health of all family members

[19] As ethnic minority immigrants are, on average, economically disadvantaged with respect to white natives (see Modood and Berthoud [1997] and Dustmann et al. [2002]).

weighs similarly when the family is confronted with migration decisions.

Results on minority natives indicate that there may be some degree of intergenerational transmission of health between first and second generation immigrants. Such evidence, however, is not conclusive, as the coefficients for the dummies are not significant for women and not strongly significant for men.

Table 4.4 reports estimates from the self-assessed indices of health. It is important to remember that the two health proxies available in the FNSEM and the HSE are different. In particular, in the FNSEM, when asked about their health status, interviewees are specifically asked to compare their health with that of individuals of their same age (what we called SARHS). On the other hand, in the HSE, there is no information about whom respondents may compare their health health status with, or if they do it at all (SAHS).

When looking at the raw differentials, for both health measures, "adult" immigrants are more likely to evaluate their health status as fair, bad or very bad than white natives. In contrast, minority natives seem to consider their health good or very good more often than white natives. In addition, according to the FNSEM, the self-assessed health of "child" immigrant women is significantly higher than white native women, but according to the HSE there is no significant difference.

Once we control for observable characteristics, including YSM, the above picture changes dramatically. Results from the SARHS in the FNSEM indicate that immigrants are more likely to evaluate their health as good or very good than natives (white and minority), and "child" immigrants more so. In contrast, results from the SAHS show no significant differences between male minority immigrants and white natives. Furthermore, white immigrants appear to be on average healthier than the other groups. However, "adult" immigrant women and minority native women seem to evaluate their health similarly and significantly worse than other groups.

Similarly to the previous results for the chronic disease measure, controlling for household income leads to higher health differentials. For the FNSEM, in particular, the increase is dramatic among male immigrants.

When YSM is excluded from the regression, all minority immigrants groups (except "child" immigrant women for the FNSEM) are more likely to assess their health as fair or very poor than natives.

Table 4.3: Health Differentials on Chronic Disease Measure (base group: Whites)

	Unconditional						
	FNSEM			HSE			
	Imm. (≥16)	Imm. (<16)	Min. Nat.	Imm. (≥16)	Imm. (<16)	Min. Nat.	White Imm
Males	-0.022	-0.181	-0.176	-0.002	-0.187	-0.215	0.001
	(1.38)	(9.63)	(8.73)	(0.10)	(8.39)	(10.93)	(0.05)
Females	-0.085	-0.193	-0.192	-0.021	-0.158	-0.204	0.002
	(5.73)	(9.91)	(10.37)	(1.43)	(7.18)	(11.38)	(0.07)
	Conditioning on age, education, and YSM						
	FNSEM			HSE			
	Imm. (≥16)	Imm. (<16)	Min. Nat.	Imm. (≥16)	Imm. (<16)	Min. Nat.	White Imm
Males	-0.220	-0.238	-0.056	-0.207	-0.261	-0.061	-0.192
	(4.60)	(5.21)	(2.03)	(5.78)	(6.26)	(2.53)	(4.46)
Females	-0.204	-0.223	-0.037	-0.176	-0.201	-0.031	-0.182
	(4.80)	(5.24)	(1.43)	(5.48)	(4.98)	(1.44)	(4.61)
	Conditioning on age, education, household income and YSM						
	FNSEM			HSE			
	Imm. (≥16)	Imm. (<16)	Min. Nat.	Imm. (≥16)	Imm. (<16)	Min. Nat.	White Imm
Males	-0.279	-0.277	-0.075	-0.275	-0.295	-0.083	-0.213
	(5.04)	(5.47)	(2.37)	(6.66)	(6.13)	(2.92)	(4.33)
Females	-0.223	-0.231	-0.025	-0.237	-0.251	-0.035	-0.208
	(4.57)	(4.61)	(0.84)	(6.44)	(5.55)	(1.46)	(4.65)
	Conditioning on age and education						
	FNSEM			HSE			
	Imm. (≥16)	Imm. (<16)	Min. Nat.	Imm. (≥16)	Imm. (<16)	Min. Nat.	White Imm
Males	-0.054	-0.097	-0.017	-0.047	-0.091	-0.041	-0.037
	(3.23)	(4.43)	(0.61)	(2.80)	(3.70)	(1.68)	(1.14)
Females	-0.075	-0.092	-0.003	-0.031	-0.025	-0.014	-0.020
	(4.79)	(3.98)	(0.13)	(1.93)	(1.03)	(0.63)	(0.74)
	Conditioning on age, education, household income						
	FNSEM			HSE			
	Imm. (≥16)	Imm. (<16)	Min. Nat.	Imm. (≥16)	Imm. (<16)	Min. Nat.	White Imm
Males	-0.063	-0.103	-0.032	-0.087	-0.087	-0.059	-0.030
	(3.25)	(4.06)	(1.00)	(4.41)	(3.09)	(2.10)	(0.84)
Females	-0.091	-0.097	0.006	-0.066	-0.048	-0.017	-0.017
	(4.92)	(3.56)	(0.20)	(3.56)	(1.72)	(0.70)	(0.57)

Note: Dependent variable = 1 if individual affected by illness. Marginal effects evaluated at sample means. Absolute t-values in brackets. Imm. (≥16) are immigrants whose age at migration was 16 years or older. Imm. (<16) are immigrants who migrated before turning 16.

Table 4.4: Health Differentials on Self-Assessed Health Measure (base group: Whites)

	Unconditional						
	FNSEM			HSE			
	Imm. (≥16)	Imm. (<16)	Min. Nat.	Imm. (≥16)	Imm. (<16)	Min. Nat.	White Imm
Males	0.153	-0.028	-0.113	0.196	0.028	-0.134	0.068
	(8.28)	(1.25)	(4.63)	(13.06)	(1.32)	(7.17)	(2.31)
Females	0.099	-0.086	-0.128	0.213	0.013	-0.052	-0.009
	(5.79)	(3.64)	(5.74)	(16.08)	(0.60)	(3.06)	(0.38)
	Conditioning on age, education, and YSM						
	FNSEM			HSE			
	Imm. (≥16)	Imm. (<16)	Min. Nat.	Imm. (≥16)	Imm. (<16)	Min. Nat.	White Imm
Males	-0.107	-0.198	-0.006	-0.009	-0.058	0.003	-0.097
	(2.13)	(3.72)	(0.20)	(0.26)	(1.46)	(0.11)	(2.56)
Females	-0.023	-0.152	-0.003	0.105	0.028	0.115	-0.087
	(0.52)	(2.91)	(0.12)	(3.45)	(0.72)	(5.38)	(2.39)
	Conditioning on age, education, household income and YSM						
	FNSEM			HSE			
	Imm. (≥16)	Imm. (<16)	Min. Nat.	Imm. (≥16)	Imm. (<16)	Min. Nat.	White Imm
Males	-0.204	-0.282	0.005	-0.071	-0.104	-0.017	-0.114
	(3.49)	(4.82)	(0.13)	(1.96)	(2.42)	(0.65)	(2.74)
Females	-0.041	-0.174	-0.001	0.020	-0.041	0.074	-0.128
	(0.82)	(2.88)	(0.05)	(0.60)	(0.96)	(3.13)	(3.24)
	Conditioning on age and education						
	FNSEM			HSE			
	Imm. (≥16)	Imm. (<16)	Min. Nat.	Imm. (≥16)	Imm. (<16)	Min. Nat.	White Imm
Males	0.125	0.055	0.054	0.145	0.123	0.025	0.034
	(6.42)	(2.18)	(1.75)	(9.38)	(5.26)	(1.05)	(1.16)
Females	0.098	0.014	0.039	0.195	0.140	0.127	-0.015
	(5.34)	(0.51)	(1.40)	(12.86)	(5.90)	(5.98)	(0.59)
	Conditioning on age, education, household income						
	FNSEM			HSE			
	Imm. (≥16)	Imm. (<16)	Min. Nat.	Imm. (≥16)	Imm. (<16)	Min. Nat.	White Imm
Males	0.114	0.037	0.085	0.106	0.104	0.012	0.056
	(5.00)	(1.25)	(2.36)	(6.06)	(3.98)	(0.45)	(1.70)
Females	0.078	-0.009	0.040	0.152	0.119	0.090	-0.021
	(3.67)	(0.29)	(1.29)	(8.77)	(4.52)	(3.83)	(0.74)

Note: Dependent variable = 1 if health assessed as fair, poor, or very poor. Marginal effects evaluated at sample means. Absolute t-values in brackets. Imm. (≥16) are immigrants whose age at migration was 16 years or older. Imm. (<16) are immigrants who migrated before turning 16.

The self-assessed measures of health show a different picture on health differentials than the CD measure. The hypothesis of self-selection of immigrants seems to be weakly supported by the SARHS, and not at all by the SAHS. Rather, immigrants' health evaluation at arrival is not significantly different from that of natives. However, the two health proxies provide information on different aspects of health with respect to the CD measure. Self-evaluation of health as poor possibly depends also on the sensitivity of the individuals to ailments which are not necessarily classified as "chronic diseases", but still affect their well-being. Such ailments may be psychologically and/or environmentally driven, and, as such, they do not affect individuals before migration or in the initial time after their arrival. Therefore, they may not determine the degree of selection of immigrants. The results including household income as a regressor seem to support this claim. Some literature finds evidence about links between harassment and negative self-evaluation of health (Karlsen and Nazroo [2002]). Furthermore, the well-being of ethnic minority immigrants may be affected by the drastic changes in culture and habitat experienced after migration. Finally, if individuals evaluate their health relatively to other people (as it certainly is for the SARHS measure), this comparison group may change over time and depend on whether and how soon immigrants mix with the native population.

Results on minority natives show sharp differences in the self-evaluation of health of women between the two surveys, but very similar coefficients for men. This seems to stress even further how the determination of somebody's health status depends on the health measure used. Further investigation is required to provide deeper insight on this issue, possibly using more information on factors affecting health.

4.6 Conclusion

Health is an important measure of well-being, equality and social integration. Ethnic minority communities (immigrants and natives) represents a significant proportion of the UK population. To analyse the welfare and the degree of social integration of ethnic minorities is important for public policy purposes. To our knowledge, this is the first study to examine health differentials among UK ethnic groups from an economic perspective.

The aim of the chapter is to determine whether immigrants are positively selected in

terms of health. We focus on ethnic minority immigrants, but also provide some evidence on white immigrants and ethnic minority natives. A model of migration decision is proposed, where health is included as a determinant. According to the model, healthier individuals migrate if their economic incentive is higher than for less healthy individuals. This happens under the assumptions that both income and the net gain from migration increase with health. This framework is then applied to a simple household decision model to understand how the health endowment of all members can affect the migration decision of the family.

To test for selection, we analyse health differentials between ethnic minority immigrants and UK-born individuals. We show that such a comparison is valid when certain assumptions on the health distributions in the UK and the countries of origin are satisfied.

Three different measures of health are used to estimate health differentials. These measures are imperfect proxies of health and, as such, may provide different aspects of health.

Results suggest several conclusions. First, according to the measure of health based on the incidence of chronic disease, individuals who migrated for economic economic reasons are positively selected in terms of health. In other words, they are healthier than the average population. This result is in line with previous literature. The positive selection seems to extend to both white and non-white immigrants. Second, still according to the same health proxy, individuals who migrated, but not necessarily for economic reasons (such as women and young children), also seem to be positively selected, and to a similar degree. This suggests that, when a family migrates, the migration decision is taken jointly in the family and in consideration of the health status of all components.

Immigrants' health advantage is significantly higher when controlling for duration of stay than when evaluated at average years since migration. More comprehensive surveys are needed to analyse the health distribution of different cohorts of immigrants.

The selection hypothesis does not seem to be supported by the SAHS measure and is weakly supported by the SARHS measure. However, these two proxies provide information on health which may not be relevant for the selection process, but still matter in measuring the degree of social exclusion. This is true if they reflect the individual's susceptibility to ailments which are less serious than chronic disease, but still affect his/her well-being, such as psychologically and environmentally driven ailments.

We find no supporting evidence to the possibility of intergenerational correlation of health between ethnic minority natives and their parents.

4.7 Appendix

Table 4.5: Health Determinants - Long-standing Illness

	FNSEM				HSE			
Variable	Males	Females	Males	Females	Males	Females	Males	Females
Degree	-0.085	-0.134	-0.095	-0.113	-0.119	-0.108	-0.125	-0.108
	(3.89)	(4.45)	(4.43)	(4.45)	(6.08)	(5.22)	(6.38)	(5.23)
A-level	-0.050	-0.051	-0.045	-0.036	-0.079	-0.072	-0.076	-0.062
	(2.32)	(2.47)	(2.09)	(1.72)	(4.24)	(4.05)	(4.12)	(3.51)
O-level	-0.053	-0.047	-0.052	-0.384	-0.061	-0.060	-0.059	-0.054
	(2.92)	(2.71)	(2.85)	(2.21)	(3.55)	(3.92)	(3.50)	(3.54)
Age	0.006	0.011	0.010	0.016	0.005	0.009	0.008	0.012
	(2.13)	(4.29)	(4.01)	(6.95)	(2.51)	(4.95)	(3.69)	(6.64)
$Age^2/100$	0.00	-0.004	-0.003	-0.008	0.003	-0.009	0.001	-0.003
	(0.06)	(1.62)	(1.15)	(3.67)	(1.34)	(0.47)	(0.63)	(1.60)
YSM	0.010	0.009	–	–	0.012	0.011	–	–
	(2.26)	(1.87)	–	–	(4.19)	(4.41)	–	–
$YSM^2/100$	-0.009	-0.004	–	–	-0.016	-0.013	–	–
	(0.94)	(0.39)	–	–	(2.96)	(2.90)	–	–
Imm.(≥16)	-0.220	-0.204	-0.054	-0.075	-0.207	-0.176	-0.047	-0.031
	(4.60)	(4.80)	(3.23)	(4.79)	(5.78)	(5.48)	(2.80)	(1.93)
Imm.(<16)	-0.238	-0.223	-0.097	-0.092	-0.261	-0.201	-0.091	-0.025
	(5.21)	(5.24)	(4.43)	(3.98)	(6.26)	(4.98)	(3.70)	(1.03)
Min. Nat.	-0.056	-0.037	-0.017	-0.003	-0.061	-0.031	-0.040	-0.013
	(2.03)	(1.43)	(0.61)	(0.13)	(2.53)	(1.44)	(1.68)	(0.63)
White Imm.	–	–	–	–	-0.192	-0.182	-0.037	-0.020
	–	–	–	–	(4.46)	(4.61)	(1.14)	(0.74)
No. Obs.	3686	4233	3686	4233	6235	7367	6235	7367
Pseudo-R2	0.1156	0.1243	0.1096	0.1175	0.0987	0.0983	0.0954	0.0952

Notes:

Dependent variable = 1 if individual affected by illness.

Marginal effects evaluated at sample means. Absolute t-values in brackets.

Table 4.6: Health Determinants - Self-assessed measure of global health

	FNSEM				HSE			
Variable	Males	Females	Males	Females	Males	Females	Males	Females
Degree	-0.203	-0.173	-0.211	-0.171	-0.193	-0.200	-0.195	-0.199
	(9.07)	(6.41)	(9.73)	(6.37)	(12.66)	(12.12)	(12.90)	(12.14)
A-level	-0.124	-0.170	-0.121	-0.153	-0.166	-0.168	-0.164	-0.165
	(5.48)	(7.68)	(5.41)	(6.94)	(11.15)	(11.65)	(10.98)	(11.45)
O-level	-0.107	-0.107	-0.105	-0.095	-0.099	-0.118	-0.098	-0.115
	(5.61)	(5.74)	(5.55)	(5.20)	(7.12)	(9.18)	(7.07)	(9.00)
Age	0.003	0.005	0.011	0.011	0.007	0.005	0.009	0.006
	(1.14)	(2.02)	(3.91)	(4.30)	(3.67)	(2.80)	(5.19)	(4.01)
$Age^2/100$	0.00	-0.001	-0.005	-0.005	0.001	0.00	-0.003	0
	(0.19)	(0.51)	(1.83)	(2.07)	(0.80)	(0.38)	(1.88)	(0.51)
YSM	0.016	0.005	–	–	0.011	0.006	–	–
	(3.39)	(0.99)	–	–	(4.18)	(3.03)	–	–
$YSM^2/100$	-0.019	0.008	–	–	-0.015	-0.009	–	–
	(1.93)	(0.68)	–	–	(3.14)	(2.45)	–	–
Imm.($\geq$16)	-0.107	-0.023	0.125	0.098	-0.009	0.105	0.145	0.195
	(2.13)	(0.52)	(6.42)	(5.34)	(0.26)	(3.45)	(9.38)	(12.86)
Imm.(<16)	-0.198	-0.152	0.055	0.014	-0.058	0.028	0.123	0.140
	(3.72)	(2.91)	(2.18)	(0.52)	(1.46)	(0.72)	(5.26)	(5.90)
Min.Nat.	-0.006	-0.003	0.054	0.039	0.003	0.115	0.025	0.127
	(0.20)	(0.12)	(1.75)	(1.40)	(0.11)	(5.38)	(1.05)	(5.98)
White Imm.	–	–	–	–	-0.097	-0.087	0.034	-0.015
	–	–	–	–	(2.56)	(2.39)	(1.16)	(0.59)
No. Obs	3682	4225	3682	4225	6236	7364	6236	7364
Pseudo-R2	0.1091	0.0781	0.1010	0.0727	0.1347	0.1161	0.1313	0.1150

Notes:

Dependent variable = 1 if health assessed as fair, poor, or very poor.

Marginal effects evaluated at sample means. Absolute t-values in brackets.

Part III

"Reception"

Chapter 5

The Local Labour Market Effects of Immigration in the UK

Abstract:

This chapter provides a comprehensive picture of the way immigration affects labour market outcomes of native born workers, embedded into a representation of the underlying theoretical mechanisms, and under the constraints given by the availability of data sources. Our investigation is the first for Britain. The analysis concentrates on employment and wage effects of immigration. We discuss the problems that may arise in empirical estimations, and suggest ways to address these problems. Our empirical analysis is based on data from the British Census and the Labour Force Survey. There is some evidence that immigration affects employment prospects negatively and that it enhances wage growth; however, in most cases, estimated effects are not significantly different from zero. We explain that it is quite compatible with standard economic theory for immigration to have no long run wage or employment effects.

5.1 Introduction

The possible negative effects of immigration on wages and employment outcomes of native workers is one of the core concerns in the public debate on immigration. The possibility that changes in the size or composition of the labour force resulting from immigration could harm the labour market prospects of some native workers is compatible with simple economic models. Not surprisingly, therefore, research on wage and employment effects of immigration is one of the core areas on migration research in economics. There is a considerable number of papers addressing this issue for the US, and some papers for other European countries. The common conclusion of this work, apart from a small number of exceptions, is that immigration has only very small, or no effect on employment and wages of native workers. No work exists for the UK.

One purpose of the current research is to fill this gap, and to conduct such an exploration. The dominant methodology in the literature, which we follow also in this report, is to seek to infer labour market effects from spatial correlations between local immigrant inflows and local changes in the labour market outcomes of natives. At the stage of empirical implementation, this methodology raises a number of important issues. Most of these relate to a clear isolation of the effect of immigration on native labour market outcomes from other associated phenomena, particularly in a context where immigrant inflows are themselves the outcome of economic decisions. Much of the empirical literature is concerned with addressing these problems. We shall discuss the appropriate empirical strategies to solve these problems, and implement them as far as our data allows us to do so.

