# On Fragile Grounds: A replication of "Are Muslim immigrants different in terms of cultural integration?" Technical documentation

Mahmood Arai, Jonas Karlsson and Michael Lundholm

# On Fragile Grounds:
# A replication of *Are Muslim immigrants different in terms of cultural integration?*
# Technical documentation

Mahmood Arai,[*] Jonas Karlsson[†] and Michael Lundholm[‡]

December 19, 2008

## Abstract

This is a technical documentation of Arai et al. (2008) which replicates "Are Muslim Immigrants Different in terms of Cultural Integration?" by Alberto Bisin, Eleonora Patacchini, Thierry Verdier and Yves Zenou, published in Journal of European Economic Association, 6, 445-456, 2008.

Bisin et al. (2008) report that they have 5963 observations in their study. Using their empirical setup, we can only identify 1901 relevant observations in the original data. After removing missing values we are left with 818 observations. We cannot replicate any of their results and our estimations yield no support for their claims.

---

[*]Corresponding author. Department of Economics and SULCIS, Stockholm University, SE 106 91 Stockholm, Sweden, mahmood.arai@ne.su.se.

[†]The Institute for Social Research and SULCIS, Stockholm University, jonas.karlsson@sofi.su.se.

[‡]Department of Economics, Stockholm University, michael.lundholm@ne.su.se.

# Contents

# 1  Introduction

This is a replication of the empirical results reported in Bisin et al. (2008). They use British data and analyse how Muslims and non-Muslims differ in cultural integration measured as (i) *Importance of Religion*, (ii) *Attitude Towards Inter Marriage* and (iii) *Importance of Racial Composition in Schools*.[1]

In the abstract of their paper they write:

> "...Muslims integrate less and more slowly than non-Muslims. ...We also find no evidence that segregated neighbourhoods breed intense religious and cultural identities for ethnic minorities, especially for Muslims." (Bisin et al., 2008, p. 245)

We wanted to check the robustness of their results when considering the ethnic and religious heterogeneity within both groups, Muslims and non-Muslims. Among other things, we were concerned about the measures of cultural values used in the paper. These measures capture ethnic and religious attributes in different degrees for different groups. For example the variable *Attitudes towards Inter-Marriage* with the majority UK population captures only inter-ethnic marriage for the Christian ethnic minorities but both inter-ethnic and inter-religious marriage for Muslims.

However, already an initial inspection of data disclosed that the number of observations in Bisin et al. (2008) exceeded the total number of observations in the ethnic minority sample. We communicated this to the authors and they answered that there were some coding errors. We have received revised codes and a revised version of their specifications and tables. Their revised codes yield fewer observations than the sample in the published version, but still more than we can identify in the relevant sample of the original data. As far as we can see, a source of the large number of observations in their revised codes is that dummy variable definitions include observations with missing values in the reference categories (defined as zeros). The underlying codes to the published paper were, however, not made available and the exact nature of the original errors are therefore unknown to us.

Bisin et al. (2008) report that they have 5963 observations in their study, whereas the ethnic minority sample in Berthoud et al. (1997) consists of 5226 observations. Implementing their empirical setup, we can only identify 1901 relevant observations in the original data. After removing missing values we are left with 818 observations. Using the remaining sample and running their specifications, we find no results that support their claims. Our replication therefore stopped here and we did not perform any sensitivity analysis. The great loss of observations implies that the remaining sample

---

[1] To facilitate comparability we use the same labels on the variables as Bisin et al. (2008).

is most likely not representative. Therefore, we hesitate to draw inference from the regressions results.

In this paper we only document the replication and report and comment results using the variable definitions, the variable names and the specifications used in Bisin et al. (2008). We choose a procedure that makes it easy to reproduce our results. Influenced by Koenker and Zeileis (2007), we use an integrated approach where data management, estimations, and the text that rely on these computations are all integrated in one single file. This strategy has the advantage that it makes is easy to adjust the codes and automatically generate a revised version of the paper.

All data analysis is made in R (R Development Core Team, 2008) and all code files related to this project can be found on `http://people.su.se/~lundh/fragile_grounds/`.

In this technical documentation we present our results in greater detail, but also all our working procedures, variables definitions etc. In addition the central part of our codes are included with typeset comments. This is done as an attempt to implement Literate Statistical Programming.

The remainder of the paper is organised as follows. The data and variable definitions are described in Section 2. Regression results are presented in Section 3. The paper is concluded in Section 4. Finally the production procedure is described in Section 5.

## 2  Data and variable description

### 2.1  Data

The data set is the *Fourth National Survey of Ethnic Minorities 1993-1994* (FNSEM); see (Berthoud et al., 1997).[2] It's main objective were:

- "to describe the social and economic conditions of Britain's main ethnic minority groups, including their health, and to compare these with the social and economic conditions of the white majority

- to assess changes over time through comparisons with other work

- to show how the position of ethnic minority groups is related to the social and ethnic compositions of the areas in which they live

- to explore diversity among different ethnic minority groups

---

[2] The data can be accessed from the UK Data Archive (UKDA) via Athens. UK Data Archive is found at `http://www.data-archive.ac.uk/` and Athens at `http://www.athens.ac.uk/`.

- to describe perceptions and experience of racial discrimina-
  tion and social harassment"

Berthoud et al. (1997)

For our coding we have used FNSEM (1993a), which contains the project instructions, and FNSEM (1993b), which is the data description file included in the files obtained when the entire data set is downloaded from the UKDA.

In the following we present how the original data are used to define the data set used in the estimations. We present extracts of our R code (R Development Core Team, 2008) with extensive comments and discussions. For details about our working procedures and how we document the research, see section 5 on page 21. In the code chunks ">" denotes the R prompt and "+" continuation of the previous line.

## 2.2   Reading data and selecting variables

We load package **foreign** to read STATA data format. Data is read from the unpacked Stata-version of the data and "_" in variable names are converted to ".".

```
> library(foreign)
> FNSEM <- data.frame(read.dta("3685.dta",
+       convert.underscore=TRUE))
```

After reading the data we select a subset of variables to be used. This code is in `araietal_source.Rnw` but not shown here. It is also available in `araietal_source.R`.

## 2.3   Defining the relevant (ethnic minority) sample

The data consist of two samples, *Ethnic Minorities* and *Whites*. We are only interested in the former and remove all Whites. The variable `ethnic` indicates ethnic group of the individual according to the British standard and is used for this purpose.

One of the three measures of cultural integration in Bisin et al. (2008) is *Importance of Religion*. Whether a respondent has a religion or belongs to a church is registered in question `s6`. Those who do not have a religion or do not belong to a church are coded 2; we remove these observations from the sample since they cannot be classified in a religious group.

A variable used in Bisin et al. (2008) concerns the role of the respondent and his or her parents about choosing the respondent's husband/wife. Since this information is only available for married and previously married persons, the unmarried persons are removed from the sample.

Furthermore, respondents were faced with one out of three question-naires (green (catageory 1), yellow (category 2) and pink (category 3)). The

questions involved in the study are only answered by individuals who were faced with the green questionnaire. Therefore we keep only these in the sample.

```
> FNSEM <- FNSEM[FNSEM$ethnic!="white" &
+       !is.na(FNSEM$ethnic),]

> FNSEM <- FNSEM[FNSEM$s6!=2,]
> FNSEM <- FNSEM[FNSEM$a1e!=3,]
> U <- FNSEM <- FNSEM[FNSEM$question==1,]
```

An issue where Bisin et al. (2008) is imprecise is whether the questions they address regard Muslims/non-Muslims or Muslim/non-Muslim immigrants. Different sample selections are possible here. The model specifications in Bisin et al. (2008) implies that White Muslims are excluded and native ethnic minority Muslims are included. This sample definition does not match Bisin et al. (2008) writing using the terms *Muslim and non-Muslim immigrants*, as the sample includes natives.