One problem with studies on the impact of immigration on labour market outcomes is that spatial information is necessary to construct measures of geographic concentration of immigrants. Many survey data sets do not include detailed spatial information - for instance, the British Labour Force Survey (LFS) includes spatial information only on regional level. A further problem is that surveys contain only small numbers on immigrants, so that the allocation of that information to spatial units, even if not detailed, may be miss-measured. Also, sample size may be an obstacle to any impact analysis that is intended to distinguish between different groups in the native population (for instance, by gender and skills). Administrative data sets like the Census solve the problem of small sample size, and, in principle, allow also to use a finer spatial allocation. On

the other hand, this data is only available once every decade, and limited information on background characteristics restricts possibilities of a detailed impact analysis for specific skill and demographic groups.

In this paper, we will use data from two sources: the LFS, and the Census. Where appropriate, we will combine these data sources. The data are complementary both in the time period they cover, and in the groups they allow us to analyse. Where they overlap, they enable us to check the robustness of our results.

We commence in the next section with a brief account of the background to this literature and explain the data sources we use. In Section 5.3 we discuss the problems which occur on the empirical level. Sections 5.4 and 5.5 report the results of our empirical analysis. Finally, Section 5.6 concludes and suggests avenues for future work.

5.2 Background

5.2.1 Theory

The theoretical analysis of the labour market effects of immigration sees effects as arising from the changes it introduces in supply of skills and consequent change in labour market equilibrium. Typically a distinction is drawn between skilled and unskilled labour. Immigration inflows affect the skill composition of the labour force if the skill composition of immigrants does not match the skill composition of natives. This change in skill composition leads to disequilibrium between supply and demand of different labour types at existing wages, prices and output levels. Restoration of equilibrium will almost certainly involve short run changes in wages and employment levels of different skill types and may or may not require long run changes[1].

The literature includes different approaches to theoretical modelling of these processes and different conclusions about the nature of long run effects. The main differences in assumptions made involve (i) differences in the number of goods produced and therefore in the flexibility of the economy to adapt through changes in mix of outputs, and (ii) differences in openness of the goods sector to trade and therefore in the extent

[1]Another less common approach (see for example Lalonde and Topel [1991]) treats immigrant and native labour as different labour types. In such a model the effect of immigration depends on substitutability between immigrant labour and native labour of different skill levels. The form of equations arising for estimation are nonetheless not dissimilar to those under the more common approach.

to which output prices are set locally or on world markets.

Models assuming limited flexibility of output mix or closedness to international trade tend to predict that immigration will have long run wage and employment effects. Such features are typical of the underlying framework used as a motivation for empirical work in this literature (see for example the models of Borjas [1999] or Card [2001]).[2].

On the other hand, models assuming a sufficiently high degree of flexibility in output mix and openness to trade predict an absence of long run effects on labour market outcomes, at least to small scale immigration. Leamer and Levinsohn [1999] refer to this as the hypothesis of *factor price insensitivity*[3]. In the context of discussion of immigration this is sometimes referred to as the *structural hypothesis.* Although it is not often a feature of the models favoured in the empirical literature on the impact of immigration, this fact is sometimes mentioned [4] (see, for example, Borjas [1999], Card [2001], Pischke and Velling [1997], Chiswick [1993]), several recent contributions lay more stress on the need for models with multiple goods and openness to trade (see, for example, Kuhn and Wooton [1991], Scheve and Slaughter [2001], Hanson and Slaughter [1999], Hanson and Slaughter [2002], Gaston and Nelson [2000], and Gaston and Nelson [2001]).

In an associated paper (Dustmann et al. [2003]), we lay out a comprehensive equilibrium model of the effects of immigration on the labour market. This model subsumes the range of models in the literature and provides a guide to specifications used in the estimation. The basic features of the model are as follows: the economy produces several goods using several labour types. Some of these goods are traded internationally at prices fixed on world markets. The number of workers of each labour type is determined by immigration. Their labour is flexibly supplied depending on the wage. In the long run, there is free entry of firms into profitable sectors.

We assume conventionally that in such an economy, wages, prices and output levels vary in the long run to maintain equilibrium between supply and demand in labour

[2]In this, these models share the features of standard models used in the broader literature on wage determination. See, for example, the influential papers of Katz and Murphy [1992] or Murphy and Welch [1992].

[3]This result is related to the well known factor price equalisation result of trade theory - see, for example, Woodland [1982], and Samuelson [1948] - although it is a weaker result.

[4]Maybe because most applications are to the US, which is less plausibly viewed as a small open economy than, say, the UK.

markets, to maintain equilibrium between supply and demand in product markets, and to maintain no incentive to further entry of firms by keeping zero profits in goods markets.

In the short run, disequilibria can exist, allowing excess demand or supply of labour and positive or negative profits in particular markets.

The nature of the labour market impact of immigration depends crucially on the scope for absorbing the impact through changes in the mix of output in the traded goods sector.

Consider, for instance, an economy with a small and homogeneous traded goods sector (and, therefore, relatively little flexibility in the output mix of traded goods). Long run responses do involve long run changes in the wage and employment structure as well as output structure. The lack of flexibility in output mix means that there are insufficient degrees of freedom to accommodate changes in the skill mix through changes in the output mix. Wage changes are therefore not zero even in the long run. This is the sort of case typically presented as theoretical background literature to empirical studies.

Now consider an economy with a large and heterogeneous traded goods sector (and, therefore, relatively high flexibility in the output mix of traded goods). In such an economy, long run wages and employment levels are insensitive to immigration. This is the Leamer and Levinsohn [1999] *long run factor price insensitivity* result already discussed. Wages are determined by world prices and technology. Rather than impacting on wages, long run effects of immigration are felt in the output mix.

However, wages can be affected in the short run. The mechanism by which the economy adjusts is as follows. Any depressive effects on wages lead to positive profits being earned in sectors using intensively labour types which become cheaper. As a consequence, output in such sectors expands, driving back up wages. In the long run, equilibrium will be restored with wages driven back to their initial levels.

This exposition shows that a variety of possible outcomes are compatible with economic theory. Immigration may depress wages and employment of natives. However, it is by no means inconsistent with economic theory to think that long run responses to immigration may involve no effect. What matters is the openness of the economy to trade[5] and the flexibility of the economy to adjust in respects other than wages and in particular through the mix of output produced.

[5]It should be noted that the empirical analysis below applies to regions within the UK. These are certainly open to trade with each other for much of their production.

5.2.2 Previous Literature

An extensive empirical literature exists on the impact of immigrants on the labour markets of host countries (see Borjas [1994] and Borjas [1999], for an overview). Most of these studies relate to the US and typically use microdata from the US census. The common consensus of most of this work is that the impact of immigration on wages and employment in local labour markets is, if at all, modest. Much less work exists for countries outside the US. Pischke and Velling [1997], De New and Zimmermann [1994], and Haisken De New and Zimmermann [1999] analyse data for Germany, Hunt [1992] analyses data for France, Carrington and de Lima [1996] analyse data for Portugal and Winter-Ebmer and Zweimüller [1996] and Winter-Ebmer and Zweimüller [1999] analyse Austria. Findings of these studies are typically in line with the US evidence, establishing only small effects of immigration on local labour markets.

The consensus in the literature is that employment and wage effects of immigration are small. Lalonde and Topel [1991] notice that"... increased immigration reduces the wages and earnings of immigrants and their close substitutes, though in our view the effects are not large ... Labor market effects on non-immigrants appear to be quantitatively unimportant." Altonji and Card [1991] conclude "Our empirical findings indicate a modest degree of competition between immigrants and less skilled natives ... We find little evidence that inflows of immigrants are associated with large or systematic effects on the employment or unemployment rates of less skilled natives." Card [2001] does find employment effects, but he states that: "the conclusion that immigrant inflows affect native employment rates is new. However, the implied effects for natives as a whole are very small. Even for workers in the bottom of the skill distribution, I find relatively modest employment effects of recent immigrant inflows in all but a few high - immigrant cities."

Conclusions of studies for Europe are very similar. De New and Zimmermann [1994] report that: "Immigration ... appears to have an overall negative effect on German wages. ... However ... the estimated effects are far from being dramatic and are well in line with economic theory." Pischke and Velling [1997] find "little evidence for displacement effects due to immigration." Finally, Winter-Ebmer and Zweimüller [1999] conclude that "The results indicate only a modest impact of immigration on the unemployment risk for native employees."

5.2.3 Data used for the analysis

The first data set we use for our analysis is the Labour Force Survey (LFS). The LFS is a household survey, conducted by the Office for National Statistics (ONS) on behalf of the Department for Education and Employment (DfEE). It provides a wide range of data on labour market statistics and related topics such as training, qualifications, income and disability. The LFS has been carried out in the UK since 1973. Between 1973 and 1983 a biennial survey was carried out during the spring. Between 1983 and 1991, the LFS was undertaken annually in the Spring of each year and before that every 2 years, beginning in 1973, originally to derive comparable labour market statistics that were required for Britain's accession to the European Union in 1975. The sample size was around 60,000 households in each survey, around 0.5% if the population. In Spring 1992, for the first time, the data were made available quarterly, with a quarterly sample size approximately equivalent to that of the previous annual data, thus becoming the Quarterly Labour Force Survey. Each quarter interviews are achieved at about 59,000 addresses with about 138,000 respondents. A core of questions covering household, family structure, basic housing information and demographic details of individuals in the households is included in every survey, together with non-core questions which vary from quarter to quarter. The British LFS contains spatial information only at regional level, except for a brief interval between 1997 and 1999 when data was made available at county level.

The Census of Population data sets is a questionnaire survey of the United Kingdom population held every ten years. The aim of the Census is to obtain a picture of the socio-economic state of the country. The three years used for this study are 1971, 1981 and 1991 (these are also the only ones available electronically). They contain information on total population, gender, age, marital status, country of birth, economic activity, employment status and various household characteristics. Additional information can be found in the more detailed 1991 version, like ethnic group, qualifications and weekly hours worked.

The information is available only in selected tables of aggregate data for geographical areas of the United Kingdom which broadly correspond to administrative areas. This implies a limited use of the data if further disaggregation is required in the analysis. In our case, for instance, we cannot obtain information on number of immigrants by qualification, gender or employment status.

The Appendix contains further details on the creation of the data set used for this analysis.

5.3 Empirical Implementation

The dominant approach to the estimation of such a model in the literature is that referred to by Borjas [1999] as the "spatial correlations" approach. Effects of immigration are identified from the spatial correlation between immigrant labour inflows and changes in native or overall labour market outcomes (or between immigrant population shares and levels of these outcomes). Spatial units are intended to correspond to geographical labour markets. In the US context, the spatial units usually used for empirical analysis are standard metropolitan statistical areas.

5.3.1 Problems in estimation

The typical empirical study regresses a measure of employment or wages of native workers in a given area on relative quantities of immigrants in that particular locality and appropriate controls. We discuss these problems and the way we intend to solve them.

Fixed effects: Levels of immigrant shares and levels of labour market outcomes may be spatially correlated because of common fixed influences, leading to a positive or negative statistical correlation between immigrant concentration and economic outcomes, even in the absence of any genuine effects of immigration. To address this problem, we use difference and within groups estimation.

Simultaneity: The direction of causality between immigrant inflows and labour market outcomes is not necessarily clear-cut. Immigrants may be attracted to those areas that are enjoying current economic success. In this case it is not only that immigrant inflows are driving labour market changes, but that labour market changes are driving inflows. This selective settlement would lead to an upwardly biased estimate of the effects of immigrants' concentration on labour market outcomes of natives.

A possible solution to this problem is instrumental variables regression. As instruments, we use measures of historic settlement patterns. The underlying assumption is that immigrants take account of existing networks and the presence of individuals with the same culture and language as themselves. Thus, besides possibly choosing areas that were subject to favourable recent economic shocks (which creates the problem), immi-

grants settle in areas with already high immigrant concentrations. Preexisting immigrant concentrations are unlikely to be correlated with current economic shocks if measured with a sufficient time lag, since existing concentrations are determined not by current economic conditions, but by historic settlement patterns of previous immigrants.[6] Of course, the assumption that lagged values of immigrant stocks are correlated with employment changes only through their relation with immigrant inflows is an identifying assumption that is not testable. It could be problematic if local economic shocks were persistent and instruments were insufficiently lagged. The strength of correlation between lagged concentrations and current inflows is observable in data and can therefore be assessed.

Measurement error: Measures of immigrant concentrations may suffer from measurement error due to small sample size. This is likely to be the case in our analysis that is based on the LFS. The consequences of any measurement error is aggravated when using difference or within groups estimation. To address this problem, we use instrumental variables regression. We use historic settlement patterns as instruments.

Native outflows: Local labour markets are not closed economies and native workers are free to move out. If immigration does drive down local wages for certain skill groups then one would expect there to be pressure for native workers of that skill type to move elsewhere. This will tend to disperse the impact of immigration through the national economy and undermine the ability to identify the impact from looking at effects within localities, leading to upward biased estimates of the effect of immigration on employment of native workers. This point has been stressed in numerous contributions. The US literature contains conflicting opinions on the seriousness of the problem. Borjas [1999] regards it as more serious than Card [2001]. The problem is one of an omitted term in the estimated equation. The most attractive resolution to this problem is available if native outflows are observable and therefore amenable to incorporation directly into the estimation, as is the case in one of our data sources. However such outflows are likely to be correlated with shocks to local economic conditions for the same reasons as immigrant flows, discussed above, creating a further simultaneity issue. These outflows therefore also need instrumenting and it is theoretically less clear what would serve as

[6]Work following this approach (see e.g. Card [2001]) has been influenced by the findings of Bartel [1989] who argued that immigrants in the US tend to settle in areas where immigrant settlement is already strong. In Chapter 3 we also find evidence of this pattern for the UK.

a suitable instrument. In practice we rely on lags.

5.3.2 Estimation Strategy

The discussion we had above on the possible problems at the empirical stage can be summarised in the following equations:

$$\ln w_{it} = \alpha_0 + \alpha_1 \pi_{it} + \alpha_2 \ln \mathbf{n}_{it} + \alpha_3 \mathbf{a}_{it} + \lambda_t^w + \mu_i^w + u_{it}^w \tag{5.1}$$

$$U_{it} = \beta_0 + \beta_1 \pi_{it} + \beta_2 \ln \mathbf{n}_{it} + \beta_3 \mathbf{a}_{it} + \lambda_t^U + \mu_i^U + u_{it}^U \tag{5.2}$$

where w_{it} denotes wage, U_{it} denotes unemployment rate, π_{it} denotes the ratio of immigrant to non-immigrant population, $\mathbf{n}_{it}$ denotes a vector of non-immigrant skill group populations and $\mathbf{a}_{it}$ denotes a vector of average ages, all in the ith region in the tth period. Here λ_t^w and λ_t^U are year effects, μ_i^w and μ_i^U are region effects and u_{it}^w and u_{it}^U are disturbance terms.

Homogeneity is imposed on the native skill group effects by omitting one skill category and expressing the others as ratios with the size of the omitted skill group.

A static model does not allow to draw distinction between short and long run effects. In the short run, disequilibria between supply and demand in labour and product markets may exist. Market responses may be slow and immigration may produce temporary situations in which nonzero profits are being earned. Inclusion of lagged values of immigrant shares and employment (wage) levels as additional regressors allows more sophisticated dynamic specifications to be recognised and short and long run effects to be distinguished. In such a context, it is possible to allow for short run effects without ruling out long run insensitivity.

All estimates are calculated in GAUSS using DPD98 (see Arellano and Bond [1991], and Arellano and Bond [1998]). Instrumental variables estimates are calculated by GMM imposing the moment restriction that Δu_{it}^w or Δu_{it}^U is uncorrelated with the chosen instruments, which in each case are two- and three-period lags of the endogenous variables π_{it} and $\mathbf{n}_{it}$. Weighting of restrictions and calculation of standard errors recognises the anticipated first order serial correlation in the differenced residuals.

Tests are reported for first and second order serial correlation of residuals and for the overidentifying restrictions implied by choice of instruments. For all IV estimates re-

ported below there is clear evidence of first order serial correlation, as should be expected given differencing of the residuals, but absence of second order serial correlation cannot be rejected at usual significance levels. The overidentifying restrictions are rejected in none of the specifications reported.

We provide estimates using a number of different estimators. Although several of these have obvious drawbacks they nonetheless offer a useful point of comparison to results of more robust methodologies and also to comparable results in the empirical literatures for other countries.

In all estimated specifications we include a full set of year effects so that aggregate time series variation is completely absorbed. Immigration may certainly have an important impact at the level of the whole economy but we do not think it wise to attempt to disentangle this from the effects of cyclical variation empirically. We are aware of no study which does this. In all estimations based on the LFS, we also include controls for average age of immigrants and natives. These are taken as given in subsequent discussion. Size of native skill groups are also entered as controls in order to allow for the effect of native outflows.[7]

We report results using the OLS estimator, a difference estimator, and the IV estimator in differences. With OLS, the effect of immigration on economic outcomes is identified from the period-by-period cross sectional correlation between relative immigrant stocks and employment levels. This offers a basic and straightforward point of comparison. However it is clearly subject to a number of serious problems, which we have discussed above. The within groups (difference) estimator adds region-specific effects to a levels regression to absorb any fixed element in the cross sectional variation. Identification of the effect is now from changes over time in the pattern of cross sectional variation. Either of these is more robust than simple OLS. However both still have problems with measurement error and simultaneity.

Combining estimation in differences with use of instrumental variables addresses both the issues of measurement error and simultaneity. In many ways this is the most attractive approach, subject to the appropriateness of the chosen instrumental variables.

For our work with the LFS, we take two- and three-period lagged values of immigrant shares and of native skill supplies as instruments. For our work with Census data, we

[7]We impose the standard assumption that equiproportionate changes in all skill groups will have no effect.

take immigrant shares at the beginning of the previous decade as instruments.

5.4 Analysis of Census Data

We commence our analysis with data from the 1971, 1981, and 1991 Censuses. The Census provides very accurate data on immigrant concentration and unemployment rates at a variety of spatial levels. We concentrate attention on the data at county level. Information is available in the form of selected published cross tabulations. The frequency of data collection is relatively low and the most recent available information is for 1991.

We have constructed variables that measure the concentration of immigrants in each census year in a particular county. We have also constructed measures of county level unemployment rates, again for each available census year. The choice of spatial unit is intended to correspond in some approximate sense to a local labour market. Choosing county gives us considerably more observations than choosing, say, region but it is arguably too small a spatial unit for the purpose. Results reported below using LFS are based on region as the spatial unit and offer an interesting point of comparison.

Information in the Census does not allow for a breakdown across native and foreign born individuals; the information we use is thus an average, including all groups. This is a serious weakness since we can not tell to what extent, if at all, employment effects reflect effects on natives. The best we can do is to conduct a simple analysis of the impact of immigration on unemployment. The information available allows further to estimate models using each of the above estimation strategies, therefore enhancing the robustness of our results towards possible contamination due to the problems we have discussed above.

Table 5.1 reports various estimates of the relationship between immigrant concentration and unemployment based on county level census data. In all cases the reported effect is that of an increase in the ratio of immigrant to native population on unemployment rate in the population as a whole.

The first column reports an estimate based on simple OLS regression. This is an estimate based on the correlations between levels of unemployment and immigrant concentration in the three census cross sections. Since this is a simple two dimensional correlation, we can use also a graphical representation for illustration. We have done

Table 5.1: Effect of immigration on total unemployment

Census 1971-1991

	Levels		Differences	
	OLS	Within groups	OLS	IV
Coefficient	-0.046	0.177	0.226	0.605
t value	(1.28)	(2.64)	(2.29)	(2.74)
Sample size	192	192	128	64

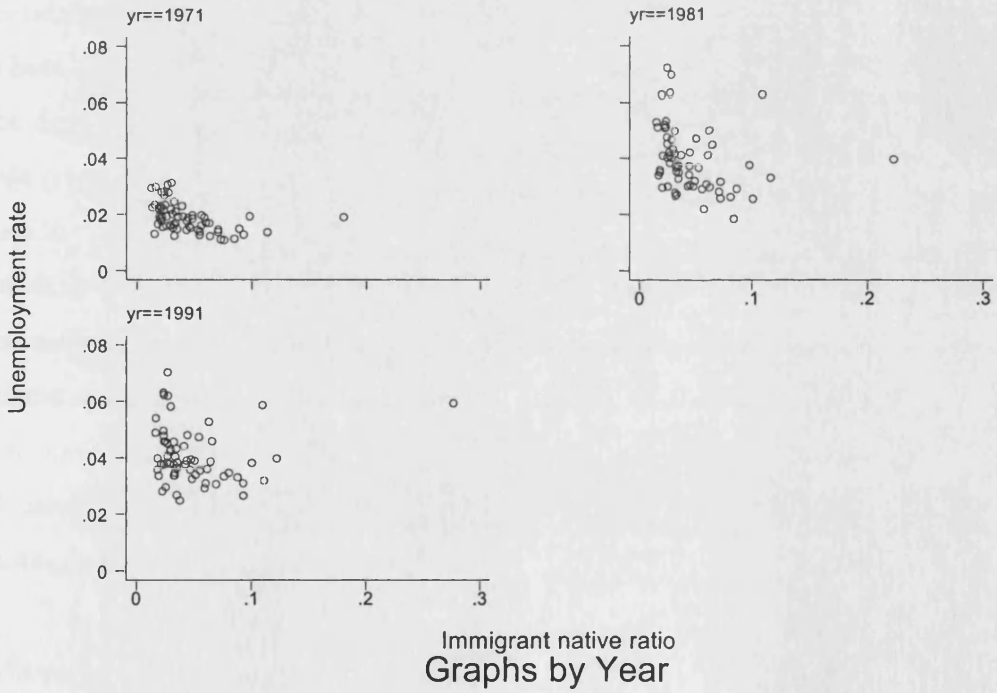


Figure 5.1: Level of unemployment and immigrant concentration: Census

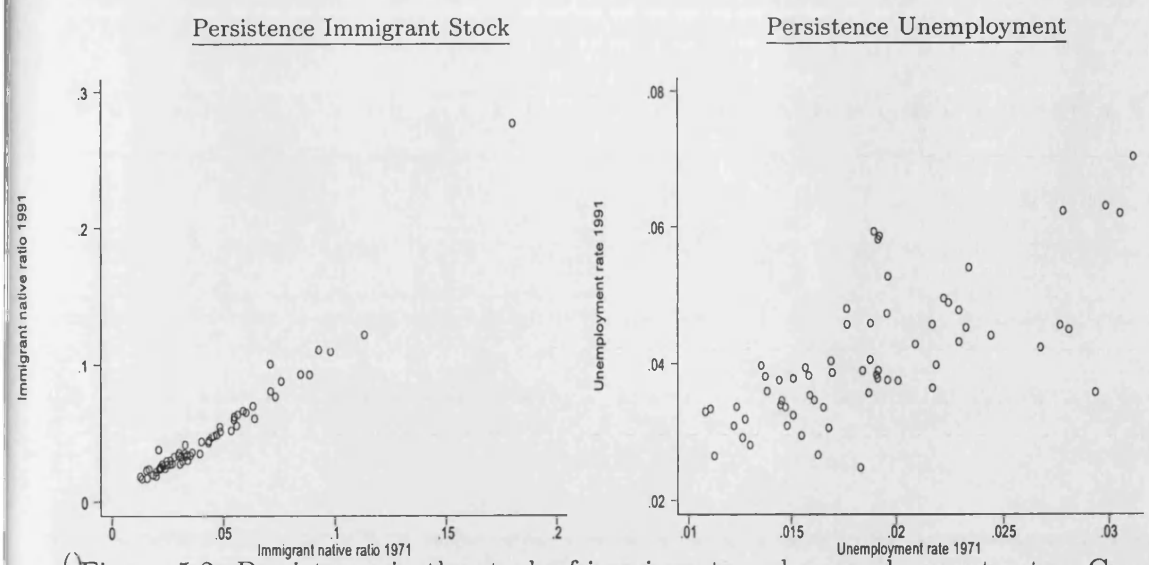


Figure 5.2: Persistence in the stock of immigrants and unemployment rates: Census

this in figure 5.1. The estimate is negative though statistically insignificant and accords with the impression from the figure that - particularly in the later two years - unemployment rates are negatively associated with immigrant concentration.

OLS may give seriously biased estimates of the effect if there are persistent and correlated county level effects in the variables under consideration. That this is the case is best illustrated in figures 5.2, where we use census data for the years 1971 and 1991. In the figures, we plot the concentration of immigrants (left panel) and the unemployment rate (right panel) in 1971 against the concentration of immigrants and the unemployment rate in 1991. Each point refers to a pair of one county at two points in time. The visual impression is of strong persistence in both the immigrant-native population ratio and the rate of unemployment. Whether this leads to positive or negative bias in estimated effects depends on whether immigrants settle predominantly in regions with high, or low unemployment. Whichever is the case, the figures suggest the potential importance of using estimation approaches which eliminate the persistence in both the stock of immigrants and economic conditions.

The second and third columns of Table 5.1 report results from within groups and difference estimation. These should both be robust to persistent correlated effects. Both estimates are positive and significant and of comparable magnitude. Figure 5.3 shows the reason for the change of sign. Although counties of high immigrant concentration have low unemployment, counties where immigrant concentrations increased - particularly between 1981 and 1991 - tended to be those where unemployment also increased. The estimated coefficients suggest a very mild effect of immigration on unemployment.

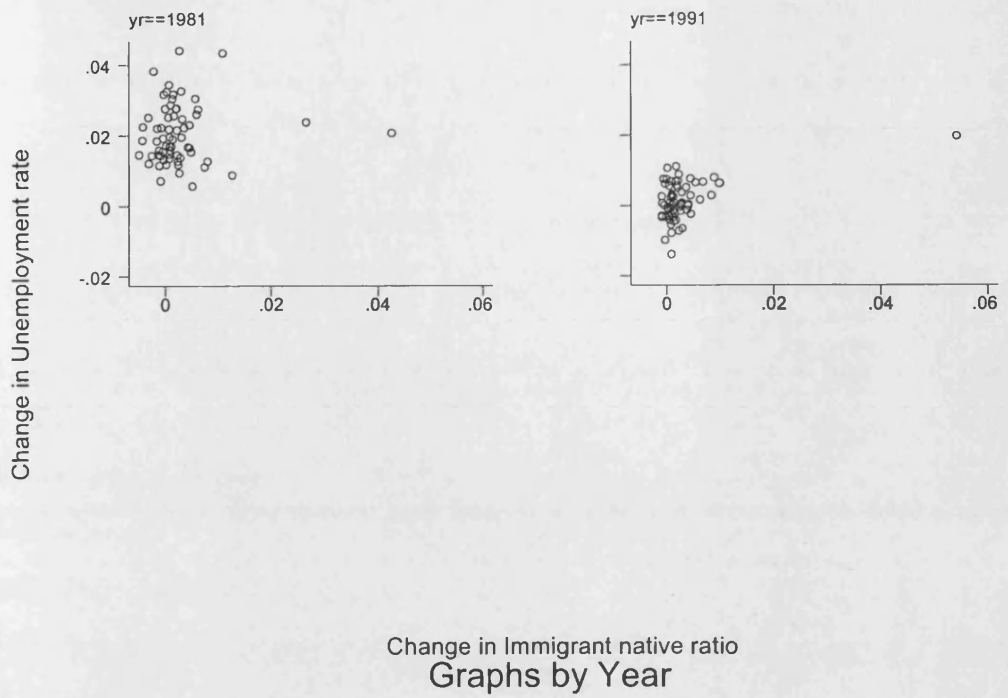


Figure 5.3: Changes in unemployment and immigrant inflow: Census

According to these estimates, an increase in the immigrant population by one percent of the native population leads to an increase in the percentage of the population unemployed of about 0.17 (for the within groups estimator) or 0.22 (for the difference estimator).

It is apparent from the figure that the relationship in the two decades may differ. Across both decades, it is also clear from the figure that one county, Greater London, enjoyed a substantially higher influx of immigrants than any other. To address concerns that this particular observation may be driving the results or that the decades may differ, we report also results based on selected samples in Table 5.2. We see firstly that the association between the changes is much stronger if we restrict attention to the more recent decade - the impact of immigration seems to double, and it is highly significant. Secondly, if we retain all census years for estimation, we see that the estimated effect does indeed fall and become statistically insignificant (though it remains positive) if we exclude London. However, if we take only the latter decade, excluding London strengthens the estimated effect.[8]

[8]The special role of the capital is something deserving greater attention and we intend to pursue it in future work. Since there is no question of mismeasurement involved in the outlying London observations, it could be persuasively argued that removing them amounts to ignoring the most informative data in

Table 5.2: Effect of immigration on total unemployment

Census 1971-1991

	OLS, Differences		
	All counties	London excluded	London excluded
	1981-1991	1971-1991	1981-1991
Coefficient	0.380	0.184	0.568
t value	(10.41)	(0.66)	(3.02)
Sample size	126	64	63

The estimates based on difference estimation may be biased if positive (or negative) shocks to local economic conditions influence immigrants' location decisions. To address this issue we use instrumental variables techniques. Essential here is that the earlier year's immigrant concentration is strongly correlated with the change in the later decade (rank condition of identification). Figure 5.4 illustrates the correlation between stock of immigrants in 1971 and the change in the immigrant population between 1981 and 1991. There is certainly a strong positive association, no matter whether we include or exclude London. This indicates, as we show in other parts of the thesis, that early settlement of immigrants attract younger immigrant cohorts.

Estimates are displayed in the final column of table 5.1. Given the validity of instruments, these estimates should not suffer from any simultaneity problem and may be regarded as the most technically robust of the estimates based on census data.

The IV estimate of the effect of immigration on unemployment is positive and statistically significant. Furthermore it is larger than the OLS estimate in differences. This is in line with what we would expect: If immigrants respond to positive shocks to local economic conditions, then the difference estimator should lead to an overly optimistic picture of the effect of immigration on unemployment.

This estimate, although being the most robust to be obtained from Census data, has a number of remaining problems that may compromise its reliability. Firstly, the dependent variable is unemployment in the whole population. Even if an association has been indicated between immigrant inflows and growth in unemployment, it is impossible on the basis of these results alone to say whether that is because the immigrants themselves are failing to find work or because native employment is declining as a consequence. Secondly, no controls have been included in the regression to capture native outflows or changes in native characteristics. For these reasons we regard these results

the sample and is not therefore desirable.

as indicative at best. Such issues are better addressed using Labour Force Survey data, as is done below.

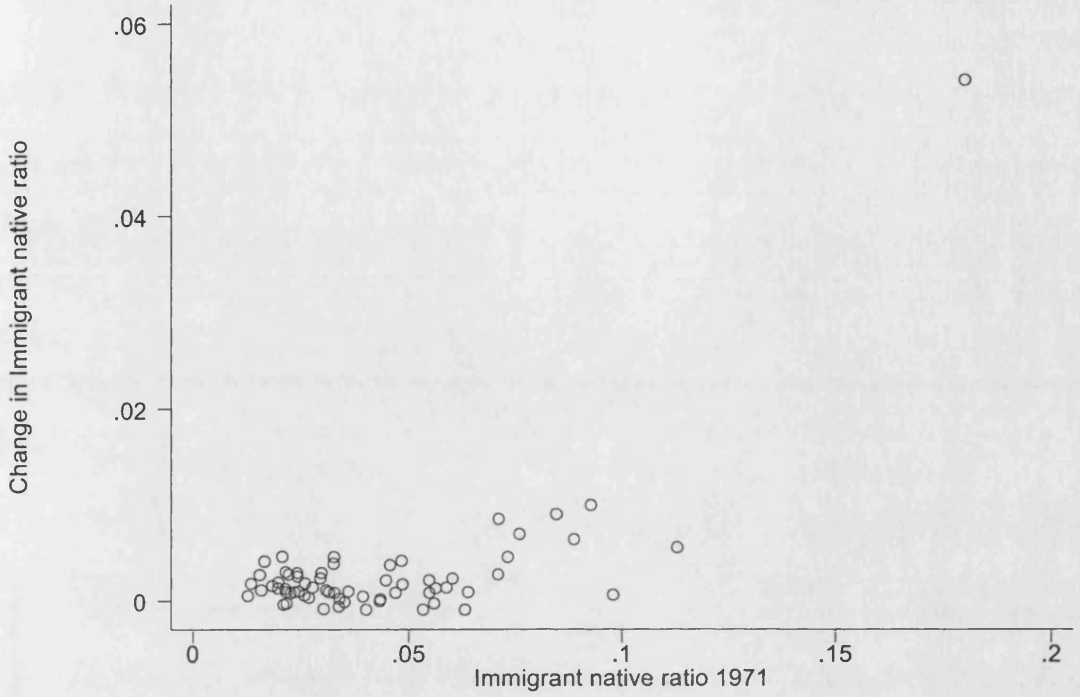


Figure 5.4: Stock of immigrants 1971 and change in immigrant population 1981-1991: Census

5.5 Analysis of LFS Data

Data from the Labour Force Survey is available at a much higher frequency providing substantially more points in the time series dimension. It also allows an analysis of changes in the 1990s (although not the 1970s). Data on employment are available from 1979 onwards and at yearly frequency from 1983 onwards.

Because raw microdata is available there is much greater scope to construct variables in ways corresponding to objects of theoretical interest. For example, native unemployment rates can be distinguished from overall unemployment rates allowing a more effective isolation of the economic effect of immigration on natives. The presence of relatively rich information on native skills also permits estimation of separate equations for different skill types as well as control for outflows of native workers by skill type.

However, sample sizes within years are much smaller and measurement errors there-

fore more pronounced, particularly as regards the key variable, inflows of immigrants. Spatial information is also weaker with only region distinguished in most years although, as argued above, that needs not be disadvantageous.

5.5.1 Unemployment

Tables 5.3 to 5.7 report a full set of a variety of regression estimates of the employment effects of immigration using LFS data. In all of these regressions, unlike those of the previous section using census data, the estimates control for the effect of flows of native workers and for changes in the age of native workers. This is potentially important, as the demographic structure across spatial units and across time may differ, and outflows of native workers may be correlated with inflows of immigrants.

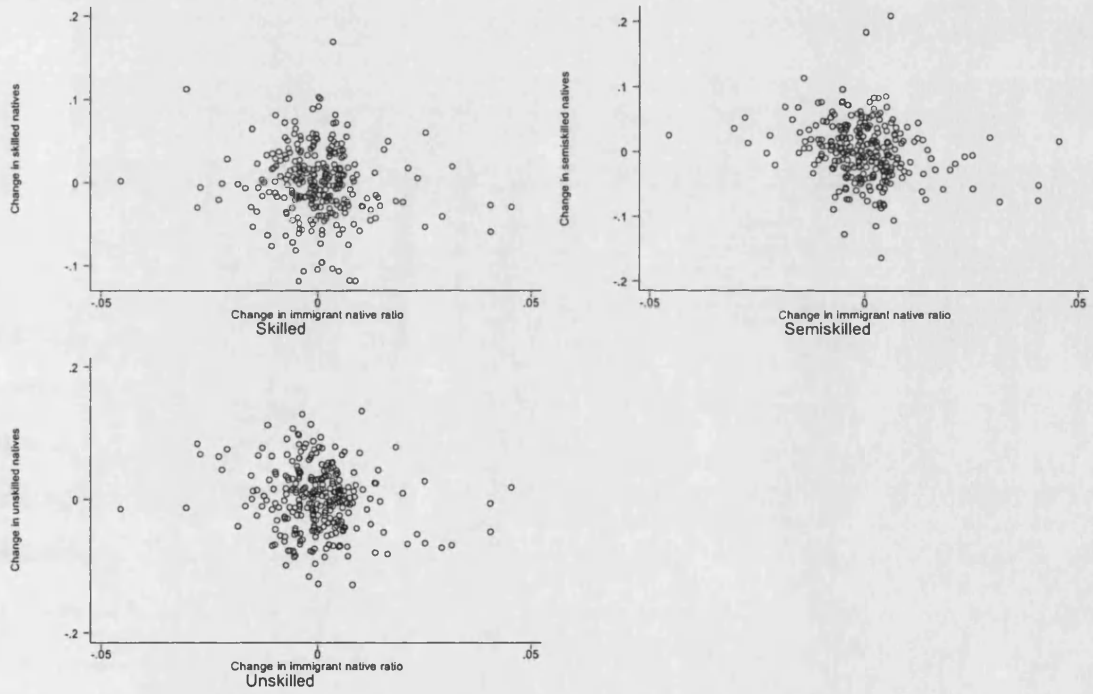


Figure 5.5: Changes in native workforce and changes in immigrant concentration: LFS

In Figure 5.5 we presents a plot of changes in native workforce by skill type against changes in immigrant concentration. Though not visually striking there is some evidence that high outflows of some labour types may be associated with growth of immigrant concentration.

Table 5.3 presents a series of different estimates of effects on total native unemployment in a way similar to Table 5.1, but based on LFS data, and adding the additional

Table 5.3: Effect of immigration on unemployment

LFS 1983-2000

	Levels				Differences			
	OLS		Within groups		OLS		IV	
Variable	Coeff	*t* value	Coeff	*t* value	Coeff	*t* value	Coeff	*t* value
Immigrant-native ratio	-0.050	-1.940	0.245	5.551	0.106	1.580	0.178	1.341
ln skilled/unskilled	-0.046	-6.059	-0.023	-1.928	-0.027	-2.451	-0.228	-1.721
ln semiskilled/unskilled	-0.044	-5.047	0.006	0.534	-0.004	-0.375	0.027	0.505
Mean native age / 100	-1.578	-5.178	-0.156	-0.673	-0.082	-0.396	0.739	1.219
Mean immigrant age / 100	-0.033	-0.510	0.177	3.670	0.063	1.392	0.083	1.054
M_1	12.858 p = 0.000		-4.489 p = 0.000		-4.685 p = 0.000		-2.049 p = 0.040	
M_2	11.496 p = 0.000		0.272 p = 0.785		0.515 p = 0.606		0.379 p = 0.705	
W_1	χ^2_5=313.642 p = 0.000		χ^2_5=351.445 p = 0.000		χ^2_5= 14.312 p = 0.014		χ^2_5= 9.853 p = 0.080	
W_2	χ^2_{17}= 234.676 p = 0.000		χ^2_{17}= 356.959 p =0.000		χ^2_{17}=715.994 p = 0.000		χ^2_{15}=220.905 p = 0.000	
S							χ^2_3= 1.833 p = 0.608	
Sample size	306		306		289		255	

Notes:

M_1 is a test for first-order serial correlation, asymptotically distributed as a standard normal

M_2 is a test for second-order serial correlation, asymptotically distributed as a standard normal

W_1 is a Wald test for joint significance of the reported regressors

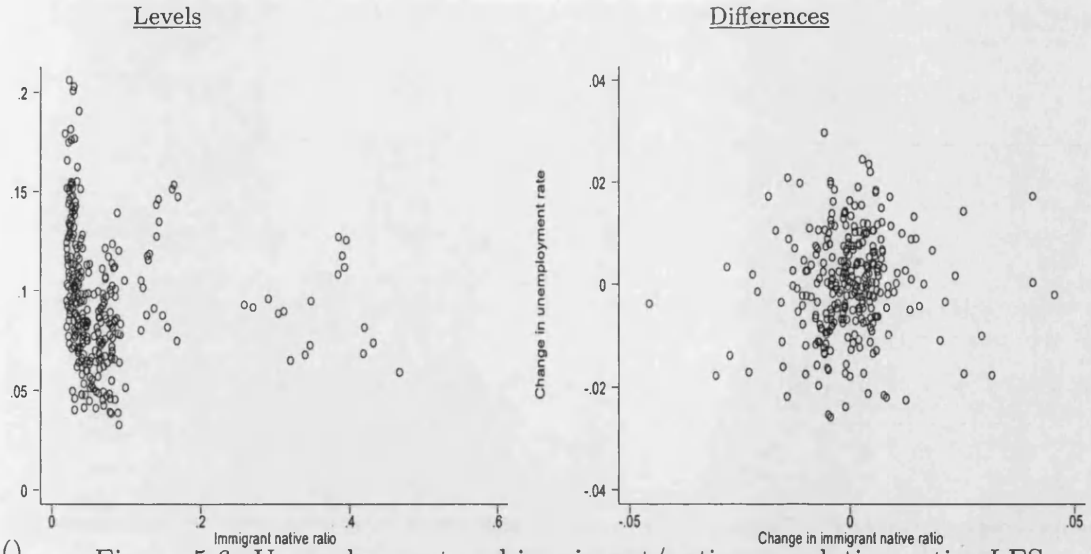
W_2 is a Wald test for joint significance of the unreported time dummies

S is a χ^2 test of the overidentifying restrictions implied by choice of instruments underlying IV estimates

controls we have just discussed. Although using different data at different frequency over a different period and looking only at unemployment of natives, the qualitative picture is remarkably similar. OLS regression shows a slight negative relationship between unemployment and immigrant native population ratio. We have illustrated this relationship in the left panel of Figure 5.6. Removing persistent correlated effects by within groups estimation or differencing switches the sign of the relationship. Immigration is now associated with a positive increase in unemployment, although, as before, it is not significant. The relationship between changes in the two variables is shown in the right hand panel of Figure 5.6.[9]

As before, these estimates may be compromised by the possible simultaneity between immigrant inflows and positive economic shocks, leading to an underestimate of the impact in simple differences. In addition, the possible presence of measurement error in the LFS immigrant flows makes the case for instrumental variables estimation even more

[9]In this figure, as in all figures in this section involving changes, year means of changes are subtracted from the data before plotting to focus attention on the cross sectional pattern of changes which drive the results.