## 2.4   Recoding of missing values

The data set contains several codes for missing values. These missing values can be of different characters: e.g., non–availables, 'can't say' or because the respondent was filtered in a previous filter question. We employ the strategy to code all these as non–availables in R; i.e. code `NA`. In the data set genuine non-availables are generally coded as $-1$. We set all $-1$ to `NA` in the entire data set:

```
> is.na(U) <- U==-1
```

In addition to $-1$ several other codes (`na`, 7, 8, 9, 98, 99, 997 and 999) are occasionally used in the data set to indicate various unknown categories. Some of $-1$ and other unknown categories are non–availables and have to be deleted in estimations. This is done after we have coded all our variables in section 2.7. In questions following a filter question `NA` may have to be set in a category. Some of these other codes used to denote unknowns are genuine `NA`'s and has to be removed. Others will be included in a category. This is done variable by variable below.

## 2.5   Variable definitions

Using the same variable names as in Bisin et al. (2008), we define the variables at the precision described by the authors. We here give our interpretation of the variable definitions in Bisin et al. (2008).

### 2.5.1 Religious affiliation

Question `s6` asks whether the respondent belongs to a church or has a religion. For those who answer yes, question `s7` asks which that church or religion is. We define two religious affiliations: `muslim`, (which are all who answered category 3 (muslim) on question `s7`) and `non-muslim` (all who did not answer category 3 (muslim) on question `s7`). All non–religious, that is those who answered category 2 on question `s6` are already removed from the sample. Observations containing `na` are recoded to `NA`. (FNSEM, 1993a, p. 112f)

```
> is.na(U$s7) <- U$s7=="na"
> U$Religion  <- ifelse(U$s7=="muslim","muslim","non-muslim")
```

### 2.5.2 Importance of religion

Question `s9` is about the importance of religion. To grade the *Importance of Religion* respondents have to choose between the following categories: 1. not at all important, 2. not very important, 3. fairly important, or 4. very important. Following standard coding practice of such questions, 1 and 2 should be one category and 3 and 4 another, but Bisin et al. (2008) choose to put 1, 2 and 3 in the same category.

Among those who have answered the question, very few have chosen the alternative 1 or 2 in their answer, implying very skewed distribution. Following Bisin et al. (2008), code 4 ('Very important') as answer on question `s9` is coded TRUE; else FALSE. Codes 8 ('Can't say') and 9 are coded as `NA`. (FNSEM, 1993a, p. 112f)

```
> is.na(U$s9) <-  U$s9==8 | U$s9==9
> U$Importance.of.Religion <- U$s9 == 4
```

### 2.5.3 Attitude towards inter-marriage

Question `s34a` is "Would you personally mind if a close relative were to marry a white person?". It serves as a filter question to `s34b` ("Would you mind very much or just a little?") which is asked to those who answered yes (cod 1) on `s34a`. We code those who answer yes on both questions (Mind & Mind very much) as TRUE. The category FALSE refers then to those who (Do not mind) or (Mind & Mind Little). Code 8 (Can't say) on `s34a` and code 9 on `s34a` and on `s34b` are assigned as `NA`. (FNSEM, 1993a, p. 125)

```
> is.na(U$s34a) <- U$s34a==8 | U$s34a==9
> is.na(U$s34b) <- U$s34b==8
> U$Attitude.Towards.Inter.Marriage <-
+       U$s34a==1 & U$s34b==1
```

### 2.5.4 Importance of racial composition in schools

Two questions are asked: `s23` is

> "If you were choosing a school for an eleven-year old child of
> yours, would your choice be influenced by how many (RESPON-
> DENT'S ETHNIC ORIGIN) children there were in the school?"
> (FNSEM, 1993a, p. 120)

and `s24a` asks that if the available school were similar in other ways would
you prefer to send this child to school with fewer than half of the pupils
(code 1), about half of the students (code 2), more than half (code 3) were
of your own ethnic origin. `s23` is not a filter question. *Importance of Racial
Composition in Schools* is set to TRUE if `s24a` is equal to 3 and FALSE
otherwise. Code 7 (No preference) is coded as FALSE. Codes 8 (Can't say)
and 9 are assigned as `NA`. (FNSEM, 1993a, p. 120)

```
> is.na(U$s24a) <- U$s24a==8 | U$s24a==9
> U$Importance.of.Racial.Composition.in.Schools <- U$s24a==3
```

### 2.5.5 Born in the UK

Defines who is born in the United Kingdom (question `a3`). Category 16 is
Northern Ireland, category 17 England and Wales and category 18 Scotland.
Code 99 is assigned as `NA`. (FNSEM, 1993a, p. 107).

```
> is.na(U$a3)<- U$a3==99
> U$Born.in.the.UK <- U$a3==16 | U$a3==17 | U$a3==18
```

### 2.5.6 Age at and years since arrival

This part defines the variables *Age at arrival* and *Years since arrival* by using
information about year of migration (question `a4n`), age (question `a1an`) and
the year (variable `year` when the interview is made). The interview is made
in `93` or `94`). The result is that some individuals get the age at arrival $-1$,
which presumably is due to rounding of years since arrival and the age.

All born in the UK are coded as 0 for `Age.at.Arrival` and `Years.Since
Arrival`. Since these two variables are related to the age of the immigrants,
one could also add an interaction variable between `Born.in.the.UK` and `Age`
to account for effect of age for the natives. The effect of age for natives is not
represented in the model as specified by Bisin et al. (2008). We experimented
with this and results were basically unchanged. The interaction variable
is insignificant in all specifications. Code 99 is assigned as `NA` for `year`.
(FNSEM, 1993a, pp. 105, 107).

```
> is.na(U$a4n) <- U$a4n==98 | U$a4n==99
> is.na(U$year)<- U$year==99
> U$Age <- U$a1an
> U$Years.Since.Arrival <- ifelse(U$Born.in.the.UK==TRUE,
+     0, U$year-U$a4n)
> U$Age.at.Arrival <- U$Age - U$Years.Since.Arrival
> U$Age.at.Arrival <- replace(U$Age.at.Arrival,
+     U$Born.in.the.UK==TRUE, 0)
```

### 2.5.7   Female

Definition of females via question `hh2a.s`. Code `na` is coded as `NA`. (FNSEM, 1993a, pp. 318).

```
> is.na(U$hh2a.s) <- U$hh2a.s =="na"
> U$Female <- U$hh2a=="female"
```

### 2.5.8   Arranged marriage

In question `s39` Indian, Pakistani, Bangladeshi and Chinese respondents who has ever been married were asked a question about the decision regarding their marriage. The question ask about the role of the respondent and his or her parents about choosing the respondent's husband/wife. In categories 1 and 2 of `s39` the respondent's parents made the final decision and these categories define the dummy where the respondent is or has been living in an arranged marriage (code 1; all other are coded 0). Notice that singles and Caribbeans have not received this question. Singles are already removed from the sample (see section 2.3. Caribbeans are coded 0. This means that the Caribbeans do not marry according to the decision of their parents. Category 8 (Can't say) and category 9 are assigned `NA`. (FNSEM, 1993a, pp. 127).

```
> is.na(U$s39) <- U$s39==8 | U$s39==9
> U$Arranged.Marriage <- ifelse(
+     U$s39==3 | U$s39==4 | U$s39==5 |
+     U$ethnic=="caribbean", FALSE , TRUE)
```

### 2.5.9   Discrimination

The *discrimination* variable is based on a series of questions related to discrimination; `v1a-v1d` about physical attacks, `v9a` about insults, `j55a` and `j63a` discrimination at work. Basically. anyone answering that they have been discriminated for any of these reasons are coded 1; else code 0.