() Figure 5.6: Unemployment and immigrant/native population ratio: LFS

convincing than for census data. Using lagged immigrant concentrations as instruments in the differenced equation increases the size of the estimated effect,[10] as we would expect. Nonetheless the final and most robust of these estimates is smaller than the census-based effect and statistically not significant. Therefore, the hypothesis of no effect can not be rejected. The value of the coefficient is modest and in line with the small size of effect typical of studies in other countries such as the US. An increase in immigration amounting to one per cent of the native population would lead, according to this result, to an increase of 0.18 percentage points in the native unemployment rate.

Distinguishing between different skill- and demographic groups

As already noted, one of the advantages of using LFS data is the ability to analyse effects on different skill groups separately. Table 5.4 reports separate regressions for unemployment among skilled, semiskilled and unskilled workers. The associated data is presented graphically in Figure 5.7. All effects are positive but individually statistically significant only for the semiskilled.[11]

Separating the workforce into demographic groups as in Table 5.5 also reveals estimated effects of similar sign and modest size, though consistently insignificant statistically. There is no strong evidence here that men or women are particularly harmed. Nor is it evident that minorities - defined here as immigrants arriving before 1981 - suffer specifically.

[10]Two- and three-period lags are used as instruments.

[11]Even this is below the critical point for the maximum of three independents t values, suggesting that the evidence for any effect is not strong.

Table 5.4: Effect of immigration on unemployment by skill group

LFS 1983-2000

	IV, Differences					
	Skilled		Semiskilled		Unskilled	
Variable	Coeff	*t* value	Coeff	*t* value	Coeff	*t* value
Immigrant-native ratio	0.104	0.915	0.390	2.219	0.026	0.112
ln skilled/unskilled	-0.084	-0.768	-0.247	-1.343	-0.233	-0.997
ln semiskilled/unskilled	-0.023	-0.529	0.090	1.269	0.003	0.036
Mean native age	0.437	0.869	0.706	1.032	-0.099	-0.086
Mean immigrant age	-0.052	-0.475	0.312	2.953	-0.093	-0.669
Mean skilled native age	0.089	0.850				
Mean semiskilled native age			0.486	0.642		
Mean unskilled native age					0.116	0.238
M_1	-4.968 p = 0.000		-2.141 p = 0.032		-4.240 p = 0.000	
M_2	0.186 p = 0.852		0.944 p = 0.345		-0.632 p = 0.527	
W_1	χ^2_6=6.739 p = 0.346		χ^2_6=14.450 p = 0.025		χ^2_6=5.536 p = 0.477	
W_2	χ^2_{15}= 200.615 p = 0.000		χ^2_{15}= 246.459 p =0.000		χ^2_{15}=60.992 p = 0.000	
S	χ^2_3= 1.187 p = 0.756		χ^2_3= 0.714 p = 0.870		χ^2_3= 0.353 p = 0.950	
Sample size	255		255		255	

Notes:

As for Table 5.3

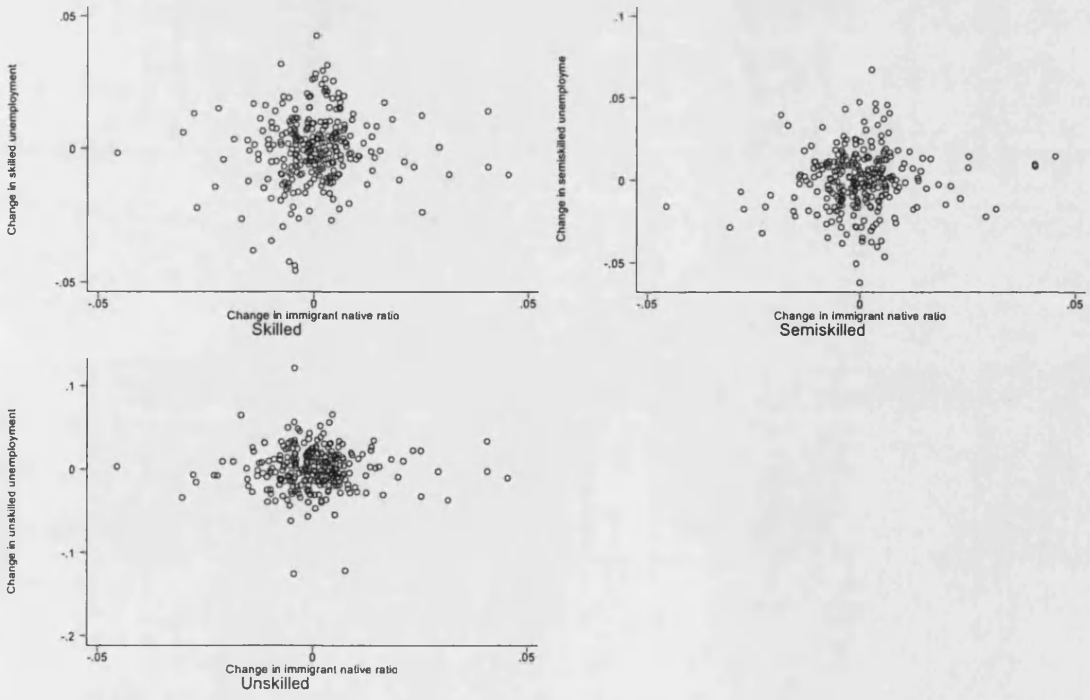


Figure 5.7: Unemployment - Immigration, different skill groups.

Table 5.5: Effect of immigration on unemployment by demographic group

LFS 1983-2000

	IV, Differences					
	Male		Female		Minority	
Variable	Coeff	*t* value	Coeff	*t* value	Coeff	*t* value
Immigrant-native ratio	0.198	1.206	0.154	1.330	0.071	0.047
ln skilled/unskilled	-0.277	-1.670	-0.154	-1.311	-1.856	-1.234
ln semiskilled/unskilled	0.018	0.273	0.041	0.873	0.250	0.416
Mean native age	1.421	1.146	0.346	0.475	7.790	1.131
Mean immigrant age	0.093	0.945	0.073	1.041	-0.772	-0.858
Mean male native age	-0.406	-0.486				
Mean female native age			0.023	0.039		
M_1	-2.006 p = 0.045		-2.886 p = 0.004		-2.314 p = 0.021	
M_2	0.621 p = 0.534		-0.449 p = 0.654		-1.719 p = 0.086	
W_1	χ^2_5=9.771 p = 0.135		χ^2_6=5.511 p = 0.480		χ^2_6=3.102 p = 0.684	
W_2	χ^2_{15}= 253.392 p = 0.000		χ^2_{15}= 141.670 p =0.000		χ^2_{15}=8.185 p = 0.916	
S	χ^2_3= 1.111 p = 0.774		χ^2_3= 2.259 p = 0.521		χ^2_3= 0.128 p = 0.988	
Sample size	255		255		255	

Notes:

As for Table 5.3

Table 5.6 separates the population into three age groups and estimates employment effects for each. The largest effect is for the oldest group but even here the coefficient is only on the margin of conventional statistical significance.

Finally, Table 5.7 returns to the effect on total unemployment, but disaggregates the immigrant inflow according to its source and gender. On the whole, these estimates are very imprecise and give no strong indication that immigration from particular source areas or of particular genders have more deleterious effects on native employment than do others.

In none of these specifications have the dynamics of the relationship been explored (as suggested in Section 5.3.2). We have been unable to find statistically reliable and well determined estimates of dynamic specifications and have therefore refrained from commenting on differences between short run and long run effects[12]. We note however that considerations of economic theory suggest that long run adjustments to immigration are likely to lower the magnitude of effects and that the estimates here, as hybrids of long and short run impact, are likely to overestimate long run responses.

[12]Section 5.7.2 reports results from the dynamic specification.

Table 5.6: Effect of immigration on unemployment by age

LFS 1983-1999

	IV, Differences					
	Age 20-35		Age 26-50		Age 51-65	
Variable	Coeff	*t* value	Coeff	*t* value	Coeff	*t* value
Immigrant-native ratio	0.207	1.463	0.070	0.366	0.292	1.961
ln skilled/unskilled	-0.134	-0.950	-0.335	-1.766	-0.089	-0.602
ln semiskilled/unskilled	-0.017	-0.302	0.065	0.853	0.032	0.540
Mean native age	0.931	1.446	0.940	1.084	-0.350	-0.515
Mean immigrant age	0.160	1.906	-0.026	-0.234	0.009	0.100
M_1	-3.773 p = 0.000		-2.310 p = 0.021		-3.871 p = 0.000	
M_2	1.340 p = 0.180		0.360 p = 0.719		-1.398 p = 0.162	
W_1	χ^2_5=12.392 p = 0.030		χ^2_5=4.527 p = 0.476		χ^2_5=9.836 p = 0.080	
W_2	χ^2_{15}= 297.494 p = 0.000		χ^2_{15}= 48.544 p =0.000		χ^2_{15}=86.942 p = 0.000	
S	χ^2_3= 3.835 p = 0.280		χ^2_3= 1.797 p = 0.616		χ^2_3= 0.234 p = 0.972	
Sample size	255		255		255	

Notes:

As for Table 5.3

Table 5.7: Effect of immigration on total unemployment by gender and source of immigration

LFS 1983-1999

	IV, Differences	
	Coefficient	*t* value
Gender of immigration		
Male	-0.207	(0.364)
Female	0.283	(0.678)
Sample size	255	
	Coefficient	*t* value
Source of immigration		
New Commonwealth	-0.057	(0.155)
Ireland	2.616	(1.768)
Other European Union	-0.255	(0.176)
Other	0.175	(0.683)
Sample size	255	

5.5.2 Wages

We now turn to the analysis of wages. We would like to interpret the results we present here with care, as the data for computing wage averages from the LFS is rather small - wages are firstly only available over the period between 1992 and 2000. Secondly, wage information is available for each individual at one or at most two interviews during the

course of the survey.

Table 5.8: Effect of immigration on wages
LFS 1992-2000

	Levels		Differences	
	OLS	Within groups	OLS	IV
Coefficient	0.644	0.863	0.159	1.869
t value	(5.049)	(1.550)	(0.222)	(2.184)
Sample size	153	136	136	102

Table 5.8 reports estimates regarding effects on wages. Figure 5.8 presents the data graphically in levels and differences. Estimates based on OLS show positive wage effects which become even larger when based on instrumental variables techniques. According to the most robust estimate, an increase in immigration amounting to one per cent of the non-immigrant population would lead to just under a two per cent increase in average non-immigrant wages.

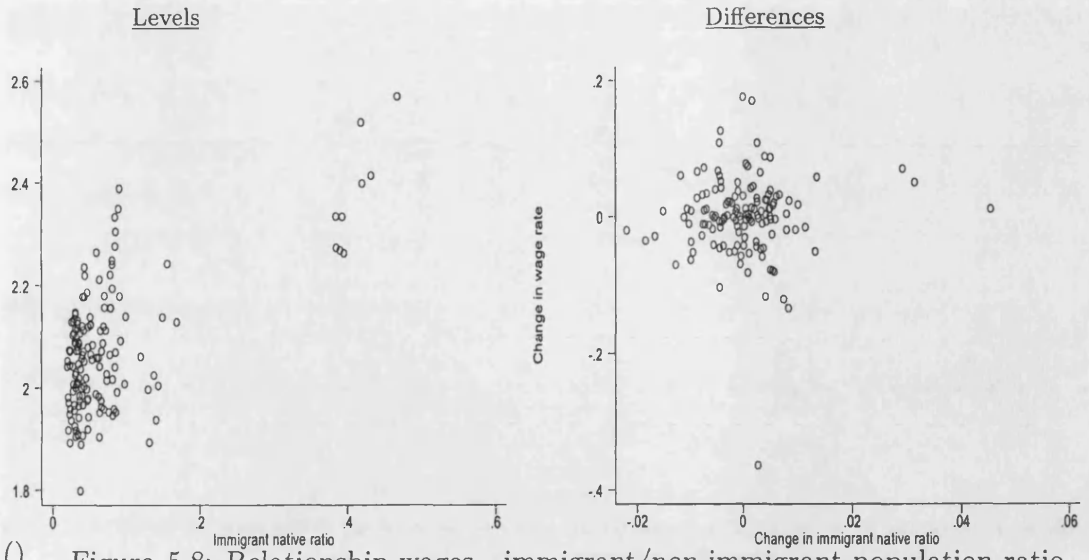
Table 5.9: Effect of immigration on wages by skill group
LFS 1992-2000

	IV, Differences		
	Skilled	Semiskilled	Unskilled
Coefficient	2.163	1.145	2.216
t value	(1.921)	(1.014)	(1.655)
Sample size	102	102	102

In table 5.9 we report coefficients for different skill groups. These estimates are less precise but effects are similarly signed in all three groups considered.

5.6 Conclusion

In this paper, we analyse the impact of immigration on labour market outcomes of native workers. Our analysis is the first for the UK. We commence by discussing the theoretical background, which suggests that there are realistic routes by which immigration can affect labour market outcomes but the absence of any long run impact is by no means implausible or inconsistent with theory for the case of an open economy with a large heterogeneous traded goods sector such as the UK.



() Figure 5.8: Relationship wages - immigrant/non-immigrant population ratio

The main result of the empirical analysis is that there is no strong evidence of large adverse effects of immigration on native employment or wages. In this respect our findings are consistent with empirical results from existing international research. There is some weak evidence of negative effects on employment but for most groups of the population it is impossible to reject the absence of any effect with the data used here. Insofar as there is evidence of any effect on wages, it suggests that immigration enhances native wage growth.

We have drawn attention to many weaknesses in the available data and conceptual problems in the empirical analysis all of which should urge caution before drawing strong conclusions. Nonetheless it seems to be fair to conclude that on current evidence fear of large and negative employment and wage effects on the resident population are not easily justifiable grounds for restrictive immigration policy. The perception that immigrant take away jobs from natives, thus contributing to large increases in unemployment, or that immigrants depress wages of native workers, do not find confirmation in the analysis of data laid out in this report.

We see our analysis as a first exploration of the available data evidence of the UK. Our analysis has identified a number of problems that are worth study in future research and possibly with future data sources. The arrival of the 2001 census will constitute a significant improvement of the available data base, allowing additional analysis of migration impact over the last decade.

The case of London is worth further study. Immigrant concentration in London as a whole far exceeds that elsewhere in any other city of the UK. Concentration and

inflows of immigrants into London also differ widely according to area. It is not unlikely that across areas, immigration has had economic effects on the resident population - a possible regularity which is only detectable with data that allows a breakdown according to smaller geographical units within the Greater London area.

Another avenue for future research is to investigate directly other dimensions through which immigration can affect the local economy, such as growth and output composition.

5.7 Appendix

5.7.1 Data Creation and Related Problems

The LFS has been carried out yearly from 1983 to 1991 and quarterly from 1992 onwards. To obtain aggregate information at regional level (the smallest geographical unit available), we create population numbers of the quantities of interest summing the (population weighted) number of individuals falling in the specific category for each region and each year. These quantities include number of natives and immigrants, broken by age, gender, country of origin and skill. This allows to have a set of quantities, reflecting the population composition, with which we can derive the ratios used in the analysis, such as the immigrants/population ratio or the unemployment rate of natives (defined as the ratio between the unemployed and the total labour force).

As it was mentioned in the main section, survey data may be characterised by very small sample sizes when analysing specific groups in the population (like immigrants, and when we want to distinguish them by gender and year of arrival). This is due to the fact that immigrants represent a small fraction (9%) of the population (LFS 2000) and that their geographical distribution in the UK appears to be very uneven (about 60% of immigrants of working age are concentrated in the Greater London and South East regions, against 29% of natives).

To give an idea of how small the sample size for certain groups and regions can be, we present a summary table from the LFS (second quarter of 2000), containing information by region on the sizes of the total sample and of some sub-samples of immigrants.

In some regions the number of observations relative to immigrants is less than a hundred. If we break the sample further, for instance because we want to focus our research on younger immigrants (column 4) or ethnic minority immigrants who arrived before or after 1981 (column 5 and 6), we further reduce the sample size into numbers that prevent from obtaining stable estimates.

Table 5.10: Sample Size by Region

LFS 1983-1999

Region	Total Sample	Immigrants		Ethnic Minority Immigrants	
		All	Less than 35	Before 1981	After 1981
Tyne and Wear	1635	53	31	6	16
Rest North Region	2978	69	32	10	14
S.Yorkshire	1913	66	22	24	16
W.Yorkshire	3129	262	80	104	78
Rest of Yorks & Humbers	2461	77	30	12	6
E.Midlands	5974	337	96	112	56
East Anglia	3138	200	78	14	25
G London	9247	2896	1054	807	919
Rest of SE	15916	1321	434	222	219
S.West	6995	391	127	39	52
W.Midlands	3537	466	134	225	134
Rest of W.Midlands	4057	122	32	25	10
Gt.Manchester	3523	251	94	70	74
Merseyside	1902	47	16	5	11
Rest of North West	3211	135	52	30	29
Wales	4076	129	58	18	34
Scotland	7839	321	147	36	37

5.7.2 Dynamic Estimates

Similarly to Table 5.3, Table 5.11 presents estimates of effects on total native unemployment based on a dynamic specification. The estimated effects of present and lagged immigrant shares are hardly significant in any of the specifications reported below. The addition of lagged variables clearly results in lower precision of the estimated coefficients. This prevents us from drawing reliable conclusions on differences in long and short run effects.

Table 5.11: Effect of immigration on unemployment (dynamic specification)

LFS 1983-2000

	Levels				Differences			
	OLS		Within groups		OLS		IV	
Variable	Coeff	*t* value	Coeff	*t* value	Coeff	*t* value	Coeff	*t* value
Unemployment(t-1)	0.841	26.89	0.531	9.77	-0.351	-5.90	0.580	0.04
Immigrant-native ratio	0.078	1.28	0.117	1.94	0.076	1.11	-1.513	-0.07
Immigrant-native ratio(t-1)	-0.096	-1.57	-0.016	-0.25	0.112	1.73	0.705	0.05
ln skilled/unskilled	-0.026	-2.77	-0.028	-2.36	-0.026	-2.44	0.608	0.15
ln skilled/unskilled(t-1)	0.019	1.99	0.004	0.39	-0.016	-1.60	1.259	0.09
ln semiskilled/unskilled	-0.007	-0.69	0.001	0.11	-0.001	-0.08	-0.367	-0.15
ln semiskilled/unskilled(t-1)	0.004	0.47	0.015	1.47	0.012	1.20	-0.398	-0.09
Mean native age / 100	-0.384	-2.30	-0.185	-0.87	-0.033	-0.19	-2.173	-0.15
Mean immigrant age / 100	-0.015	-0.44	0.090	2.07	0.054	1.50	0.067	0.13

Chapter 6

Ethnic Concentration, Prejudice and Racial Harassment of Minorities in the UK

Abstract:

In this chapter, we analyse the association between ethnic minority concentration, hostile attitudes towards minorities, and the probability of ethnic minorities experiencing racial hostility. Our main focus is on the relationship between ethnic concentration of minorities on the one side, and hostile attitudes as well as acts of racial harassment on the other. Our approach recognises the role precautionary behaviour may play in distorting this link. Other than much of the existing literature, we understand the formation of hostile attitudes and the realisation of acts of racially motivated violence as two distinct processes. We develop a general empirical model that subsumes many of the existing theories. We estimate a reduced version of that model, which allows us to derive conclusions about the relationship between racial harassment and precautionary behaviour of minorities, and ethnic concentration and hostile attitudes. Our data sources are the fourth National Survey of Ethnic Minorities for the UK and the 1981 and 1991 UK Census.

6.1 Introduction

Over the last 5 decades, Europe has experienced an unprecedented inflow of immigrants of distinguishably different ethnic composition leading to the existence of substantial communities of ethnic minorities, distributed within countries with spatially heterogeneous concentrations. Acts of intimidation and harassment aimed at ethnic minority individuals are quite commonly reported. Interracial conflict bears high social costs, and discourages long term integration. It often manifests itself through social exclusion and deterioration of welfare of the ethnic groups subjected to it (for instance, see Karlsen and Nazroo [2002]). Hostility towards minorities of any form may seriously affect the process of social and economic integration of immigrant minorities and their offsprings. Maintenance of good ethnic relations has therefore been a prime motivation of race relations and immigration policies. The persistence of racial harassment experienced by the resulting minority communities is nonetheless a continuing problem.

One key question for social scientific enquiry is the circumstances under which racial harassment and racially motivated crime occur. Below we summarise existing theory and seek to set it within a framework distinguishing different components of the processes generating harassment. Within this model, we discuss the possible identification of different theories. We then estimate an identifiable version of this model, which focuses on the relationship between individual and demographic characteristics, and the incidence of racial harassment. In particular, we investigate the association between racial harassment and ethnic concentration. We expect ethnic concentration to be associated with the incidence of harassment in several ways. Firstly, ethnic context affects the probability of minority individuals meeting majority individuals. Secondly, it affects the attitudes of the majority population. Thirdly it affects the probability of hostility finding expression in acts of racial harassment.

Fears of racial harassment, resulting in precautionary behaviour, may distort this relationship. We model the interdependence of precautionary behaviour and racial harassment, and we analyse the association between precautionary behaviour and ethnic concentration. Finally, theoretical and empirical literature implicitly assumes that ethnic concentration relates to hostile attitudes and racial harassment in the same way. Instead, we allow these as two independent processes, that may affect each other, and that may be related to ethnic concentration in a fundamentally different way. Our

empirical analysis supports this view.

Racially motivated crimes represent the most extreme form of victimisation against ethnic minorities. However, there are less violent but nonetheless socially disruptive behaviours against these communities. We analyse in this paper acts of racial harassment, which still express hostility and inter-racial tensions and may affect ethnic minorities' welfare and socio-economic integration. We base our investigation on the Fourth National Survey on Ethnic Minorities (FNSEM), which has been collected between 1993 and 1994. This survey contains a "boost" sample of ethnic minorities and a random sample of white natives, and provides a wide range of socio-economic information on respondents. In addition, the FNSEM contains specific information on different forms of racial harassment and abuse experienced by ethnic minorities. The FNSEM also contains information on how ethnic minority individuals take precautions, and change some of their habits as a response to racial harassment.

The structure of the paper is as follows. In the next section, we discuss some theories that explain racial conflict. In Section 6.2 we structure this discussion into a formal model encapsulating features of the main theories. In Section 6.4, we develop the empirical model that we use for estimation, and discuss identification. Section 6.5 introduces the data. Finally, in Section 6.6 we present the results, and provide a discussion in Section 6.7.

6.2 Theories

Harassment occurs when there is a meeting between an ethnic minority individual and a prejudiced white person who chooses to harass. There are therefore three elements to such events: the hostility, the meeting and the decision to express hostility aggressively. We can theorise fruitfully about each of these aspects, drawing on papers in this literature that review different theories on the formation of racial harassment or racial prejudice. Green et al. [1998], for example, provide a useful categorisation of theories which we draw on below. We restrict our discussion here to the essential features of some of these models and their empirical implications, and refer the reader to the literature for more details.

6.2.1 Formation of hostile attitudes

The first element, the formation of hostile attitudes between ethnic groups is the subject of a large theoretical literature. An influential strand emphasises group conflict as at the heart of patterns of hostility. *Realistic group conflict theory* suggests that racial prejudices derive from "a threat to real resources and accepted practices" of the majority population posed by minorities (Bobo [1983]). The *power-threat hypothesis* (Blalock [1956], Blalock [1957], and Blalock [1967]) says that intolerance of the white majority population is due to minorities being seen as economic competitors and as a challenge to social and political dominance of the majority population. Individuals from the white majority aim to keep a "social distance" with ethnic minorities. The level of intolerance will increase as this distance is threatened by growing concentration of ethnic minorities, suggesting therefore an increasing relationship between racial prejudice or racism, and ethnic minority concentration.

In linking racial prejudice to the competition for scarce resources, group conflict theories come possibly closest to economic explanations for prejudice and opposition towards immigration. These economic theories are based on equilibrium models that predict adverse effects for groups that compete most intensively with newcomers in the local labour market. Scheve and Slaughter [2001], Gang et al. [2002] among others analyse the determinants of individual preferences over immigration policies in the US and Europe respectively. In these papers, an empirical association between labour market status and attitudes is established and argued to be consistent with a determining role for labour market competition[1] In essence this approach suggests hostility which is not so much related to the ethnicity of the minority population as to the threat any newcomers impose to sharing of resources perceived as finite, as well as to social and political hegemony. It predicts higher sensitivity of the majority population towards any threats of economic and political hegemony in times and at places where competition for economic resources is most intense. Empirical implications are that indicators that reflect economic hardship at a geographical level should be positively correlated with the intensity of prejudice, or acts of harassment.

It is not only economic competition however that can establish grounds of interracial hostility, but also the broader collective threat to their established social and political

[1]In this context, Chapter 5 evaluates one aspect of such a competition, with an analysis of the impact of immigration on the local labour market.

prerogatives that the majority population perceives as coming from the minority population (Blumer [1958]). According to the theory of *defended neighbourhoods*, hostility towards ethnic minorities is based on a fear of loss of social identity. According to Rieder [1985] and De Sena [1990], residents in ethnically homogeneous neighbourhoods define their identity through the exclusion of other ethnic groups. Racism and racial harassment are according to this theory highest in areas where one ethnicity has been dominant for a long time, but suddenly experiences inflows of a new group or is threatened by encroachment. This hypothesis suggests that it may be the *change* in ethnic composition that is a catalyst for racist prejudice and action (Green et al. [1998]).

Such processes may be moderated by induced population flows. As ethnic minority density increases, some whites adapt to integration, whereas the most hostile individuals may leave the neighbourhood after their attempts to stop the minority inflow fail (Kinder and Mendelberg [1995]). This may then lead to an improvement in race relationships. A similar prediction is offered by the literature on *"white flight"* (Clark [1993]), where intolerant whites move out of the neighbourhood when ethnic density reaches a *"tipping-point"*. Such ideas suggest we may expect to find nonlinearities in the relationship between ethnic balance and indicators of hostility.

Racial prejudices are sustained by acceptance of stereotyping and misrepresentation of minority practices and characteristics. The *contact hypothesis* draws attention to the weakening impact of social contact with minorities (Rothbart and John [1993]). Since the frequency of such contact increases with minority concentration this may provide a countervailing force through which higher minority density reduces hostility.

6.2.2 The meeting

These ideas all offer useful insight into the determination of attitudes in the majority community. However, the existence of hostility towards minorities does not necessitate the incidence of harassment.

For harassment to occur, it is necessary that majority and minority individuals come into contact. This has been recognised by the *random interaction hypothesis* (Blau [1977]), which suggests that inter-ethnic frictions are proportional to the amount of interracial contact in the area. Other things being equal, the frequency with which ethnic minority individuals encounter whites decreases as ethnic minority concentration increases in an area, meaning that there are less opportunities for harassment to occur.

For a fixed level of prejudice, the number of racial incidents (as opposed to the probability of a minority individual experiencing harassment) should therefore reach its maximum when the two groups reach the same size (and power) and then decline as the size and power balance becomes increasingly in control of the minority community.

There is good reason to think interaction may not be random however. The frequency with which minorities visit locations where they are likely to encounter whites may well be affected by perceived levels of prejudice. Also, while, on the one hand, weak racists may avoid areas where they are likely to meet ethnic minorities, on the other hand, extreme racists within the majority community may seek out opportunities for confrontation. Local social characteristics such as density of housing, availability of public space and so on, may also affect the nature and intensity of social practices which bring different ethnic communities into contact.

6.2.3 Aggression

Finally, it is necessary that prejudice find violent or aggressive expression in an act of harassment. Intensity of hostility presumably predisposes majority individuals towards harassment and to that extent the theories discussed above may also serve as theories of harassment. Harassment, however, is not simply a more extreme form of prejudice but a particular mode of manifestation. For any given level of hostility in white attitudes, the likelihood of this translating into harassment may itself depend upon the circumstances of the encounter and the characteristics of the potential perpetrator and victim. Specifically, the tendency to harass, as a conscious choice of the harasser, may be expected to depend not only on the strength of the desire to harass but also on the costliness of harassment to the perpetrator and on the availability of substitutes.

The *power-differential hypothesis* points out that minorities can protect themselves better in neighbourhoods at high ethnic density and therefore white perpetrators will fear more in terms of reprisal or punishment (LeVine and Campbell [1972], and Levin and McVine [1993]). This suggests there may be "safety in numbers" for minority individuals who may be less likely to suffer harassment, albeit that white hostility may be greater, in areas of higher density.

The likelihood that whites choose to express hostility through harassment may also depend upon the availability of other means of releasing dissatisfaction. More affluent, more articulate and more educated whites may, for example, be less inclined to resort

to violent expression of discontent against minorities.

6.2.4 Implications

All these hypotheses establish a link between ethnic concentration on the one side, and expressions of racial intolerance on the other, and it is this link which has been most extensively explored in previous work. To briefly summarise the main implications, theories where competition for scarce resources underlies the conflict between minorities and majorities (group threat theories) predict a positive correlation between concentration of minorities and hostile attitudes, which may result in acts of racial violence. These theories also predict that conflict is harsher the more intense the competition for resources, i.e. the more unfavourable are the economic conditions.

An increasing intensity of conflict at least initially is also predicted by the tipping point hypothesis. Here however, out-migration of the most aggressive individuals may lead to a turning point in the relationship with increasing minority density. The defended neighbourhood hypothesis suggests that both the level and the change in minority concentration should matter to racial aggression. It is the sudden increase in minority concentration in areas previously unexposed to minorities which leads most strongly to conflict. The random interaction hypothesis is again predicting an increase in conflict with ethnic concentration. The argument here however is that an increase in concentration does not increase the intensity of the conflict, but the probability of encounters between individuals of the different communities. The contact hypothesis comes to similar conclusions, but here it is ignorance that creates aggression, and enhanced knowledge by contact that reduces aggression.

In contrast to the previous theories, which, at least initially, predict a positive relationship between racial aggression and ethnic concentration, the power differential hypothesis comes to opposite conclusions. Here racial aggression decreases with concentration as majorities find it harder to harass in areas where they may have to fear reprisal.

These theories are not exclusive, pointing as they do to effects which can coexist. To the extent that predictions are unambiguous and uncontested by the predictions of other theories, they can be tested. But where theories suggest counteracting effects we can aim only to estimate the balance between them.

6.2.5 Empirical studies

There are numerous studies that investigate several aspects of the formation of attitudes, as well as acts of intimidation directed at ethnic minorities. Many of these analyse the link to ethnic minority concentration, and attempt to explain emerging results within theories as those discussed above.

Early work by Studlar [1978] for the UK establishes a non-linear relationship between negative opinion on immigration, and immigrant concentration which sometimes, but not always supports the tipping point hypothesis. Kinder and Mendelberg [1995] investigate the relationship between prejudice against minorities and opinion on racial policy. Their study concludes that racial isolation enhances this link significantly. Green et al. [1998] investigate more directly the incidence of racially motivated crime, and in-migration of minorities as well as economic conditions. They find a rising racial crime rate when non-whites move into a particular area, but falling crime rates where non-whites have long resided. Krueger and Pischke [1997] investigate the link between crime against minorities and ethnic concentration for Germany, which has experienced a fast and large inflow of ethnic Germans from former Eastern countries and refugees from the former Yugoslavia in the 1990's. Their evidence suggests that high concentrations of minorities in areas of Germany have caused a rise in hostility and criminal acts against minorities.

To summarise this literature, although some studies do find a positive relationship between measures of hostility and ethnic concentration, this link seems to be by no means undisputed. Below we draw on theories discussed in this section, and develop a more complete framework for analysis than many of the previous studies have employed.

6.3 Ethnic Concentration and Harassment

We formulate a model to investigate the mechanism behind the relationship between ethnic composition and racial harassment at a local level. Our analysis will not explicitly attempt to test one of the above mentioned theories against another; we believe that each of these hypotheses contributes in some way to explain racial aggression. We will however develop a model that subsumes, and is motivated by a variety of theoretical explanations. The central but not sole focus of our analysis will be the relationship between both attitudes and harassment on the one hand, and ethnic concentration on the other. It is this relationship where the literature makes most explicit predictions, as

we have discussed above.

Our main aim is the estimation of an inclusive harassment equation capturing the impact of individual and contextual variables through *all* channels, including influences on precautionary activities, internal migration and so on. While it would be desirable to separate effects arising through different channels it is not easy to think of plausible restrictions allowing identification.

It is also unlikely that all individuals belonging to a minority population are equally affected by racially motivated aggression. Individual-specific features may explain a lot of variation in being victimised. For instance, male minority individuals may be more or less exposed to racial harassment than females, as a consequence either of harassers' attitudes or of differences in the social and occupational situation.

Our approach tries to take account of some of these issues. We define a rate of arrival of harassment incidents λ_i, which is the product of the probability of meeting a white racist μ_i and the probability of that person choosing to harass or insult ν_i given the encounter. Both these probabilities depend in principle on individual specific characteristics, as well as local minority concentration, and the change in local minority concentration, both through the impact on white attitudes and directly:

$$\lambda_i = \mu_i \nu_i \tag{6.1}$$

$$\mu_i = f(A_{j(i)}, \pi_{j(i)}, X_i, Z_{j(i)}, n_i) \tag{6.2}$$

$$\nu_i = g(A_{j(i)}, \pi_{j(i)}, X_i, Z_{j(i)}, n_i) \tag{6.3}$$

$$\Rightarrow \lambda_i = h(A_{j(i)}, \pi_{j(i)}, X_i, Z_{j(i)}, n_i) \tag{6.4}$$

where $j(i)$ is the area in which the individual lives, A_j is the propensity to racism of white individuals in the jth area, π_j denotes the concentration of ethnic minorities in the jth area[2], Z_j denotes other characteristics of the jth area, n_i is the degree of precaution taken against meeting white racists and X_i denotes other individual characteristics. Attitudes themselves depend on area characteristics including π_j

$$A_j = F(\pi_j, Z_{j(i)}) \tag{6.5}$$

Notice that this formulation acknowledges different sources of racial aggression, as put forward by above mentioned theories. The random interaction hypothesis concentrates on the fact that λ_i increases with ethnic concentration because μ_i increases. The

[2]Note that lagged values of concentration may also be important

power threat hypothesis is not explicit about μ_i, but predicts that ν_i increases with ethnic concentration. Similarly the defended neighbourhood hypothesis suggests that ν_i increases if the change in ethnic concentration is large within a low concentration areas, but does not specifically concern μ_i. Selective out-migration leads both μ_i and ν_i to decrease. Finally, the power differential hypothesis, again, refers to ν_i rather than μ_i.

Assimilating points from the multiplicity of theories discussed earlier, we expect $\partial f/\partial A$ to be ambiguous: the presence of more racists in the area means there are more of them to be met but that may be counterbalanced if white racists prefer to avoid contact with ethnic minorities and therefore shy away from contact. We also expect $\partial f/\partial \pi$ to be positive (encounters with whites are more frequent in areas of low ethnic minority density) and $\partial f/\partial n$ to be negative it is in the nature of the sort of precautions we consider that they reduce frequency of confrontation). The probability of the white racist to harass, ν_i should be increasing in prejudices A, since prevalence of racism may be expected to create a cultural environment in which harassment is more acceptable. The sign of $\partial g/\partial \pi$ is expected to be ambiguous since high minority density may either aggravate white racists or make them more defensive for all of the many reasons outlined. Finally, we expect $\partial g/\partial n$ to be negative since precautions may be directed at avoidance of confrontation as much as avoidance of encounters with racists. As for attitudes, we expect $\partial F/\partial \pi$ to be ambiguous for the many reasons outlined in the theories discussed in previous sections.

On the whole theories are quiet on the way minority individuals may react to perceived threat from majorities. Precautionary behaviour such as going out less frequently, making the home safer and so on, is a decision which will be motivated both by the prevalence of harassment in the area and personal circumstances and characteristics which make harassment harmful,

$$n_i = G(\lambda_i, X_i, Z_{j(i)}) \tag{6.6}$$

with $\partial G/\partial \lambda > 0$.

A partially reduced form for this system relates harassment and precaution jointly to local characteristics, including local white attitudes, and personal circumstances

$$\lambda_i = L(A_{j(i)}, \pi_{j(i)}, X_i, Z_{j(i)}) \tag{6.7}$$

$$n_i = N(A_{j(i)}, \pi_{j(i)}, X_i, Z_{j(i)}). \tag{6.8}$$

The fact that $A_{j(i)}$ and $\pi_{j(i)}$ enter (6.6) only through λ_i implies cross-equation (pro-

portionality) restrictions on the way that these enter (6.7) and (6.8). These equations pick up effects of ethnic context π_j on harassment intensity given white prejudice. However, we can also substitute from (6.5) to develop a fully reduced form capturing total dependence of harassment on π_j incorporating its effect on white attitudes

$$\lambda_i = \Lambda(\pi_{j(i)}, X_i, Z_{j(i)}) \tag{6.9}$$

$$n_i = H(\pi_{j(i)}, X_i, Z_{j(i)}). \tag{6.10}$$

Structural identification of (6.4) and (6.6) requires exclusion restrictions in the harassment equation that are unlikely to be plausible. Characteristics which might encourage precaution without affecting harassment propensities directly might be those which affect the costliness to the victim of being harassed but in so far as these are observable by potential harassers it is difficult to justify the exclusion required. Racism may, for instance, encourage people to direct harassment at the most vulnerable. An example of a putative instrument might be presence of children since this might make parents particularly keen to avoid confrontation. However it cannot be ruled out that harassers themselves might be reluctant to insult in the presence of children and also that having children brings individuals into social interactions with other communities, at school for example, that provides opportunities for harassment to occur.