Questions `v1a-v1d` is a series of filter questions: Question `v1a` asks if the respondent have been attacked (yes or no), question `v1b` how many attacks

9

the respondent has been enduring and question, `v1c` asks those who have been attacked once if they believe the attack had to do with reasons to do with race or colour and `v1d` asks the same question and regards those who have been attacked more than once. Generally code 8 and code 9 are assigned `NA`, except for question `v1b` where also code 7 is assigned `NA`. (FNSEM, 1993a, pp. 154ff, 163, 195 and 199).

```
> vjlist <- c(paste("v1",letters[1:4],sep=""),
+       "v9a","j55a","j63a")

> is.na(U[vjlist]) <- U[vjlist]==8 |
+       U[vjlist]==9
> is.na(U$v1b)  <- U$v1b==7 |  U$v1b==8 |
+       U$v1b==9
> U$Discrimination   <-
+       (U$v1a==1 &  U$v1b==1 &  U$v1c==1) |
+       (U$v1a==1 &
+       (U$v1b >= 2 & U$v1b <= 6)  &
+       U$v1d==1) | U$v9a==1 |  U$j55a==1 |
+       U$j63a==1
```

### 2.5.10 Children

No question about the number of children is asked. Instead the number of children has to be calculated indirectly via the number of children not living at home (questions `f16a` and `f16b1n`-`f16b3n`) and the relation between the respondent and other persons living in the household (questions `hh2c.b`-`hh2.c.m`).

The number of children out of home is calculated in the following way: If children out of home is TRUE (`f16a=1`), then the number of children equals the sum of `f16b1n`, `f16b2n` and `f16b3n` (the number of children not living at home below 5 years, between 5 and 15 years and above 15 years of age). Else, if there are no children out of home (i.e., if `f16a=2`), then the number of children out of home is set to 0. Missing values are coded as below. (FNSEM, 1993a, p. 57).

```
> is.na(U$f16a)    <- U$f16a==8 | U$f16a==9
> is.na(U$f16b1n)  <- U$f16b1n==99
> is.na(U$f16b2n)  <- U$f16b2n==99
> is.na(U$f16b3n)  <- U$f16b3n==98 |
+       U$f16b3n==99
> U$Child.not.at.Home <- ifelse(U$f16a==1,
+       U$f16b1n+U$f16b2n+U$f16b3n,0)
```

Questions `hh2c.b`-`hh2.c.m` are about the relationship between the respondent and other individuals in the household (person b, c, d etc to person

m); category 5 being child of the respondent. First we check if the person is a child to the respondent and then all children are summed over the respondents household adding the variable measuring number of children not at home. Generally codes 98 and 99 are assigned `NA`. (FNSEM, 1993a, p. 318).

```
> hhlist <- c(paste("hh2c.", letters[2:13], sep=""))

> is.na(U[hhlist]) <- U[hhlist]==98 | U[hhlist]==99
> U$Children <-
+       U$Child.not.at.Home + apply(apply(subset(U,
+       select=c(hh2c.b:hh2c.m)), 2,
+       function(x) x==5),1, function(x) sum(x,
+       na.rm=TRUE))
```

### 2.5.11   No British education

Question `q1` asks whether the respondent has any British education. Code 2 is no. Code 8 (Can´t say) is kept in the alternative category since these individuals will answer the question `q3` about foreign education. Code 9 in `q1` is assigned `NA`. (FNSEM, 1993a, p. 96).

```
> is.na(U$q1) <- U$q1==9
> U$No.British.Education <- U$q1==2
```

### 2.5.12   British basic education

We could not exactly see how this variable was defined in Bisin et al. (2008). They define the British high education as A–level and above. One interpretation is then that O-level are educations included in the basic level. This interpretation is implemented here. `NA` is assigned to all observations for which the filter question *No British Education* was `NA`. (FNSEM, 1993a, pp. 96ff).

```
> q2alist <- c(paste("..q2a", c(1:8,12:18), sep=""))

> U$British.Basic.Education <-
+       apply(apply(U[q2alist] ,2,function(x)
+       x==1),1, function(x) sum(x, na.rm=TRUE))!=0
> U$British.Basic.Education <- ifelse(
+       is.na(U$No.British.Education),
+       NA,U$British.Basic.Education)
```

### 2.5.13 British higher education

Bisin et al. (2008) explicitly defined British higher education as A-level. Given the definition of British Basic education, the reference group will include trade apprenticeships as well as university educations. `NA` is assigned to all observations for which the filter question *No British Education* was `NA`. (FNSEM, 1993a, pp. 96ff).

```
> U$British.Higher.Education <- apply(apply(subset(
+       U,select=c(..q2a9:..q2a11,..q2a19,..q2a20))
+       ,2,function(x) x==1),1,
+       function(x) sum(x, na.rm=TRUE))!=0
> U$British.Higher.Education <- ifelse(
+       is.na(U$No.British.Education),
+       NA,U$British.Higher.Education)
```

### 2.5.14 Foreign education

Foreign educations is question `q3` asked to all who answered 'no' or 'Can't say' on question `q1`. The answer yes is coded as (code 1) and no is coded as(code 2). Contrary to the above educational variables `NA` is *not* assigned to all `NA` on *No British Education* since some of them (code 8) actually was asked the question `q3`. Instead `NA` is assigned to all observations for which the filter question `q1` was 1 or 9 and to all code 8 (Can't say) and code 9 on question `q3`. (FNSEM, 1993a, p. 99).

```
> is.na(U$q3) <- U$q3==8 | U$q3==9
> U$Foreign.Education <- U$q3==1
```

### 2.5.15 Labour market status

We code the labour market status using `j1b` (in paid work last week or not) and `j3occ` (classification of activity; either last week's activity or potential activity during the last ten years). The variable `j1b` takes the value 1 for paid work last week and 2 otherwise. The variable `j3occ` is coded as follows: (FNSEM, 1993a,b, the former pp. 81f)

1. Self-employed (25+ employees)
2. Self-employed (1-24 employees)
3. Self-employed (no employees)
4. Self-employed (employees not known)
5. Manager (establishment of 25+ employees)
6. Manager (establishment of 1-24 employees)
7. Manager (employees not known)
8. Foreman/supervisor
9. Other employee

10. Employee status unknown
11. Not known/not answered

**Employee**  In order to be classified as an employee the individual has to have answered yes (value 1) in `j1b` and be classified as employee in `j3occ` (value 9) and have the value `NotAssigned.NA` is assigned to categories 10 and 11 in `j3occ`.

```
> is.na(U$j3occ) <- U$j3occ==10 | U$j3occ==11
> U$Labour.Market.Status <- ifelse(U$j1b==1 &
+      U$j3occ==9 & !is.na(U$j1b==1 &
+      U$j3occ==9), "Employee","NotAssigned")
```

**Self Employed**  Self–employed are also coded using `j1b` and `j3occ`; above. Categories $1 - 4$ in `j3occ` are defined as self-employed. We also require that `j1b` is equal 1 (Self–Employed).

```
> U$Labour.Market.Status <- replace(
+      U$Labour.Market.Status,
+      (U$j3occ==1 | U$j3occ==2 |
+      U$j3occ==3  | U$j3occ==4) &
+      U$j1b==1,"SelfEmployed")
```