Restrictions which might allow separation of effects through μ_i and through ν_i are also difficult to imagine. Since several ambiguities have been identified in effects coming through $f(\cdot)$ and through $g(\cdot)$, the overall impact of white attitudes and of ethnic densities on harassment is difficult to sign.

We may worry about endogeneity of location choice $j(i)$ if our interest is in identifying effects conditional on fixed location. Moving house is after all one extreme form of precaution. This could motivate instrumenting $A_{j(i)}$ and $\pi_{j(i)}$ if suitable instruments exist but arguments for these tend to be tenuous.

6.4 Empirical Implementation

Let H_i denote the frequency of harassment. If harassment arrives at rate λ_i then the probability of being harassed k times is

$$\Pr(H_i = k|\lambda_i) = e^{-\lambda_i}\lambda_i^k/k! \tag{6.11}$$

We let $\ln \lambda_i = X_i\beta + \epsilon_i$ where X_i includes all relevant observed characteristics and ϵ_i captures unobserved influences on harassment propensity.

We observe a discrete indicator of precautionary activity n_i which we take to reflect a latent underlying disposition to precaution n_i^* where $n_i^* = X_i\alpha + \eta_i$ and η_i captures unobserved influences on precautions taken. We partition the range for n_i^* such that observed precaution falls into the dth of D observed categories if $\delta_{d-1} < n_i^* \leq \delta_d$ where $\delta_0 = -\delta_D = -\infty$.

Let the joint density of ϵ and η be denoted $p_{\epsilon,\eta}(\epsilon,\eta)$, the conditional density of ϵ given η be $p_{\epsilon|\eta}(\epsilon|\eta)$ and the marginal density of η be $p_\eta(\eta)$. Then the likelihood contribution for the ith observation is

$$\begin{aligned}\Pr(H_i = k, \delta_{d-1} < n_i^* \leq \delta_d | X_i) = \\ &\frac{1}{k!}\int_{\delta_{d-1}-X_i\alpha}^{\delta_d - X_i\alpha}\int_{-\infty}^{\infty} e^{-\exp(X_i\beta+\epsilon)}e^{(X_i\beta+\epsilon)k}p_{\epsilon,\eta}(\epsilon,\eta)\mathrm{d}\epsilon\mathrm{d}\eta \\ &= \int_{\delta_{d-1}-X_i\alpha}^{\delta_d - X_i\alpha}\left[\frac{1}{k!}\int_{-\infty}^{\infty} e^{-\exp(X_i\beta+\epsilon)}e^{(X_i\beta+\epsilon)k}p_{\epsilon|\eta}(\epsilon|\eta)\mathrm{d}\epsilon\right]p_\eta(\eta)\mathrm{d}\eta \quad (6.12)\end{aligned}$$

We choose a normal distribution for η and a conditional gamma distribution for e^ϵ

$$\begin{aligned} e^\epsilon|\eta, X_i &\sim \Gamma(e^{\psi\eta}, \zeta) && (6.13)\\ \eta|X_i &\sim \mathcal{N}(0,1). && (6.14)\end{aligned}$$

Here

- ψ captures correlation between harassment and precaution arising either from the influence of one on the other or from correlation in unobserved influences on the two. If $\psi = 0$ then the specification reduces to the combination of an independent ordered probit and negative binomial count model.

- nonzero ζ allows for two things. Firstly, it permits unobserved variation in harassment propensity λ_i independent of precautionary behaviour. Secondly, it divorces the mean and variance of the harassment process, allowing for "overdispersion" or "underdispersion" in the harassment equation relative to a Poisson model. If $1/\zeta = 0$ then the specification reduces to one in which harassment follows a Poisson process with unobservable influences perfectly correlated in the two latent specifications.

With such a specification, we can integrate within the bracketed component of (6.12) to derive[3]

$$\Pr(H_i = k, \delta_{d-1} < n_i^* \leq \delta_d | X_i) = \int_{\delta_{d-1}-X_i\alpha}^{\delta_d - X_i\alpha} \left[\frac{\Gamma(k+\zeta)}{\Gamma(k+1)\Gamma(\zeta)} \left(\frac{\zeta}{\zeta + X_i\beta + \psi\eta} \right)^{\zeta} \left(\frac{X_i\beta + \psi\eta}{\zeta + X_i\beta + \psi\eta} \right)^{k} \right] \phi(\eta) \mathrm{d}\eta \quad (6.15)$$

This formula involves only a single integral which we compute numerically (by Gauss-Legendre quadrature). In cases where precaution behaviour is unrecorded we integrate over the whole real line.

There are cross equation restrictions involving proportionality of coefficients on variables hypothesised to enter the precaution equation only through λ_i, namely those involving white attitudes and ethnic density. We report both unrestricted estimates and estimates on which these are imposed by minimum distance methods.

6.5 Data

The data we use for our analysis is survey based, and stems from the Fourth National Survey for Ethnic Minorities (FNSEM). The FNSEM is a cross-section survey collected between 1993 and 1994. It consists of a main sample of respondents belonging to ethnic minorities, and a reference sample of individuals belonging to the white majority population. In the survey, 59% of the ethnic minority sample was selected from wards where, according to the 1991 Census data, ethnic minorities represent at least 10 percent of the whole population. About 38% were selected from areas with ethnic concentration between 1 and 5 percent and the rest in areas with concentration of less than 1 percent. In contrast, the white reference sample was a random sample in the population.

One crucial advantage of the FNSEM is to provide our analysis with demographic information at ward level.[4] This allows to capture the wide diversity in the ethnic composition which characterises areas belonging to the same region. Consequently, it allows to have sufficient variation across different geographical units. According to the 1991 Census of population, in the UK, almost 80% of ethnic minorities live in the South

[3]See Cameron and Trivedi [1998].

[4]In the UK, a ward is the smallest geographical area identified in the Population Census. See Chapter 3 for more details on geographic ethnic composition.

East (mainly Greater London) and the Midlands regions. Inside these regions, however, ethnic concentration widely varies across smaller areas, such as wards.

In addition, the FNSEM contains extensive information on socioeconomic characteristics of the interviewees. It also focuses on ethnic and cultural issues, and, among these, on interracial relations. In particular, one section of the survey is dedicated to the potential "victimisation" experienced by ethnic minority individuals in the year previous to the interview. Various incidents are recorded, such as personal attacks, property damage, and insult and whether the victim believed such incidents were based on reasons due to race or colour. Furthermore, the white respondents' sample contains a set of questions on self-reported prejudice towards different types of ethnic minorities. These can be used to estimate the relationship between the average attitudes surrounding ethnic minority individuals and the probability that they are harassed.

We base our estimations on the indicator of a milder form of harassment, whether the respondent has been insulted "for reasons to do with race or colour". Information on more serious forms of harassment is available only for less than half of the victimised sub-sample and would heavily reduce sample size. Furthermore, the racial component in an insult should be relatively straightforward to determine [5]. Although less serious, the incidence of such milder forms of harassment is more common and still likely to have disruptive consequences on the degree of integration of minorities in the society.

Table 6.1: Harassment: Annual frequency of occurrence

Frequency	Percentage of sample
None	90.2
1	2.2
2	2.2
3	1.0
4	0.6
5	0.6
6 or more	2.9
Number of cases	4935

The data gives information both on whether or not the individual has been harassed and, if so, how often. We wish to make use of the latter information to ensure we take fullest account of differences in frequency of insult. However we also have to appreciate

[5]The wording of the question in the survey is as follows: "In the last twelve months, has anyone insulted you for reasons to do with race or colour? By insulted, I mean verbally abused, threatened or been a nuisance to you?"

that there is considerable bunching and rounding[6] in this data at higher frequencies as well as an imprecise category corresponding to a frequency too high to count. We therefore group this with all frequencies of 6 times or above, calculating the likelihood contribution appropriately. We provide frequencies in table 6.1.

Table 6.2: Precautions

Precautions	Percentage of sample
Avoiding going out at night	9.0
Making home safer	7.8
Visiting shops only at certain times	5.4
Avoiding going out alone	4.1
Stopping children playing	4.5
Avoiding white areas	3.1
Changing travel routes	2.5
Worshipping less frequently	1.9
Stopping going to pubs	1.8
Changing telephone number	1.4
Making business premises safer	1.3
Stopping use of public transport	1.2
Moving home	0.5
Moving school	0.2
Number of cases	2263

About half of the sample were also asked about precautions taken in response to concern about harassment. Fourteen different possible precautions were suggested, some more commonly undertaken than others. The full list is given in Table 6.2. We focus on the four most commonly taken of those potentially available to all respondents (including the childless) and add up the number of precautionary activities entered into as our measure (see table 6.3).

Table 6.3: Number of key precautions undertaken

Number	Percentage of sample
None	90.5
1	3.0
2	2.8
3	2.6
4	1.2
Number of cases	2263

Table 6.4 reports the means and standard deviations of variables that measure characteristics of the respondent, as well as attitudes and ethnic densities. About 11 percent

[6]There is a small isolated spike at 52 times a year, for instance, clearly corresponding to (roughly) once a week.

of the respondents reports being racially harassed in the year previous to the interview.

Table 6.4: Descriptive Statistics

Variable	Mean	St. D.
Male	0.488	0.500
Degree	0.113	0.317
A-level	0.208	0.406
Vocational	0.182	0.386
Age	38.704	15.155
Has Children	0.592	0.491
Caribbean	0.234	0.423
Indian	0.247	0.431
African-Asian	0.140	0.347
Pakistani	0.228	0.420
Bangladeshi	0.110	0.313
Chinese	0.040	0.196
Foreign Born	0.776	0.416
Living in London	0.402	0.490
Harassed	0.098	0.297
County White Attitudes	0.935	0.197
Ward Ethnic Density	0.331	0.218

In our estimations, we include three education dummies indicating whether individuals have university education, A-levels or vocational qualifications. Education is likely to influence the type of socio-economic environment in which the individual interacts. Therefore, it can affect the probability of being harassed. Attitudes and, in particular, propensity to harass may vary in different socio-economic environments. In addition, individuals with different qualifications may come in contact with white people to different extents. Moreover, education may also reflect different degrees of sensitivity to harassment.

The average age of minority individuals in the sample is 39 years. Age may be another determinant of the propensity to be harassed. Older individuals, for example, may tend to go out less or to go to places less frequented by white individuals. In addition, potential harassers may prefer to target certain cohorts rather than others.

We consider how the ethnic groups identified in the sample, namely Black Caribbean, Indian, African-Asian, Pakistani, Bangladeshi and Chinese may be victims of harassment to a different extent. Ethnicity may indicate the extent to which cultures differ from the English one and the extent to which different ethnic groups have integrated in English society. Racial abuse may be experienced particularly by individuals whose look and behaviour are perceived as radically different from those of the white majority population. Accordingly, we also look at different harassment experiences for ethnic

minority immigrants and native born ethnic minorities. Natives may tend to mix with whites more than immigrants. In the sample, 78 percent of ethnic minorities were born abroad.

We have information on spatial variation in certain dimensions of white prejudice and can therefore explore the extent to which frequency of harassment depends on the prejudicial attitudes of the white community in the area where the individual lives by including this as conditioning information. Our data here concerns broad expressions of prejudice against minorities. As Smith [1989], p.150, notes, it is possible that such "low-level" attitudes "provide a reservoir of procedural norms that not only tacitly inform routine activity, but are also available to legitimize more purposive, explicitly racist, practice." Specifically, to investigate this, we add a variable indicating the average attitudes against minority individuals at county level. Information about attitudes of white individuals towards minorities is available in the FNSEM. However, such information cannot be used as a regressor at ward level. Due to the survey sampling design described above, a large part of the white sub-sample live in different wards from ethnic minority respondents. This leaves too small a number of observations which would match the minority sample. The same is true for average attitudes at district level. Therefore, we use average attitudes at county level. How this variable is constructed is explained in the Appendix. The least prejudiced counties appear to be Cheshire, Gloucestershire, and Norfolk, whereas the most prejudiced are Northamptonshire, the West Midlands and Essex. The impact of ethnic context on attitudes of this type is investigated in many papers, including, for the UK, Dustmann and Preston [2001]. We should avoid thinking of the estimates including this measure of white attitudes as an estimate however of (6.7) and (6.8) rather than of (6.9) and (6.10), since the questionnaire responses on which the data is based can only hope to pick up a limited subset of relevant white attitudes.

Our main interest is in the impact of ethnic concentration at ward level and this information comes from the UK Census.

The sociological literature points to arguments for expecting both the level and the recent change in minority concentration to matter to white hostility (see discussion above). While the 1991 Census contains information on racial identities of the population this information is unavailable in the previous 1981 Census for the purpose of constructing information on dynamics of ethnic concentration. We therefore prefer to base our measure of ethnic density on the percentages of immigrants from South Asia

and the West Indies, which is a variable consistently available. These particular sources are the main geographical origin of ethnically different immigration to the UK (excluding only East Asia).

In table 6.5 we display percentages on the incidence of experiencing racial harassment, worries about being racially harassed, and precautions, for different quartiles of the attitude distribution and the distribution of ethnic concentration. The first panel of the table refers to attitudes. The first row suggests that the relationship between harassment as well as worries of being harassed, and hostile attitudes is inverse U-shaped, first increasing, but then declining. This pattern seems to be very similar with respect to precautions. It is surprising that harassment and worries of being harassed decrease at highest levels of hostility.

The lower panel of the table distinguishes between different quartiles of ethnic concentration. Here the incidence of harassment is clearly declining, with individuals in areas with highest concentrations reporting lowest incidences. Precautions seem to be more inverse U-shaped, increasing first, but then decreasing.

Table 6.5: Precautions and harassment, quartiles of hostile attitudes and ethnic concentration

	All	1st Q	2nd Q	3rd Q	4th Q
		Quartiles of hostile attitudes			
Insulted in last 12 months	10.56	12.14	10.23	13.49	9.83
Worried about being racially harassed	22.56	22.12	25.53	29.32	20.90
Avoiding going out at night	8.96	7.62	12.25	10.41	7.67
Making home safer	7.83	7.59	10.05	7.32	6.87
Visiting shops only at certain times	5.44	4.83	7.10	6.23	5.19
Avoiding going out alone	4.05	4.12	6.14	4.58	3.41
		Quartiles of Ethnic concentration			
Insulted in last 12 months	10.56	13.45	12.15	9.34	7.13
Worried about being racially harassed	22.56	25.0	25.12	25.16	14.72
Avoiding going out at night	8.96	8.10	10.90	10.0	6.88
Making home safer	7.83	7.16	8.51	9.53	6.14
Visiting shops only at certain times	5.44	5.07	7.27	6.02	3.41
Avoiding going out alone	4.05	3.	4.30	5.94	3.01
Number of cases	2263				

6.6 Results

Tables 6.6, 6.3 and 6.9 presents estimates from a variety of specifications. Tables 6.6 and 6.8 present independent estimates of the harassment and precaution equations under the

assumption $\psi = 0$. Table 6.9 presents joint estimates.

6.6.1 Harassment

Table 6.6 presents estimation results from the negative binomial count model, where we model the number of occurrences of racial harassment.

Table 6.6: Harassment: Independent Negative Binomial

Variable	Coeff	t	Coeff	t	Coeff	t	Coeff	t
Ethn.Conc	-3.221	-9.25	-2.982	-7.86	-5.551	-5.30	-5.903	-4.20
Ethn.Conc^2	.	.	.	.	6.778	2.87	6.875	2.14
ΔEthn.Conc	.	.	.	.	.	.	-0.602	-0.55
White Att.	.	.	-1.134	-7.04	2.194	1.79	3.615	2.26
White Att.^{2}	.	.	.	.	-1.860	-2.66	-2.361	-2.57
Unemp.Rate	1.641	3.44	1.461	2.87	1.706	3.09	1.711	2.22
Male	0.425	7.91	0.401	7.14	0.394	7.00	0.302	4.77
Age	2.364	1.94	2.390	1.88	2.301	1.79	2.172	1.47
Age^2	-3.655	-2.53	-3.762	-2.48	-3.661	-2.39	-2.672	-1.53
Has children	0.100	1.26	0.080	0.97	0.090	1.07	-0.018	-0.19
No. children	-0.045	-0.18	0.006	0.02	-0.019	-0.07	0.554	1.91
Degree	0.502	6.37	0.499	5.93	0.489	5.81	0.565	5.91
A level	0.377	5.49	0.438	6.22	0.443	6.16	0.441	5.17
Vocational	0.277	3.52	0.255	3.10	0.263	3.18	0.339	3.60
Immigrant	-0.471	-5.71	-0.460	-5.38	-0.458	-5.28	-0.587	-6.20
Caribbean	0.521	4.56	0.639	5.25	0.652	5.34	0.887	5.65
Indian	0.253	2.20	0.392	3.15	0.415	3.28	0.700	4.07
Afro-asian	0.658	5.78	0.785	6.31	0.804	6.36	1.117	6.46
Pakistani	0.470	4.29	0.611	5.19	0.605	5.08	1.032	6.43
Chinese	0.653	4.70	0.830	5.57	0.796	5.33	1.193	6.49
London	0.474	8.24	0.609	8.92	0.571	7.26	0.625	6.05
Const	-2.578	-10.44	-1.734	-6.03	-3.014	-5.48	-4.188	-5.93
ζ	1.470	31.97	1.480	30.38	1.481	30.27	1.495	26.46
Mean log-likelihood	-0.554		-0.558		-0.553		-0.589	
Number of cases	4935		4640		4640		3435	

All specifications condition on a set of individual observed characteristics, as well as a dummy variable for London. The first pair of columns presents results where we condition on a linear ethnic concentration variable, measured on ward level. The second set of columns conditions on white attitudes in addition. Columns 3 and 4 are the same specifications, but we allow for nonlinear relationships by adding squared terms on both variables.

There are several well determined demographic and socioeconomic effects. Men are more likely to suffer harassment. There is a nonlinear relationship with age typically peaking for individuals in their 30s. The more educated are more harassed. Furthermore, those born outside the UK are less likely to be harassed. These effects may come from the

different milieux frequented by persons with different characteristics or from differences in demeanour which attract or repel the attention of harassers. Different ethnic groups suffer harassment of differing intensity.

The main effect of interest is the role of ethnic density. The results point clearly towards lower harassment in areas of higher minority concentration, consistently with the random interaction story or with a "safety in numbers" effect as predicted by the power differentials hypothesis. There is some evidence of curvature with the marginal effect diminishing as minority concentration increases though it does not flatten off within ranges of ethnic density typically found in the data[7]. No evidence of any impact from the rate of change in ethnic density is apparent, contrary to the predictions of the defended neighbourhoods argument.

When we control for hostile white attitudes, the estimate of the ethnic concentration variable decreases slightly in absolute value.