**Manager**  Managers are also coded using `j1b` and `j3occ`; see above. Categories $5 - 8$ in `j3occ` are defined as managers (including supervisors).

```
> U$Labour.Market.Status <-
+      replace(U$Labour.Market.Status,
+      U$j1b==1 & U$j3occ>4 & U$j3occ<9,
+      "Manager")
```

**Unemployed**  The question `hh5d.s` describes the respondent's labour market status. Unemployment is defined via this variable. `hh5d.s` is coded in the following way (FNSEM, 1993b):

1. Full-time education
2. Govt. training programme
3. Full-time paid work
4. Part-time paid work
5. Waiting to take up paid work
6. Registered unemployed
7. Unemployed, not registered
8. Permanently sick or disabled
9. Wholly retired from work

10. Looking after the home
11. Doing something else
12. NA

We define unemployed as category 6 and 7. (Bisin et al., 2008, p. 324). There are few cases where the individual is classified as Employee according to our definition above and is reported to be unemployed in `hh5d.s`. This can for example be part-time unemployment. We classify these individuals as having `Unclear` labour market status. These will be checked later and be removed if they are few in the final sample.

```
> U$Labour.Market.Status <-
+       replace(U$Labour.Market.Status,
+       (U$hh5d.s==6 | U$hh5d.s == 7) &
+       U$Labour.Market.Status!="NotAssigned",
+       "Unclear")
> U$Labour.Market.Status <-
+       replace(U$Labour.Market.Status,
+       (U$hh5d.s==6 | U$hh5d.s == 7) &
+       U$Labour.Market.Status=="NotAssigned",
+       "Unemployed")
```

**Out of labour force**  The category out of labour force is defined as those having values (1,2,5,8,9,10,11) in `hh5d.s` or value 2 in `j1b` or `j2`.

```
> U$Labour.Market.Status <-
+       replace(U$Labour.Market.Status,
+       (U$hh5d.s==1 | U$hh5d.s == 2 |
+       U$hh5d.s==5  | U$hh5d.s == 8 |
+       U$hh5d.s==9  | U$hh5d.s == 10|
+       U$hh5d.s==11 | U$j1b    == 2 |
+       U$j2     == 2 ) &
+       U$Labour.Market.Status=="NotAssigned",
+       "OutOfLabourForce")
```

Remaining observations with the value `NotAssigned` in `Labour Market.Status` will be assigned `NA`. At this point there are 4 observations coded as `Unclear`. These are now recoded as `NA`.

```
> is.na(U$Labour.Market.Status) <-
+       U$Labour.Market.Status == "NotAssigned"
> is.na(U$Labour.Market.Status) <-
+       U$Labour.Market.Status == "Unclear"
```

We will create dummy variables using this variable before we run our models.

### 2.5.16 No parents

Variable *No parents* means that the respondent is not living with his or
her parents. This variable is coded with question `f14a` (which takes code 1
for both alive, code 2 father alive, code 3 for mother alive and 8 for both
dead) and `f14b` (which takes code 2 if both living parents do not live with
the respondent and code 6 if the only living parent does not live with the
respondent; else it takes one of the values 1 or $3-5$). *No parents* should
be coded TRUE if either both parents are dead or the respondent does not
live with any living parents. However, since we follow Bisin et al. (2008), we
code this variable including only those who have both their parents dead or
both paprents live away from the respondent. This definition implies that
those who have a parent living away and one parent dead are assigned the
value FALSE. Code 8 and 9 are assigned `NA` (FNSEM, 1993a, p. 56)

```
> is.na(U$f14a) <- U$f14a==9
> is.na(U$f14b) <- U$f14b==9
> U$No.Parents   <- U$f14a==8 | U$f14b==2
```

### 2.5.17 Contacts with parents

The three variables measuring contacts with parents are defined via three
questions asking about the number of physical contacts (question `f15bn`),
the number of contacts via telephone (question `f15cn`) and the number of
contacts via letters (question `f15dn`) that the respondent has had with his
or her parents during the last four weeks conditional on not both parents
being dead. All three takes the value of the underlaying variable if at least
one parent is alive and the value 0 if both parents are dead. Code 999 on
all three variables and code 997 on `f15bn` are assigned `NA`. (FNSEM, 1993a,
p. 56)

```
> is.na(U$f15bn) <- U$f15bn==997 | U$f15bn==999
> is.na(U$f15cn) <- U$f15cn==999
> is.na(U$f15dn) <- U$f15dn==999
> U$Parents.Physical.Contacts <- ifelse(
+       U$f14a!=8,  U$f15bn,0)
> U$Parents.Telephone.Calls <- ifelse(
+       U$f14a!=8, U$f15cn,0)
> U$Parents.Letters <- ifelse(
+       U$f14a!=8, U$f15dn,0)
```

### 2.5.18 English language

There are several language variables measuring whether the respondent is
speaking English with different individuals; at home with older, at home

with younger, at work and with friends. We construct theses variables using question `s12a` which asks whether the respondent regularly speak to anyone in Britain in any other language than English and `..s12f`, `..s12g`, `..s12h` and `..s12i` which asks which language is spoken to the above mentioned categories of individuals. Each of `..s12f`, `..s12g`, `..s12h` and `..s12i` comes in 18 versions (e.g., `..s12f1`,…,`..s12f18`) where each question 1–15 is coded `yes` if the respondent speaks the language. Question 16 is "Never speaks to these people/Not ap[plicable]", question 17 NA and question 18 " None of the above answered positive". Question 1 is always regarding English.

The respondent is coded as English speaker if either `s12a` is answered negatively or `..s12aX1`, where `X=f,g,h,i`, is answered positively. Only respondents for which either `s12a` is `NA` or all of `..s12X1`,…,`..s12X16` are answered negatively are coded as `NA`. Below is the code for *English Spoken At Home With Older*:

```
> is.na(U$s12a) <- U$s12a== 8 | U$s12a == 9
> s12flist <- c(paste("..s12f", 2:16, sep=""))
> U$oOLD <- apply(U[s12flist]=="yes", 1, sum)
> U$English.Spoken.at.Home.with.Older <-
+      ((U$s12a==1 | is.na(U$s12a)) &
+      U$..s12f1=="yes") | U$s12a==2
> is.na(U$English.Spoken.at.Home.with.Older) <-
+      U$oOLD==0 &
+      U$English.Spoken.at.Home.with.Older==FALSE
> U$English.Spoken.at.Home.with.Older <-
+      replace(U$English.Spoken.at.Home.with.Older,
+      U$oOLD>0 &
+      is.na(U$English.Spoken.at.Home.with.Older),FALSE)
> U$DO.NOT.SPEAK.WITH.OLDER <- ifelse(U$..s12f16=="no",0,1)
```

The codes for *English Spoken At Home With Younger*, *English Spoken At Work* and *English Spoken With Friends* are equivalent. These codes are in `araietal_source.Rnw` but not shown here. It is also available in `araietal_source.R`.

### 2.5.19 Household income

The question `hh40` provides information in which interval the household income of the respondent's household is. We assign the midpoints in these intervals as the household income. For the lowest bracket this income is the midpoint of $[0, 77]$. For the highest bracket we assign the income which is the lowest income in the bracket plus the income interval down to the midpint of the second highest bracket; i.e., $789 + \frac{788-731}{2} = 817.5$. This method

16

underestimates the range of the highest bracket but to a lesser extent than the lower limit 789.

```
> is.na(U$hh40) <- {U$hh40=="refused" |
+       U$hh40=="can't say"  | U$hh40=="na"}
> U$Household.Income <- c(38.5,96.5,135.0,
+       173.5,230.5,260.0,318.0,366.0,414.0,
+       471.5,529.5,587.0,649.5,702.0,759.5,
+       817.5)[U$hh40]
```

### 2.5.20   Ward variables

Ward density of own ethnic group is measured by the variable `wown` in the original data. This variable is coded $1 - 7$ depending on the density of the respondent's own ethnic group is in the ward of the respondent (FNSEM, 1993b):

1. Up to 1.99%
2. 2-4.99%
3. 5-9.99%
4. 10-14.99%
5. 15-24.99%
6. 25-32.99%
7. 33% or more

We recode the variable to take the midpoints of the density intervals in the same fashion as the household income variable was recoded. This means that if the respondent is the lowest interval the density is set to be 1 etc. In the highest interval we set the density to be the lowest density in the interval plus the density distance down to the midpoint of the second highest density interval; i.e., $33 + \frac{33-25}{2} = 37$.

The variable `wunemp` is coded $1 - 6$ depending on unemployment rate in the ward of the respondent (FNSEM, 1993b):

1. Up to 1.99%
2. 2-4.99%
3. 5-9.99%
4. 10-14.99%
5. 15-20%
6. 20% or more

We recode this variable to instead take the midpoints of the intervals as we did for the household income. This means that if the respondent is the lowest interval the the rate is set to be 1 etc. In the highest interval we set the rate to be the lowest rate in the interval plus the distance down to the midpoint of the second highest rate interval; i.e., $20 + \frac{20-15}{2} = 22.5$.

```
> U$Ward.Density.Own.Ethnicity <- (c(1.0,3.5,
+       7.5,12.5,20.0,29.0,37.0)[U$wown])/100
> U$Ward.Unemployment.Rate <- c(1.0,3.5,7.5,12.5,
+       17.5,22.5)[U$wunemp]
```

## 2.6 Discrimination own ethnicity

Finally we define a variable describing the discrimination against the own ethnic group. It is defined as the average of the variable *Discrimination* over ethnic groups after the removal of non–availables.

```
> GroupDiscrimination <-
+       tapply(U$Discrimination,U$ethnic,
+               function(x) mean(x, na.rm=TRUE))
> U$Discrimination.Own.Ethnicity <-
+       GroupDiscrimination[U$ethnic]
```

## 2.7 Defining the subset

We define dummy variables for labour market status to have the same variable labels as in Bisin et al. (2008).

```
> U$Employee <- as.numeric(U$Labour.Market.Status=="Employee")
> U$Manager <- as.numeric(U$Labour.Market.Status=="Manager")
> U$Self.Employed <-
+       as.numeric(U$Labour.Market.Status=="SelfEmployed")
> U$OUT.OF.LABOUR.FORCE <-
+       as.numeric(U$Labour.Market.Status=="OutOfLabourForce")
> U$Unemployed <-
+       as.numeric(U$Labour.Market.Status=="Unemployed")
```

We save a data set keeping all observations containing non–availables:

```
> U.Original <- U
```

We then choose the variables to keep in U. This code is in `araietal_source.Rnw` but not shown here. It is also available in `araietal_source.R`.

Finally we remove all observations containing non–availables from U:

```
> U <- na.omit(U)
```

## 2.8 Sample statistics

Table 1 compares the number of observations in this sample *before* (see section 2.7; i.e. data frame U.Original) and *after* (see section 2.7; i.e. data frame U) non–availables are removed with the numbers of observations

reported by the Bisin et al. (2008) study. The number of observations for various groups in the non–Muslim category are not reported in the Bisin et al. (2008) paper. These numbers are therefore missing in the table. The category definitions are from the original dataset and involves no recoding on our part.

After removing observations with missing values on all variables of interest ("After" in Table 1), we are left with 330 Muslims and 488 non-Muslims. The sample selections induced by the choice of variables and the missing values in these variables lead to a loss of 57 percent of the relevant sample of the original data.[3]

The sample means reported in Bisin et al. (2008) seem to be unweighted. Since data instructions says that the data should always be weighted, tables 2 and 3 report weighted and unweighted sample means before and after removal of non-availables and the Bisin et al. (2008) data.[4] Comparing means, the Bisin et al. (2008) data seem to be different from the original sample. The variables *Attitude Towards Inter Marriage* and *Importance of Racial Composition in Schools* in Bisin et al. (2008) data deviate largely from corresponding averages in the original data. The deviation is extreme in case of *Importance of Racial Composition in Schools*. The original sample has a mean for this variable that is 2 percent for non-Muslims (compare with 33 percent in Bisin et al. (2008)) and 6 percent for Muslims (compare with 65 percent in Bisin et al. (2008)). Due to this extremely skewed distribution. it is hardly meaningfull to run a regression on this variable,

Notice that also the distribution of the variable *Importance of Religion* would be extremely skewed using standard coding of this type of variables. Such a coding would imply that religion is important when the respondent answer "Very Important" and "Fairly Important", to the question "How important is religion to the way you live your life?".

The sample means in our data after removing accumulated missing values due to all variables in the estimations deviate marginally in general from the original data. The similarities here are partly due to the fact that the statistics are based exactly on the same variable definition in our implementation. In some respects, the deviations are larger. For further comparisons we refer to Tables 2 and 3.

Due to the fact that the large majority of observations from the original data are lost, the remaining sample is likely to be contaminated with sample selection bias. To compare the characteristics of the remaining sample with the original sample says something about systematic attrition with respect to observables. The sample selection bias with respect to unobservables cannot, however, be resolved.

---

[3]The variables written in capital letters are created to ensure well–defined reference categories. They are included in our regressions, but we cannot say whether they are included in the regressions of Bisin et al. (2008).

[4]See FNSEM (1993b).

# 3 Regression Results

We use linear probability models (LPM).[5] Our results are presented in Tables 4 and 5. Bisin et al. (2008) write that:

1. "Muslims integrate less and more slowly than non-Muslims." (abstract, p. 445) and

2. "...there is no evidence that segregated neighbourhoods breed intense religious and cultural identities. On the contrary, ...intense identities in our data are more prominent in relatively mixed neighbourhoods." (p. 446)

The first claim is based on their reported results concerning the variable *Years Since Arrival*. In this way Bisin et al. (2008) compare cohorts of Muslims and non Muslims and attempt to say something about the evolution of values over time. They do not follow individuals over time but nonetheless call these cohort differences "Integration over time". They report negative coefficients for *Years Since Arrival*, but the estimates are smaller in absolute value for Muslims than for non-Muslims. In our case, the coefficients for *Years Since Arrival* reported in Tables 4 and 5 are insignificant in all cases except in the regression for *Importance of Religion* for Muslims, where it is negative. This is opposite to what Bisin et al. (2008) claim.

The second claim is based on their reported results concerning the variable *Ward Density Own Ethnicity*. Bisin et al. (2008) report negative and significant estimates for *Ward Density Own Ethnicity* in all six specifications. Their negative coefficient for this variable would imply that ethnic minorities put more weight on religion, mind more about inter-ethnic marriage and have stronger taste for ethnically profiled schools, as we move from neighbourhoods (Wards) with high density of their own ethnicity to neighbourhoods where people from their own ethnicity are scarce. This is not at all what we find in our replication.

---

[5]Bisin et al. (2008) use probit estimations. Our attempts to use probit run into convergence problems. The convergence problems are severe for the model using *Importance of Racial Composition in Schools* as dependent variable. Hence, our choice of LPM.