The relationship to attitudes of white people in the same area also appears curved, initially rising with expressions of white racial prejudice but also flattening off or even turning down at higher levels. A negative impact is perhaps surprising but not beyond rationalisation. It may be that the sort of attitudes picked up by the questions on prejudice asked to white respondents are those which discourage contact with ethnic minorities rather than aggressive confrontation with them. We should also remember that what this measures is willingness to admit prejudice to interviewers and the sort of hostility which sustains harassment may be the sort of hostility which either does not recognise itself as prejudice or which is reluctant to admit it to interviewers.

Local unemployment seems to be associated with higher harassment, even conditional on white attitudes. The results in Dustmann and Preston [2001] (see also discussion in next section) point to no identifiable impact of white unemployment on prejudice or hostility to minorities. It is interesting that these results are indicative of greater harassment, perhaps because unemployment provokes greater hostility in the expression of negative attitudes or because it puts a pool of unemployed individuals into contact with others in circumstances where hostile outcomes can easily occur.

[7]The effect flattens out at an ethnic minority concentration of about 41 percent, which is above the sample median (mean) of 33 (31) percent.

Table 6.7: White Attitudes: Ordered Probit

Variable	Coeff	t
Ethn.Conc	3.158	3.41
Ethn.Conc^2	-7.518	-3.75
ΔEthn.Conc	.	.
Unemp.Rate	-0.032	-0.07
Male	0.144	3.12
Age	0.082	1.20
Age^2	-0.008	-1.18
Degree	-0.232	-2.25
Vocational	0.014	0.21
A Level	-0.012	0.20
Const	-0.786	4.20
$d_2 - d_1$	0.208	.
$d_3 - d_2$	0.321	.
$d_4 - d_3$	0.447	.
Mean log-likelihood	-1.051	
Number of cases	2750	

Note: d_i are threshold parameters.

6.6.2 Attitudes of white respondents

We now turn to analysing the effect of ethnic concentration on the attitudes of white respondents. Table 6.7 reports the result of an ordered probit regression of white prejudice on ethnic density and other variables. There emerges very clear evidence of a positive but concave association with ethnic density as well as correlation particularly with education. The peak is at 21 percent of ethnic minority concentration, which is about the 95th percentile of the distribution of ethnic minority concentrations for the white population. These results are similar to those in Dustmann and Preston [2001] where responses from several years of the British Social Attitudes Survey are analysed.

Interesting is the stark contrast with the harassment estimates in the previous section. While harassment seems to decrease with increased ethnic minority concentrations, the opposite is the case for the formation of hostile attitudes. This suggests strongly that racial abuse is not just an intensification of hostile prejudice towards minorities. In fact the two measures relate to ethnic concentrations in the opposite way, indicating that, although whites in areas with higher ethnic concentration tend to be more prejudiced towards minorities, incidents of racial harassment occur less frequently. While the formation of attitudes seems therefore compatible with theories that predict a positive correlation between concentration of minorities and attitudes, the relationship between racial abuse and concentration points rather towards explanations that emphasise the opposite. The findings on attitudes exclude theories like the contact hypothesis, as this

should lead to similar findings for the harassment equation. It is not unlikely that attitudes are formed consistent with group threat theories, while harassment is reduced with concentration due to decreasing contact, as well as safety in numbers serving as a deterrent of aggression, as predicted by the power differential hypothesis.

6.6.3 Precautions

In table 6.8 we present results from the precaution equation. Some coefficient estimates are very similar to those of the harassment equation. This is implied by our model, where precautionary behaviour is conditioned on the (expected) rate of arrival of harassment incidents. As harassment, precautions seem to decrease with ethnic concentration. This result, in conjunction with the finding on attitude formation in the previous section, supports further power differential theory, suggesting that minority individuals feel safer when they are in larger numbers. There is however an initially positive effect of white attitudes on precautionary behaviour, suggesting that increased prejudice leads in fact to more precautions. As before, the effect of changes in ethnic composition on precautions is positive, but not significant.

Interesting are the effects of various measures of individual characteristics on precautions. Males seem to take less precautions than females, explaining possibly in part why males are harassed more. Similar considerations hold for the number of children. Similarly, Caribbeans seem to take less precautions than the Bangladeshi reference group, and experience more harassment. Immigrants and natives, on the other hand, do not seem to differ in the amount of precautionary behaviour, but the incidence of harassment is higher for immigrants.

6.6.4 Joint Estimation

In table 6.9 we report results from joint estimations. For lack of credible identifying instruments, we do not estimate the full structural form of the model, where precautionary behaviour is conditional on the rate of harassment incidents.

The estimation results in Table 6.9 allow for correlation between the unobservables in the way we have explained in the model section. Results point very conclusively towards a positive value for ψ showing that harassment and precautionary behaviour are positively associated either because of correlation in unobserved influences or because

Table 6.8: Precaution: Independent Ordered Probit

Variable	Coeff	t	Coeff	t	Coeff	t	Coeff	t
Ethn.Conc	-2.268	-4.05	-2.227	-3.82	-1.967	-1.20	-2.925	-1.29
Ethn.Conc^2	.	.	.	.	-0.479	-0.14	0.186	0.04
ΔEthn.Conc	.	.	.	.	.	.	1.806	1.13
White Att.	.	.	-0.148	-0.65	1.565	0.65	6.207	1.72
White Att.^{2}	.	.	.	.	-0.963	-0.71	-3.375	-1.69
Unemp.Rate	2.371	3.09	2.472	3.04	2.382	2.65	2.893	2.33
Male	-0.115	-1.37	-0.101	-1.18	-0.100	-1.16	-0.161	-1.54
Age	2.859	1.49	2.601	1.29	2.658	1.32	3.304	1.44
Age^2	-3.384	-1.51	-3.059	-1.32	-3.116	-1.34	-3.262	-1.25
Has children	-0.022	-0.18	0.017	0.14	0.019	0.15	0.066	0.46
No. children	0.570	1.92	0.524	1.73	0.510	1.67	0.592	1.65
Degree	0.022	0.18	0.063	0.49	0.061	0.47	0.009	0.06
A level	0.197	1.90	0.234	2.19	0.234	2.18	0.352	2.92
Vocational	-0.128	-1.09	-0.109	-0.90	-0.105	-0.87	-0.124	-0.81
Immigrant	0.111	0.79	0.124	0.86	0.116	0.80	0.053	0.29
Caribbean	-0.540	-3.09	-0.568	-3.15	-0.564	-3.12	-0.642	-2.96
Indian	-0.002	-0.01	-0.007	-0.04	0.003	0.02	0.110	0.61
Afro-asian	0.216	1.30	0.218	1.28	0.227	1.33	0.288	1.44
Pakistani	0.046	0.31	-0.007	-0.05	-0.019	-0.12	0.010	0.06
Chinese	-0.307	-1.20	-0.248	-0.93	-0.248	-0.91	-0.404	-1.24
Const	-2.147	-5.53	-2.035	-4.83	-2.752	-2.61	-5.135	-3.22
$d_2 - d_1$	0.210	7.67	0.196	7.21	0.196	7.20	0.203	0.89
$d_3 - d_2$	0.281	7.85	0.286	7.73	0.287	7.68	0.316	13.10
$d_4 - d_3$	0.495	7.26	0.480	7.06	0.480	6.99	0.475	0.60
Mean log-likelihood	-0.415		-0.416		-0.416		-0.437	
Number of cases	2231		2105		2105		1545	

Reference individual is of Bangladeshi origin. d_i are threshold parameters.

precaution responds to the prevalence of harassment. The main influences on harassment remain reasonably well identified when estimated jointly with the precaution equation.

Table 6.10 reports the outcome of imposing the cross equation restrictions which require the influence of ethnic density and white attitudes on precaution to come only through their influence on harassment intensity. The acceptability of the restriction differs between specifications and little is gained in the precision of the estimated harassment effects by imposing the restriction.

6.7 Conclusion

In this paper, we analyse the association between ethnic minority concentration, hostile attitudes towards minorities, and the probability of ethnic minorities experiencing racial hostility. Our main focus is on the relationship between ethnic concentration of minorities on the one hand, and hostile attitudes as well as acts of racial harassment on the other. Our approach recognises the role precautionary behaviour may play in distorting this link. Other than much of the existing literature, we understand the formation of hostile attitudes and the realisation of acts of racially motivated violence as two distinct processes.

We develop a general empirical model that subsumes many of the existing theories. We estimate a reduced version of that model, which allows us to derive conclusions about the relationship between racial harassment and precautionary behaviour of minorities, and ethnic concentration and hostile attitudes.

Our findings are interesting in several respects. First, we find strong evidence that racial harassment is not simply a stronger form of racial prejudice. Our results show that, although racial prejudice increases with ethnic concentration over most of the concentration distribution, acts of racial harassment as well as induced precautionary behaviour decrease. Accordingly, these measures can not be interchangeably used to test different theories against each other, as often suggested in the literature. Our results exclude a number of theories, like the contact hypothesis. Results are compatible with attitudes being formed according to considerations as suggested by group threat theory. Acts of harassment however follow a different process, where, in addition, the frequency of contact, as well as deterrence by power through numbers on the side of the minority population play an important role, as suggested by the power differential hypothesis.

Table 6.9: Harassment and Precaution: Joint Model

	Harassment		Precaution		Harassment		Precaution		Harassment		Precaution		Harassment		Precaution	
Variable	Coeff	t	Coeff	t	Coeff	t	Coeff	t	Coeff	t	Coeff	t	Coeff	t	Coeff	t
Ethn.Conc	-5.553	-6.10	-1.431	-3.48	-4.915	-5.37	-1.569	-3.62	-6.790	-2.77	-1.543	-1.27	-7.848	-2.32	-2.975	-1.82
Ethn.Conc^2	.	.	.	.	.	.	.	.	5.368	0.97	0.477	0.19	5.128	0.67	2.134	0.62
ΔEthn.Conc	.	.	.	.	.	.	.	.	.	.	.	.	-1.413	-0.53	1.324	1.08
White Att.	.	.	.	.	-1.252	-3.33	0.008	0.04	4.926	1.45	5.338	2.87	7.678	1.81	6.046	2.42
White Att.^{2}	.	.	.	.	.	.	.	.	-3.465	-1.81	-2.988	-2.88	-4.526	-1.87	-3.126	-2.25
Unemp.Rate	2.972	2.45	1.498	2.70	2.759	2.15	1.802	2.99	2.624	1.94	1.600	2.44	3.251	1.72	2.524	2.88
Male	0.595	4.56	-0.110	-1.79	0.585	4.31	-0.107	-1.66	0.584	4.27	-0.106	-1.64	0.474	2.81	-0.131	-1.72
Age	2.803	0.92	2.208	1.58	2.865	0.95	1.867	1.28	4.337	1.42	2.285	1.57	3.370	0.94	0.974	0.58
Age^2	-4.842	-1.38	-2.511	-1.54	-4.975	-1.43	-2.094	-1.24	-6.727	-1.92	-2.566	-1.51	-4.743	-1.14	-0.850	-0.44
Has children	0.049	0.25	0.016	0.18	-0.020	-0.10	0.035	0.38	-0.010	-0.05	0.036	0.39	-0.148	-0.61	0.065	0.60
No. children	0.519	0.92	0.652	2.81	0.589	1.02	0.673	2.80	0.486	0.84	0.612	2.51	1.516	2.29	0.798	2.90
Degree	0.657	3.21	0.005	0.05	0.577	2.63	0.024	0.25	0.515	2.32	0.012	0.12	0.629	2.37	-0.023	-0.20
A level	0.578	3.37	0.111	1.42	0.565	3.23	0.134	1.63	0.584	3.33	0.141	1.72	0.639	2.99	0.248	2.65
Vocational	0.362	1.89	-0.086	-1.01	0.340	1.77	-0.071	-0.80	0.393	2.04	-0.049	-0.56	0.436	1.83	-0.113	-1.03
Immigrant	-0.597	-2.84	-0.057	-0.58	-0.557	-2.58	-0.042	-0.41	-0.557	-2.54	-0.054	-0.53	-0.731	-2.87	-0.014	-0.12
Caribbean	0.730	2.54	-0.334	-2.58	0.845	2.81	-0.341	-2.49	0.865	2.86	-0.342	-2.50	1.195	3.15	-0.484	-3.03
Indian	0.542	1.96	0.074	0.61	0.639	2.17	0.098	0.76	0.690	2.34	0.129	1.00	1.310	3.37	0.114	0.78
Afro-asian	1.313	4.54	0.235	1.83	1.357	4.41	0.274	2.00	1.391	4.51	0.300	2.20	2.092	5.36	0.281	1.78
Pakistani	0.959	3.72	0.244	2.13	1.061	3.91	0.226	1.84	1.037	3.79	0.190	1.55	1.780	5.04	0.113	0.81
Chinese	1.088	2.96	-0.094	-0.57	1.307	3.41	-0.022	-0.13	1.255	3.27	-0.045	-0.25	1.984	4.32	-0.224	-1.06
London	0.725	4.86	0.342	5.01	0.879	5.33	0.369	4.93	0.746	3.93	0.220	2.34	0.752	2.94	0.194	1.68
Const	-4.888	-7.60	-1.948	-6.78	-3.872	-5.46	-1.999	-6.15	-6.596	-4.37	-4.253	-5.22	-8.739	-4.72	-4.573	-4.05
ψ	2.427	20.76	.	.	2.358	19.59	.	.	2.373	19.94	.	.	2.519	18.54	.	.
ζ	1.384	22.14	.	.	1.408	20.87	.	.	1.394	21.05	.	.	1.374	17.51	.	.
$d_2 - d_1$	.	.	0.161	7.84	.	.	0.152	7.33	.	.	0.151	7.32	.	.	0.153	6.51
$d_3 - d_2$	.	.	0.208	8.04	.	.	0.214	7.86	.	.	0.215	7.76	.	.	0.238	6.97
$d_4 - d_3$	.	.	0.360	7.57	.	.	0.356	7.33	.	.	0.355	7.25	.	.	0.362	6.37
Mean log-likelihood	-0.671				-0.673				-0.672				-0.707			
Number of cases	4935				4640				4640				3435			

Note:d_i are threshold parameters.

Table 6.10: Harassment: Restricted Estimates

	Harassment		Harassment		Harassment	
Variable	Coeff	t	Coeff	t	Coeff	t
Ethn.Conc	-4.893	-5.34	-6.386	-2.65	-7.474	-2.21
Ethn.Conc^2	.	.	4.642	0.85	3.582	0.47
ΔEthn.Conc	.	.	.	.	-1.121	-0.42
White Att.	-1.127	-3.08	4.686	1.38	8.109	1.92
White Att.^{2}	.	.	-3.410	-1.78	-4.750	-1.97
Cross equation restrictions	$\chi^2_1 = 1.988$		$\chi^2_3 = 11.997$		$\chi^2_4 = 11.056$	
P values	$p = 0.159$		$p = 0.007$		$p = 0.026$	

The segregation of mechanisms leading to hostile attitudes and harassment is further suggested by findings on other regressors. For example, while unemployment in the local community does lead to higher frequencies of harassment, it is not related to attitude formation.

6.8 Appendix: Average Attitudes

We base our measure of racial attitudes on the response deriving from the following questions:

Would you describe yourself as very prejudiced against...

Chinese

Asian

Caribbean

Muslim

...people, a little prejudiced, or not prejudiced at all?

We define a discrete variable, taking values from 0 to 4, to describe a general degree of prejudice. This variable is set equal to zero if the respondent declared to have no prejudice against any of the above groups, equal to 1 if the prejudice is against one group, and so on. About 68 percent of the white sample reports to be not prejudiced against any ethnic group, and 7 percent to be prejudiced against all four groups. The white sub-sample from which this variable is derived has a size of 2780 observations.

The average of this variable at county level is then used as a measure of the attitudes characterising the individuals living in the same county as the ethnic minority

respondents.

Conclusion

It is in the interest of the host country to understand the process of socio-economic integration of immigrants and its determinants. In this thesis, the welfare of immigrants and their process of adaptation into the British society and the UK labour market have been investigated from several perspectives, to provide a comprehensive picture of immigration in the UK. For this purpose, in the empirical analysis, we have combined numerous data sources (some of which oversample immigrants). Most of the analysis has focused on ethnic minority immigrants. This chapter summarises main results, discusses the limitations in the available data sources, and suggests avenues for future research.

We have seen that, from after the Second World War, the UK has received a large number of immigrants, many of whom have come from British ex-colonies. Most immigrants from New Commonwealth countries and Pakistan entered the country between the 1950s and the early 1980s. This flow of immigration was influenced by the high labour shortage which characterised the post-war period. Public and private employers actively recruited their employees abroad. Furthermore, for much of this period, immigrants from Commonwealth countries had free access to the UK.

We have seen that ethnic minority immigrants differ from white natives in their education and demographic background more than other immigrants (who mainly come from other European countries). These differences probably were (and still are) at the heart of inter-ethnic tensions. Such tensions, together with the fall in labour demand brought about by the early 1970s recession induced the British government to impose increasingly severe limitations to immigration. From 1981, following these restrictions, immigration from New Commonwealth countries became almost exclusively due to family reunification.

An important finding of this thesis is that ethnic minority immigrants are a very diverse group. Therefore, it is very important not to treat them as a homogeneous group. Minorities of different origin differ in education, age structure, time of residence in the

UK, and geographic distribution. Moreover, these observable differences explain only a part of the differences in economic performance, language ability, and incidence of racial harassment. Related work (see Dustmann et al. [2002]) shows that, with respect to white natives, the differentials in participation, employment, and wages are significantly smaller for Caribbeans, Indians and African Asians than for Black Africans, Bangladeshis, and Pakistanis. These differentials are only partly explained by observable characteristics, such as age, education, family status, and region of residence.

Diverging economic success may also be explained by differences in language performance. We have shown that language is an important determinant of economic performance. However, as with other observable characteristics, differences in economic outcomes among immigrant groups persist even when controlling for language ability. Our results indicate that language proficiency is lowest among those groups that exhibit the largest disadvantages in the labour market.

It is not simple to answer why there are large differences between immigrants of different origins, even conditional on observed characteristics. With regard to labour market outcomes, one possible reason may relate to discrimination. We have not sought to investigate discrimination, and we do not provide any hard evidence for this hypothesis. However, evidence of large employment differentials may be indicative of demand factors playing a role. This may be an important subject for future research. One popular hypothesis is that immigrants choose to become self-employed because they are discriminated against in the labour market. Previous literature has shown that some ethnic minority groups are more likely to be self-employed than others (Clark and Drinkwater [2000] and Dustmann et al. [2002]). Consequently, it would be interesting to analyse self-employment rate differentials (and their determinants) across ethnic groups and to quantify the potential contribution made to the UK economy by self-employed immigrants.

In the case of racial harassment, we have found that Bangladeshis and Indians are harassed less than other ethnic minority groups. Again, it is difficult to understand why some groups are more targeted than others. A tentative explanation may relate to differences in external appearance. Immigrants' looks reflect their differences in origin, religious affiliation, and traditional customs. Negative reactions in hostile individuals may set in against those individual whose appearance is perceived as least conforming to that of the majority population. However, the effect of these differences is difficult to

identify.

Another important result is that significant differences exist between male and female immigrants. Possibly the most important aspect that characterises ethnic minority females, in particular Pakistani and Bangladeshi women, is the low rate of participation in the labour market. Overall, ethnic minority immigrant women are on average less likely to be proficient in English and less likely to experience harassment. These results are likely to be related to each other. Low participation in the labour market implies that women have less opportunity to interact with natives. This is consistent with findings indicating that ethnic concentration at arrival affects the language proficiency of women, but not that of men. Cultural and religious characteristics may explain why these differences between genders are larger than for other immigrants and white natives. However, it is very difficult to establish the causal links with the available data.

One further relevant result relates to the effect of time of residence in the UK. We have found that longer duration of stay implies higher language proficiency, and better economic outcomes. We have also found that, if we cannot reject the hypothesis that immigrants are positively selected in terms of health, this advantage seems to decrease with time in the UK. Most of our data sources are cross-sectional, and do not allow to separately identify cohort and assimilation effects. The problem could only be addressed in the analysis of the labour market performance differentials, using several waves from the Labour Force Survey. Our results show that employment and participation differentials converge with years of stay in the UK. However, due to small sample size, the analysis has been conducted jointly for all ethnic minority groups, and our results for wages are not robust.

Another important contribution of this thesis is a detailed analysis of the geographic distribution of immigrants and the relevance of this for several variables related to immigrants' welfare and immigration effects. We have analysed the relationship between ethnic concentration and language proficiency. We have found no significant evidence of immigrants' location choice being based on their level of language fluency. As we mention above, we have also found that ethnic concentration has a negative effect on female immigrants' accumulation of language capital. We have investigated the association between ethnic concentration and racial harassment. Evidence shows that higher ethnic density implies lower incidence of racial harassment, possibly because of higher risk of reprisal for potential perpetrators. However, our findings also indicate that higher ethnic

concentration implies higher hostile attitudes. These two results combined suggest that (differently from what previous theories implicitly assume) it is of considerable importance to consider overt acts of hostility and racial prejudices as independent processes. Finally, we have estimated the impact of immigration on the local labour market. Individual preferences over migration policies depend, among other things, on the expected economic impact of immigration. Our results indicate that there is no strong evidence of adverse effects of immigration on native employment or wages. These findings are consistent with previous research for the US and other European countries[8].

Throughout the thesis, we have drawn attention to many weaknesses in the available data and conceptual problems in the empirical analysis. One problem, repeatedly observed and affecting a large part of the analysis, relates to the small sample size of the available data sources. Immigrants are a small fraction of the total population, therefore surveys which randomly sample the population have few observations on immigrants. The problem becomes particularly serious when we need information on wages and want to compare groups within the immigrant population. To make the problem worse, in the survey used, some questions were asked only of a sub-sample of respondents, reducing the number of observations even further.

Small sample size also hampers the investigation of economic performance across immigrant generations. We have observed large differences in the economic outcomes of ethnic minority natives. For policies aimed at the integration of native minority communities it is important to fully understand the determinants of these differentials. Also, due to limited sample size we have not been able to investigate health differentials for different ethnic minority groups.

One other problem is the quality of the information on local demographic composition. In Chapters 3, 5, and 6 we have seen that access to this information is limited. Information provided by the Population Census is detailed at geographical level, but is available only every ten years and in selected published cross-tabulations. On the other hand, information provided by the Labour Force Survey is yearly and allows subdivision of the population at smaller levels, but is only available at regional level. Access to more detailed information would, for example, enable us to analyse the impact of migration for London. Information on the internal mobility of natives and immigrants would also

[8]See, for example, Borjas [1999] for the US, Pischke and Velling [1997] for Germany, and Hunt [1992] for France.

be most useful in much of our analysis.

The creation of a longitudinal survey sampling immigrants would answer some of the questions raised in this thesis. This survey should address issues specific to migration status such as those already contained in the FNSEM, but provide more information on wages and economic activities and possibly address a larger portion of immigrants. For example, little evidence exists on the performance and adaptation of white immigrants in the UK.

Bibliography

Alexander, Z. (1999). Study of black, Asian and ethnic minority issues. The Department of Health.

Altonji, J. and Card, D. (1991). *Immigration, Trade and Labor*, chapter The Effects of Immigration on the Labor Market Outcomes of Less-skilled Natives. University of Chicago Press.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, 58:277–297.

Arellano, M. and Bond, S. (1998). Dynamic panel data estimation using DPD98 for Gauss: A guide for users. mimeo.

Bartel, A. (1989). Where do the new U.S. immigrants live? *Journal of Labor Economics*, 7:371–391.

Bell, B. (1997). The performance of immigrants in the United Kingdom: Evidence from the GHS. *The Economic Journal*, 107:333–344.

Benzeval, M., Judge, K., and Solomon, M. (1992). *The Health Status of Londoners.* King's Fund Institute.

Berman, E., Lang, K., and Sriniver, E. (2000). Language-skill complementarity: Returns to immigrant language acquisition. mimeo, Boston University.

Blackaby, D., Clark, K., Leslie, D., and Murphy, P. (1994). Black-white male earnings and employment prospects in the 1970s and 1980s evidence for Britain. *Economics Letters*, 46:273–279.

Blackaby, D., Drinkwater, S., Leslie, D., and Murphy, P. (1997). A picture of male and female unemployment among Britain's ethnic minorities. *Scottish Journal of Political Economy*, 44:182–197.

Blalock, H. (1956). Economic discrimination and negro increase. *American Sociological Review*, 21:548–588.

Blalock, H. (1957). Percent non-white and discrimination in the South. *American Sociological Review*, 22:677–682.

Blalock, H. (1967). *Toward a Theory of Minority Group Relations.* New York: Wiley.

Blau, P. (1977). *Inequality and Heterogeneity: A Primitive Theory of Societal Structure.* New York: Free Press.

Blumer, H. (1958). Race prejudice as a sense of group position. *Pacific Sociological Review*, 1:3–7.

Bobo, L. (1983). Whites' opposition to busing: Symbolic racism or realistic group conflict? *Journal of Personality and Social Psychology*, 45:1196–1210.

Borjas, G. (1985). Assimilation, changes in cohort quality, and the earnings of immigrants. *Journal of Labor Economics*, 3:463–489.

Borjas, G. (1987). Self-selection and the earnings of immigrants. *American Economic Review*, 77:531–553.

Borjas, G. (1999). *Handbook of Labor Economics*, volume 3, chapter The Economic Analysis of Immigration, pages 1697–1760. Elsevier.

Borjas, G. and Bronars, S. (1991). Immigration and the family. *Journal of Labor Economics*, 9:123–148.

Borjas, G. J. (1994). The economics of immigration. *Journal of Economic Literature*, 32:1667–1717.

Bound, J. (1991). Self-reported versus objective measures of health in retirement models. *Journal of Human Resources*, 26:106–138.

Bound, J. and Burkhauser, R. (1999). *Handbook of Labor Economics*, chapter Economic Analysis of Transfer Programs Targeted on People with Disabilities, pages 3417–3528. Elsevier.

Cameron, C. and Trivedi, P. (1998). *Regression Analysis of Count Data.* Cambridge University Press.

Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics*, 19:22–63.

Carrington, W. and de Lima, P. (1996). The impact of 1970s repatriates from Africa on the Portuguese labor market. *Industrial and Labor Relations Review*, 49:330–347.

Chen, J., Wilkins, R., and Ng, E. (1996). Health expectancy by immigrant status, 1986 and 1991. *Health Reports*, 8:29–38.

Chiswick, B. (1980). The earnings of white and coloured male immigrants in Britain. *Economica*, 47:81–87.

Chiswick, B. (1991). Speaking, reading, and earnings among low-skilled immigrants. *Journal of Labor Economics*, 9:149–170.

Chiswick, B. (1993). Review of immigration and and the work force: Economic consequences for the United States and source areas. *Journal of Economic Literature*, 31:910–911.

Chiswick, B. (1994). Language choice among immigrants in a multi-lingual destination. *Journal of Population Economics*, 7:119–131.

Chiswick, B., Cohen, Y., and Zach, T. (1997). The labor market status of immigrants: Effects of the unemployment rate at arrival and duration of residence. *Industrial and Labor Relations Review*, 50:289–303.

Chiswick, B. and Miller, P. (1995). The endogeneity between language and earnings: International analyses. *Journal of Labor Economics*, 13:246–288.

Chiswick, B. and Miller, P. (1996). Ethnic networks and language proficiency among immigrants. *Journal of Population Economics*, 9:19–35.

Chiswick, B. and Miller, P. (2002). Do enclaves matter in immigrant adjustment? IZA Discussion Paper No. 449.

Chiswick, B. R. (1978). The effect of americanization on the earnings of foreign-born men. *The Journal of Political Economy*, 86:897–921.

Clark, K. and Drinkwater, S. (2000). Pushed out or pulled in? self-employment among ethnic minorities in England and Wales. *Labour Economics*, 7:603–628.

Clark, W. (1993). Neighborhood tipping in multiethnic/racial context. *Journal of Urban Affairs*, 15:161–172.

Currie, J. and Madrian, B. (1999). *Handbook of Labor Economics*, chapter Health, Health Insurance and the Labor Market, pages 3417–3528. Elsevier.

Dearden, L. (1999). Qualifications and earnings in Britain: How reliable are conventional OLS estimates of the returns to education? IFS working paper No. W99/7.

Dearden, L., Ferri, J., and Meghir, C. (2002). The effect of school quality on educational attainment and wages. *Review of Economics and Statistics*, 84:1–20.

Deaton, A. (1999). Inequalities in income and inequalities in health. NBER Working Paper no. 7141.

De New, J. and Zimmermann, K. (1994). Native wage impacts of foreign labor: A random effects panel analysis. *Journal of Population Economics*, 7:177–192.

De Sena, J. (1990). *Protecting One's Turf: Social Strategies for Maintaining Urban Neighborhoods.* Univ. Press Am.

Dobson, J., Koser, K., McLaughlan, G., and Salt, J. (2001). International migration and the United Kingdom: Recent patterns and trends. Home Office RDS Occasional Paper no. 75.

Duffield, M. (1985). *Rationalization and the Politics of Segregation: Indian Workers in Britain's Foundry Industry, 1945-62.*

Duncan, O. and Duncan, B. (1955). A methodological analysis of segregation indexes. *American Sociological Review*, 20:210217.

Dustmann, C. (1994). Speaking fluency, writing fluency and earnings of immigrants. *Journal of Population Economics*, 7:133–156.

Dustmann, C. (1999). Temporary migration, human capital, and language fluency of migrants. *Scandinavian Journal of Economics*, 101:297–314.

Dustmann, C., Fabbri, F., and Preston, I. (2003). The local labour market effects of immigration in the UK. UCL mimeo.

Dustmann, C., Fabbri, F., Preston, I., and Wadsworth, J. (2002). The performance of immigrants in the UK. Home Office Report.

Dustmann, C. and Preston, I. (2001). Attitudes to ethnic minorities, ethnic context and location decisions. *The Economic Journal*, 111:353–373.

Dustmann, C. and van Soest, A. (2001). Language fluency and earnings estimation with misclassified language indicators. *Review of Economics and Statistics*, 83:663–74.

Gang, I., Rivera-Batiz, F., and Yun, M. (2002). Economic strain, ethnic concentration and attitudes towards foreigners in the European Union. IZA Discussion Paper no. 578.

Gaston, N. and Nelson, D. (2000). Immigration and labour-market outcomes in the United States: A political-economy puzzle. *Oxford Review of Economic Policy*, 16:104–114.

Gaston, N. and Nelson, D. (2001). The employment and wage effects of immigration: Trade and labour economics perspectives. University of Nottingham, Leverhulme Centre for Research on Globalisation and Economic Policy Research Paper 2001/28.

Goldman, D. and Lakdawalla, D. (2001). Understanding health disparities across education groups. NBER Working Paper no. 8328.

Green, D., Strolovich, D., and Wong, J. (1998). Defended neighborhoods, integration, and racially motivated crimes. *American Journal of Sociology*, 104:372–403.

Grossman, M. (1972). The demand for health - a theoretical and empirical investigation. NBER Working Paper.

Haisken De New, J. and Zimmermann, K. (1999). *International Trade and Employment: The European Experience*, chapter Wage and Mobility Effects of Trade and Migration, pages 139–160. Oxford University Press.

Hanson, G. and Slaughter, M. (1999). The rybczinski theorem, factor-price equalization, and immigration: Evidence from u.s. states. NBER Working Paper No. 7074.

Hanson, G. and Slaughter, M. (2002). Labor market adjustment in open economies: Evidence from U.S. states. *Journal of International Economics*, 57:3–29.

Hargreaves, S. (2000). System to detect tuberculosis in new arrivals to UK must be improved. *British Medical Journal*, 320.

Hatton, T. and Wheatley Price, S. (2002). *European Migration, What do we Know?*, chapter Migration, Migrants and Policy in the United Kingdom. CEPR/OUP. ed. K.F. Zimmermann.

Hawker, J., Bakhshi, S., Ali, S., and Farrington, C. (1999). Ecological analysis of ethnic differences in relation between tuberculosis and poverty. *British Medical Journal*, 319:1031–1034.

Heath, T. and Hill, R. (2002). Asylum statistics, United Kingdom 2001. Home Office HOSB 9/02.

Heckman, J. (1997). Instrumental variables. *Journal of Human Resources*, 32:441–462.

Heckman, J., Ichimura, H., Smith, J., and Todd, P. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies*, 64:605–654.

Heckman, J., Ichimura, H., and Todd, P. (1998). Matching as an econometric evaluation estimator. *Review of Economic Studies*, 65:261–294.

Hunt, J. (1992). The impact of the 1962 repatriates from Algeria on the French labor market. *Industrial and Labor Relations Review*, 45:556–72.

James, A. (2000). The return of an old enemy. The Guardian.

Jasso, G., Massey, D., Rosenzweig, M., and Smith, J. (2002). Immigrant health - selectivity and acculturation. paper presented at the National Academy of Science Conference on Racial and Ethnic Disparities in Health.

Karlsen, S. and Nazroo, J. (2002). The relationship between racial discrimination, social class and health among ethnic minority groups. *American Journal of Public Health*, 92:624–631.

Katz, L. and Murphy, K. (1992). Changes in relative wages, 1963-1987: Supply and demand factors. *Quarterly Journal of Economics*, 107:35–78.

Kennedy, B., Kawachi, I., Glass, R., and Prothrow-Stith, D. (1998). Income distribution, socio-economic status, and self-rated health in the United States: multilevel analysis. *British Medical Journal*, 317:917–921.

Kinder, D. and Mendelberg, T. (1995). Cracks in american apartheid: The political impact of prejudice among desegregated whites. *Journal of Politics*, 57:402–424.

Krueger, A. and Pischke, J. (1997). A statistical analysis of crime against foreigners in unified germany. *Journal of Human Resources*, 32:182–209.

Kuhn, P. and Wooton, I. (1991). *Immigration, Trade and Labor*, chapter Immigration, International Trade and the Wages of Native Workers. University of Chicago Press.

Lalonde, R. and Topel, R. (1991). *Immigration, Trade and Labor*, chapter Labor Market Adjustments to Increased Immigration. University of Chicago Press.

Lavender, M. (1997). Screening immigrants for tuberculosis in Newcastle upon Tyne. *Journal of Public Health Med*, 19:320–323.

Lazear, E. P. (1999). Culture and language. *The Journal of Political Economy*, 107(6):S95–S126.

Leamer, E. and Levinsohn, J. (1999). *Handbook of International Economics*, volume 3, chapter International Trade Theory: The Evidence, pages 1339–1394. North Holland Press.

Levin, J. and McVine, J. (1993). *Hate Crime: The Rising Tide of Bigotry and Bloodshed.* New York: Plenum.

LeVine, R. and Campbell, D. (1972). *Ethnocentrism.* New York: Plenum Press.

Marmot, M., Adelstein, A., and Bulusu, L. (1984). Lessons from the study of immigrant mortality. *Lancet*, pages 1455–1457.

McCallum, J., Shadbolt, B., and Wang, D. (1994). Self-rated health and survival: 7 years follow-up study of Australian elderly. *American Journal of Public Health*, 847:1100–1105.

McManus, W. S. (1990). Labor market effects of language enclaves. *The Journal of Human Resources*, 25(2):228–252.

Mincer, J. (1978). Family migration decisions. *The Journal of Political Economy*, 86:749–773.

Modood, T. and Berthoud, R. (1997). *Ethnic Minorities in Britain.* Policy Studies Institute.

Murphy, K. and Welch, F. (1992). The structure of wages. *Quarterly Journal of Economics*, 107:285–326.

Murray, R. (2000). UK life blamed for ethnic schizophrenia. BBC News. http://news.bbc.co.uk/2/hi/health/807945.stm.

Nickell, S. (1980). A picture of male unemployment in Britain. *Economic Journal*, 90:776–794.

Pischke, J. and Velling, J. (1997). Employment effects of immigration to Germany: An analysis based on local labor markets. *The Review of Economics and Statistics*, 79:594–604.

Propper, C. (2000). The demand for private health care in the UK. *Journal of Health Economics*, 19:855–876.

Raleigh, V. (1997). Diabetes and hypertension in Britain's ethnic minorities: implications for the future of renal services. *British Medical Journal*, 314.

Ratcliffe, P. (1981). *Racism and Reaction.* London, Routledge & Kegan Paul.

Razum, O., Zeeb, H., and Rohrmann, S. (2000). The 'healthy migrant effect' - not merely a fallacy of inaccurate denominator figures. *Journal of International Epidemiology*, 29:191–192.

Rieder, J. (1985). *Canarsie: the Jews and Italians of Brooklyn Against Liberalism.* Harvard University Press.

Rivera-Batiz, F. (1990). English language proficiency and the economic progress of immigrants. *Economics Letters*, 34:295–300.

Rosenbaum, P. and Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55.

Rothbart, M. and John, O. (1993). *Immigration and Ethnicity: Integration of America's Newest Immigrants*, chapter Immigration and race: recent trends. Urban Inst.P.,U.S.

Rumbaut, R. and Weeks, J. (1996). Unraveling a public health enigma: Why do immigrants experience superior prenatal health outcomes? *Research in the Sociology of Health Care*, 13B:337–391.

Samuelson, P. (1948). International trade and the equalization of factor prices. *Economic Journal*, 48:163–184.

Scheve, K. and Slaughter, M. (2001). Labor market competition and individual preferences over migration policy. *Review of Economics and Statistics*, 83:133–45.

Shields, M. and Wheatley Price, S. (1998). The earnings of male immigrants in england: Evidence from the quarterly LFS. *Applied Economics*, 30:1157–1168.

Shields, M. and Wheatley Price, S. (2001). Exploring the economic and social determinants of psychological and psychosocial health. IZA Working Paper No. 396.

Shields, M. and Wheatley Price, S. (2002). The English language fluency and occupational success of ethnic minority immigrant men living in English metropolitan areas. *Journal of Population Economics*, 15:137–160.

Smith, A. (2001). A healthy colour. The Guardian.

Smith, J. (1999). Healthy bodies and thick wallets: The dual relation between health and economic status. *Journal of Economic Perspectives*, 13:145–166.

Smith, P. and Prior, G. (1996). The Fourth National Survey of Ethnic Minorities: Technical report. Social and Community Planning Research.

Smith, S. (1989). *The Politics of Race and Residence: Citizenship, Segregation and White Supremacy in Britain.* Cambridge: Polity Press.

Stewart, M. (1983). On least squares estimation when the dependent variable is grouped. *Review of Economic Studies*, 50:737–753.

Studlar, T. (1978). Policy voting in Britain: The colored immigration issue in the 1964, 1966 and 1970 elections. *The American Political Science Review*, 72:46–64.

Wheatley Price, S. (2001). The employment adjustment of male immigrants to the English labour market. *Journal of Population Economics*, 14:193–200.

Winter-Ebmer, R. and Zweimüller, J. (1996). Immigration and the earnings of young native workers. *Oxford Economics Papers*, 48:473–491.

Winter-Ebmer, R. and Zweimüller, J. (1999). Do immigrants displace young native workers: The Austrian experience. *Journal of Population Economics*, 12:327–340.

Woodland, A. (1982). *International Trade and Resource Allocation.* North-Holland.

Wooldridge, J. (2003). *Introductory Econometrics.* Thomson South-Western.

Zimmermann, K. and Bauer, T. (2002). *The economics of migration.* Elgar Reference Collection.