Another issue is that Bisin et al. (2008) should have included dummy variables indicating religious affiliation: Christians, Sikhs and others in the non-Muslim category to check similarities and differences among non-Muslims as well. In this respect we follow their model specification.

Moreover, Bisin et al. (2008) should have adjusted for within ward correlations. This might matter for their standard errors, which might be underestimated. In our case, with almost no significant results, this would not matter much. The variable is not available in the data set and we did not make much effort to obtain it.

All estimated models include 7 UK-region dummies, and the variables `DO.NOT.SPEAK.WITH.OLDER`, `DO.NOT.SPEAK.AT.WORK`, and `DO.NOT.SPEAK.WITH.FRIENDS`. It turned out that the variable `DO.NOT.SPEAK.WITH.YOUNGER` is TRUE for few observations and cannot be included in the model.

In our estimations, the estimated coefficients for this variable are all positive but far from significant. The P-values are 0.69, 0.97 and 0.59 for Muslims and 0.15, 0.14 and 0.3 for non-Muslims contradicting the Bisin et al. (2008) results.

Inspecting the results presented in Tables 4 and 5, there are many similarities and few differences in the estimated coefficients for Muslims and non-Muslims. Our results are generally very different from results reported by Bisin et al. (2008). We are, however, doubtful whether it is possible to draw any reliable inference from these results due to great loss of observations and possible sample selection bias, together with the problem of endogeneity (also mentioned by Bisin et al. (2008)).

## 4 Concluding remarks

The Bisin et al. (2008) paper rests on fragile grounds. Our examination of the data using their variable definitions and the same set-up indicates that their claims about differences between Muslims and non-Muslims, and their conclusion that strong Religious/Ethnic identities are found in mixed neighbourhoods does not hold. There is no systematic relation between ethnic minorities' views on religion, inter-ethnic marriage or ethnic profile of schools and the density of their own ethnic minority in their neighbourhood. However, we hesitate to draw inference from these results since the great loss of observations (57 percent) implies that the remaining sample is most likely not representative.

## 5 Production notes

To facilitate reproducibility and save others timely interpretations of what is done in this paper, we attempt to follow Literate Statistical Programming procedures. For documenting our results we have used **Sweave** by Leisch (2002) in combination with the LaTeX family of programs using the packages `inputenc`, `fontenc`, `natbib`, `Sweave`, `fancyvrb`, `color`, `url`, `hyperref` and `multirow`.

All code (R code, LaTeX code and BibTeX data base code) used to do the econometric estimations, to produce this technical documentation including all tables and to produce the companion paper Arai et al. (2008) is contained in the file `araietal_source.Rnw`.

The estimations and the documentation can be reproduced in by following the instructions at the top of the file http://people.su.se/~lundh/fragile_grounds/araietal_source.Rnw

Our results were obtained on a i486-pc-linux-gnu platform using R version 2.8.1 beta (2008-12-12 r47173) (R Development Core Team, 2008) with

packages **lmtest** 0.9-22 (2008-12-09), **sandwich** 2.1-0 (2008-01-26), **zoo** 1.5-4 (2008-07-09), **foreign** 0.8-29 (2008-08-07) and **xtable** 1.5-4 (2008/10/03).

# References

Mahmood Arai, Jonas Karlsson, and Michael Lundholm. On fragile grounds: A replication of *Are Muslim immigrants different in terms of cultural integration?* Unpublished manuscript, 2008.

R.G. Berthoud, T. Modood, P. Smith, and G. Prior. *Fourth National Survey of Ethnic Minorities, 1993-1994 [computer file].* Colchester, Essex: UK, 1997. URL http://www.data-archive.ac.uk/doc/3685%5Cmrdoc%5CUKDA%5CUKDA_Study_3685_Information.htm. SN: 3685.

Alberto Bisin, Eleonora Patacchini, Thierry Verdier, and Yves Zenou. Are muslim immigrants different in terms of cultural integration? *Journal of the European Economic Association*, 6:445–456, 2008.

FNSEM. *P1312 Fourth National Study of Etnic Minorities. Project instructions.* Social and Community Planning Research, London, 1993a. URL http://www.data-archive.ac.uk/doc/3685/mrdoc/pdf/a3685uab.pdf. Project 3685.

FNSEM. *UK Data Archive Data Dictionary.* Social andCommunity Planning Research, London, 1993b. An RTF file called UKDA-3685-tab/mrdoc/allissue/3685_UKDA_Data_Dictionary.rtf available when the FNSEM data set is downloaded from the UK Data Archive.

Roger Koenker and Achim Zeileis. Reproducible econometric research (a critical review of the state of the art). Report 60, Department of Statistics and Mathematics, Wirtschaftsuniversität Wien, Research Report Series, 2007. URL http:epub.wu-wien.ac.at/dyn/virlib/wp/eng/mediate/epub-wu-01_c75.pdf?ID=epub-wu-01_c75.

Friedrich Leisch. Sweave user manual. 2002. URL http://www.ci.tuwien.ac.at/~leisch/Sweave/.

R Development Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2008. URL http://www.R-project.org. ISBN 3-900051-07-0.

# Appendix: Tables

Table 1: Religious affiliation (absolute (#) and relative (%) numbers), before (columns 1 and 2) and after (columns 3 and 4) removal of `NA` compared with Bisin et al. (2008) (columns 5 and 6).

| Religious affiliation | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Before | | After | | Bisin et al. | |
| | $n = 1901$ | | $n = 818$ | | $n = 5963$ | |
| | # | % | # | % | # | % |
| hindu | 359 | 18.88 | 149 | 18.22 | | |
| sikh | 288 | 15.15 | 86 | 10.51 | | |
| muslim | 852 | 44.82 | 330 | 40.34 | 2369 | 39.73 |
| christian | 357 | 18.78 | 232 | 28.36 | | |
| buddhist | 17 | 0.89 | 9 | 1.10 | | |
| confucian | 1 | 0.05 | 1 | 0.12 | | |
| jain | 7 | 0.37 | 3 | 0.37 | | |
| parsi/zorastrian | 3 | 0.16 | 2 | 0.24 | | |
| rastafarian | 2 | 0.11 | 1 | 0.12 | | |
| jewish | 1 | 0.05 | 0 | 0.00 | | |
| other | 10 | 0.53 | 5 | 0.61 | | |
| na | 0 | 0.00 | 0 | 0.00 | | |
| NA's | 4 | 0.21 | 0 | 0.00 | | |
| All non-Muslims | 1045 | 55.09 | 488 | 59.66 | 3594 | 60.27 |

NOTE: The row names shows exactly how the original data is coded, so that e.g., 'NA's' are true missing values whereas 'na' is coded as religious affiliation 'na'. On the last line non–Muslims are calculated excluding na and NA.

Table 2: Weighted and Unweighted Means for Muslims and non–Muslims before and after removal of `NA` compared with Bisin et al. (2008).

| | Muslim | | | | | Non-Muslim | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Weighted | | Unweighted | | | Weighted | | Unweighted | | |
| | Before | After | Before | After | Bisin et al. | Before | After | Before | After | Bisin et al. |
| Importance of Religion | 0.79 | 0.81 | 0.80 | 0.85 | 0.79 | 0.49 | 0.48 | 0.48 | 0.48 | 0.42 |
| Attitude Towards Inter Marriage | 0.45 | 0.39 | 0.45 | 0.39 | 0.70 | 0.19 | 0.15 | 0.21 | 0.17 | 0.37 |
| Importance of Racial Composition in Schools | 0.06 | 0.08 | 0.06 | 0.07 | 0.65 | 0.02 | 0.01 | 0.02 | 0.02 | 0.33 |
| Age at Arrival | 22.49 | 20.73 | 22.36 | 20.32 | 39.18 | 22.31 | 17.85 | 22.83 | 18.71 | 42.57 |
| Age | 40.42 | 42.01 | 40.18 | 40.56 | | 44.40 | 43.98 | 44.90 | 44.59 | |
| Female | 0.50 | 0.47 | 0.50 | 0.45 | 0.47 | 0.52 | 0.53 | 0.52 | 0.52 | 0.48 |
| Born in the UK | 0.08 | 0.08 | 0.06 | 0.06 | 0.21 | 0.13 | 0.18 | 0.12 | 0.16 | 0.28 |
| Arranged Marriage | 0.65 | 0.63 | 0.65 | 0.63 | 0.22 | 0.27 | 0.18 | 0.30 | 0.22 | 0.12 |
| Discrimination | 0.23 | 0.20 | 0.20 | 0.16 | 0.17 | 0.38 | 0.41 | 0.35 | 0.39 | 0.19 |
| Children | 3.12 | 3.41 | 3.23 | 3.43 | 2.17 | 2.37 | 2.32 | 2.44 | 2.36 | 1.68 |
| Years Since Arrival | 19.78 | 19.27 | 18.99 | 18.79 | 26.43 | 24.19 | 20.68 | 24.06 | 20.95 | 27.08 |
| No British Education | 0.80 | 0.79 | 0.83 | 0.82 | 0.81 | 0.56 | 0.50 | 0.59 | 0.53 | 0.52 |
| British Basic Education | 0.14 | 0.13 | 0.12 | 0.11 | 0.06 | 0.30 | 0.36 | 0.29 | 0.35 | 0.13 |
| British Higher Education | 0.05 | 0.06 | 0.03 | 0.03 | 0.08 | 0.10 | 0.12 | 0.09 | 0.12 | 0.16 |
| Foreign Education | 0.25 | 0.25 | 0.23 | 0.22 | 0.25 | 0.26 | 0.24 | 0.28 | 0.25 | 0.29 |
| Employee | 0.19 | 0.21 | 0.19 | 0.22 | 0.38 | 0.40 | 0.42 | 0.40 | 0.42 | 0.59 |
| Manager | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.07 | 0.08 | 0.06 | 0.07 | 0.04 |

Table 3: Table 2 continued. Weighted and Unweighted Means for Muslims and non–Muslims before and after removal of `NA` compared with Bisin et al. (2008).

| | Muslim | | | | | Non-Muslim | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Weighted | | Unweighted | | | Weighted | | Unweighted | | |
| | Before | After | Before | After | Bisin et al. | Before | After | Before | After | Bisin et al. |
| Self Employed | 0.09 | 0.08 | 0.07 | 0.06 | 0.09 | 0.12 | 0.12 | 0.10 | 0.10 | 0.14 |
| OUT OF LABOUR FORCE | 0.50 | 0.49 | 0.52 | 0.48 | | 0.33 | 0.29 | 0.36 | 0.32 | |
| Unemployed | 0.19 | 0.19 | 0.20 | 0.22 | 0.19 | 0.08 | 0.10 | 0.08 | 0.09 | 0.08 |
| No Parents | 0.65 | 0.70 | 0.67 | 0.70 | 0.34 | 0.64 | 0.65 | 0.66 | 0.66 | 0.32 |
| Parents Physical Contacts | 2.69 | 2.72 | 2.80 | 2.88 | 3.05 | 2.75 | 3.02 | 2.83 | 2.97 | 3.87 |
| Parents Telephone Calls | 2.10 | 2.51 | 2.28 | 2.85 | 3.38 | 3.34 | 3.34 | 3.15 | 3.26 | 4.74 |
| Parents Letters | 0.47 | 0.41 | 0.48 | 0.42 | 0.67 | 0.25 | 0.24 | 0.25 | 0.27 | 0.37 |
| English Spoken at Home with Older | 0.12 | 0.12 | 0.10 | 0.09 | 0.03 | 0.43 | 0.53 | 0.42 | 0.52 | 0.08 |
| DO NOT SPEAK WITH OLDER | 0.02 | 0.04 | 0.01 | 0.02 | | 0.02 | 0.01 | 0.02 | 0.02 | |
| English Spoken at Home with Younger | 0.46 | 0.46 | 0.41 | 0.40 | 0.20 | 0.77 | 0.81 | 0.75 | 0.79 | 0.25 |
| DO NOT SPEAK WITH YOUNGER | 0.00 | 0.00 | 0.00 | 0.00 | | 0.00 | 0.00 | 0.00 | 0.00 | |
| English Spoken at Work | 0.46 | 0.53 | 0.41 | 0.47 | 0.19 | 0.79 | 0.85 | 0.77 | 0.82 | 0.27 |
| DO NOT SPEAK AT WORK | 0.45 | 0.39 | 0.48 | 0.44 | | 0.16 | 0.11 | 0.18 | 0.14 | |
| English Spoken With Friends | 0.50 | 0.55 | 0.46 | 0.51 | 0.22 | 0.79 | 0.86 | 0.77 | 0.83 | 0.27 |
| DO NOT SPEAK WITH FRIENDS | 0.12 | 0.07 | 0.12 | 0.06 | | 0.04 | 0.02 | 0.05 | 0.02 | |
| Household Income | 216.19 | 215.48 | 195.40 | 194.94 | 200.74 | 334.66 | 333.83 | 307.69 | 313.46 | 330.26 |
| Ward Density Own Ethnicity | 0.13 | 0.13 | 0.16 | 0.16 | 0.15 | 0.11 | 0.10 | 0.12 | 0.11 | 0.11 |
| Ward Unemployment Rate | 16.07 | 15.72 | 17.97 | 18.05 | 16.57 | 12.19 | 12.35 | 13.44 | 13.52 | 12.60 |
| Discrimination Own Ethnicity | 0.23 | 0.22 | 0.21 | 0.19 | 0.21 | 0.34 | 0.37 | 0.34 | 0.36 | 0.18 |

Table 4: Regression Results for Muslims 330 and non-Muslims 488 to be compared with Table 2 in Bisin et al. (2008). Heteroskedasticity corrected (HC1) Standard Errors are in parentheses. P-values < 0.05 are marked with *.

| | Importance of Religion | | Inter Ethnic Marriage | | Ethnic Composition of Schools | |
|---|---|---|---|---|---|---|
| | Muslims | non–Muslims | Muslims | non–Muslims | Muslims | non–Muslims |
| (Intercept) | 0.92* | 0.94* | 0.83* | 0.27 | 0.04 | -0.02 |
| | (0.30) | (0.26) | (0.37) | (0.26) | (0.18) | (0.04) |
| Age at Arrival | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| Female | -0.04 | 0.11* | 0.14 | 0.05 | -0.04 | -0.01 |
| | (0.08) | (0.05) | (0.09) | (0.03) | (0.05) | (0.01) |
| Born in the UK | -0.24 | 0.26 | -0.23 | 0.08 | -0.03 | -0.04 |
| | (0.16) | (0.15) | (0.15) | (0.09) | (0.08) | (0.03) |
| Arranged Marriage | 0.11 | -0.08 | 0.09 | 0.20* | -0.06 | 0.01 |
| | (0.06) | (0.07) | (0.07) | (0.07) | (0.05) | (0.03) |
| Discrimination | -0.09 | 0.04 | 0.01 | 0.01 | 0.00 | 0.00 |
| | (0.07) | (0.05) | (0.08) | (0.04) | (0.05) | (0.01) |
| Children | 0.02* | 0.02 | -0.01 | 0.01 | -0.01* | 0.00 |
| | (0.01) | (0.02) | (0.01) | (0.01) | (0.01) | (0.00) |
| Years Since Arrival | -0.01* | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| No British Education | 0.01 | -0.14 | -0.06 | 0.04 | 0.08 | 0.01 |
| | (0.12) | (0.09) | (0.13) | (0.05) | (0.07) | (0.01) |
| British Basic Education | 0.05 | -0.12 | -0.13 | -0.09 | 0.00 | 0.00 |
| | (0.14) | (0.10) | (0.17) | (0.06) | (0.07) | (0.01) |
| British Higher Education | -0.16 | -0.13 | -0.19 | 0.01 | -0.02 | 0.00 |
| | (0.18) | (0.08) | (0.15) | (0.06) | (0.08) | (0.01) |
| Foreign Education | 0.00 | 0.05 | -0.15* | -0.01 | 0.01 | -0.02 |
| | (0.07) | (0.07) | (0.07) | (0.05) | (0.05) | (0.02) |
| Employee | -0.14 | -0.04 | 0.01 | 0.02 | -0.04 | -0.04 |
| | (0.10) | (0.08) | (0.10) | (0.05) | (0.06) | (0.02) |
| Manager | -0.29 | -0.06 | 0.09 | -0.01 | -0.17* | -0.03 |
| | (0.24) | (0.12) | (0.22) | (0.07) | (0.08) | (0.02) |

Table 5: Table 4 continued. Regression Results for Muslims 330 and non-Muslims 488 to be compared with Table 2 in Bisin et al. (2008). Heteroskedasticity corrected (HC1) Standard Errors are in parentheses. P-values < 0.05 are marked with *.

| | Importance of Religion | | Inter Ethnic Marriage | | Ethnic Composition of Schools | |
|---|---|---|---|---|---|---|
| | Muslims | non–Muslims | Muslims | non–Muslims | Muslims | non–Muslims |
| Self Employed | 0.00 | -0.06 | 0.11 | 0.06 | -0.04 | -0.05 |
| | (0.13) | (0.10) | (0.15) | (0.07) | (0.09) | (0.03) |
| Unemployed | -0.07 | 0.07 | 0.12 | 0.07 | -0.07 | -0.04 |
| | (0.08) | (0.10) | (0.10) | (0.08) | (0.07) | (0.02) |
| No Parents | -0.07 | 0.09 | 0.04 | 0.06 | 0.09* | 0.00 |
| | (0.05) | (0.05) | (0.07) | (0.04) | (0.03) | (0.01) |
| Parents Physical Contacts | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00* |
| | (0.00) | (0.00) | (0.01) | (0.00) | (0.00) | (0.00) |
| Parents Telephone Calls | 0.00 | 0.00 | -0.01* | 0.00 | 0.00 | 0.00 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| Parents Letters | -0.02 | 0.07* | 0.03 | 0.01 | -0.01 | 0.00 |
| | (0.03) | (0.03) | (0.03) | (0.02) | (0.02) | (0.00) |
| English Spoken at Home with Older | -0.34* | -0.08 | -0.05 | -0.09 | 0.00 | -0.05 |
| | (0.12) | (0.07) | (0.10) | (0.07) | (0.06) | (0.03) |
| English Spoken at Home with Younger | -0.01 | -0.20* | 0.11 | 0.03 | 0.02 | 0.02 |
| | (0.07) | (0.09) | (0.08) | (0.09) | (0.05) | (0.03) |
| English Spoken at Work | -0.04 | -0.19 | -0.08 | 0.03 | 0.10 | 0.04 |
| | (0.08) | (0.13) | (0.14) | (0.12) | (0.06) | (0.03) |
| English Spoken With Friends | -0.04 | 0.06 | -0.12 | -0.16 | -0.02 | 0.02 |
| | (0.07) | (0.09) | (0.10) | (0.09) | (0.07) | (0.02) |
| Household Income | 0.00 | 0.00 | 0.00 | 0.00* | 0.00 | 0.00 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| Discrimination Own Ethnicity | 0.26 | 0.13 | 0.07 | -0.28 | 0.10 | 0.20 |
| | (0.41) | (0.39) | (0.43) | (0.31) | (0.24) | (0.10) |
| Ward Density Own Ethnicity | 0.13 | -0.41 | -0.01 | 0.36 | 0.10 | -0.06 |
| | (0.32) | (0.29) | (0.30) | (0.24) | (0.19) | (0.06) |
| Ward Unemployment Rate | 0.01* | 0.01* | 0.00 | 0.00 | 0.00 | 0.00 |
| | (0.01) | (0.01) | (0.01) | (0.00) | (0.00) | (0.00) |
| Adjusted R-square | 0.20 | 0.09 | 0.18 | 0.18 | 0.05 | 0.08 |

NOTE: All estimated models include 7 UK-region dummies, and the variables `DO.NOT.SPEAK.WITH.OLDER`, `DO.NOT.SPEAK.AT.WORK`, and `DO.NOT.SPEAK.WITH.FRIENDS`. It turned out that the variable `DO.NOT.SPEAK.WITH.YOUNGER` is TRUE for few observations and cannot be included in the model.

SULCIS is a multi-disciplinary research center focusing on migration and integration funded by a Linnaeus Grant from the Swedish Research Council (VR). SULCIS consists of affiliated researchers at the Department of Criminology, the Department of Economics, the Department of Human Geography, the Department of Sociology and the Swedish Institute for Social Research (SOFI). For more information, see our website: **www.su.se/sulcis**

## SULCIS Working Paper Series

2007:1     Arai, M & Skogman Thoursie, P., "Giving Up Foreign Names: An empirical Examination of Surname Change and Earnings"

2007:2     Szulkin, R. & Jonsson, J.O., "Immigration, Ethnic Segregation and Educational Outcomes: A Multilevel Analysis of Swedish Comprehensive Schools"

2007:3     Nekby, L. & Özcan, G., "Do Domestic Educations Even Out the Playing Field? Ethnic Labor Market Gaps in Sweden"

2007:4     Nekby, L. & Rödin, M., "Acculturation Identity and Labor Market Outcomes"

2007:5     Lundborg, P., "Assimilation in Sweden: Wages, Employment and Work Income"

2007:6     Nekby, L., Rödin, M. & Özcan, G., "Acculturation Identity and Educational Attainmnet"

2007:7     Bursell, M., "What's in a name? A field experiment test for the existence of ethnic discrimination in the hiring process"

2007:8     Bygren, M. & Szulkin, R., "Ethnic Environment during Childhood and the Educational Attainment of Immigrant Children in Sweden"

2008:1     Hedberg, C., "Entrance, Exit and Exclusion: Labour Market Flows of Foreign Born Adults in Swedish "Divided Cities"

2008:2     Arai, M, Bursell, M. & Nekby, L. "Between Meritocracy and Ethnic Discrimination: The Gender Difference"

2008:3     Bunar, N., "Urban Schools in Sweden. Between Social Predicaments, the Power of Stigma and Relational Dilemmas"

2008:4     Larsen, B. and Waisman G., "Who is Hurt by Discrimination?"

2008:5     Waisman, G. and Larsen, B., "Do Attitudes Towards Immigrants Matter?"

2009:1     Arai, M., Karlsson, J. and Lundholm, M., "On Fragile Grounds: A replication of "Are Muslim immigrants different in terms of cultural integration?"

2009:2     Arai, M., Karlsson, J. and Lundholm, M., "On Fragile Grounds: A replication of "Are Muslim immigrants different in terms of cultural integration?" Technical documentation.