

```

%<*paper|techreport|present>
%
%
% TITLE:   On Fragile Grounds:
%          A replication of "Are Muslim immigrants
%          different in terms of cultural integration?"
% AUTHORS: Mahmood Arai, Jonas Karlsson and Michael
%          Lundholm
% CONTACT: mahmood.arai@ne.su.se
%          jonas.karlsson@sofi.su.se
%          michael.lundholm@ne.su.se
% NOTE:    The text is *marginally* revised 2009-09-02 compared
%          with the Working paper verion.
%
%
% This batch file can be used to generate the relevant
% files performing the estimations, presentations and
% documentations of the project by using the R package
% Sweave, the LaTeX DOCSTRIP utility and the standard LaTeX
% family of software using the UN*X commands
%
% echo 'Sweave("araietal_source_bis.Rnw")' | R --vanilla --quiet
% echo 'source("script.R")' | R --vanilla --quiet
%
% or from within R using
%
% Sweave("araietal_source_bis.Rnw")
% source("script.R")
%
% The file araietal_source.Rnw contains all source code
% and run through Sweave it generates the files
% araietal_source.tex, script.R, araietal_source.R,
% araietal.ins and araietal.bib. When the file script.R is
% run first LaTeX's DOCSTRIP utility is run via
% araietal.ins and the files araietal_paper.tex,
% araietal_techreport.tex and araietal_present.tex are
% generated. The script then continues to process these
% files with the LaTeX family of programs, using the
% BibTex bibliography database file araietal.bib, to
% generate the corresponding PDF-files for paper,
% technical documentation and presentation.
%
% Accordingly the file contains three levels of markup
% 1. noweb markup to define the code chunks evaluated
%    when "araietal_source.Rnw" is run through R via Sweave.
% 2. DOCSTRIP markup to define the conditional LaTeX code
%    to be shipped to the different generated LaTeX files.
%    Conditioning is made on the tags "paper" (which
%    generates araietal_paper.tex and contains our
%    research result in an article, "techreport" (which
%    generates araietal_techreport.tex and contains a
%    complete technical documentation of our research) and
%    "present" (which generates araietal_present.tex and
%    contains a slide presentation of our research).
%    when "latex araietal.ins" is run.
% 3. Standard LaTeX markup to be considered when the
%    generated LaTeX files are run with latex/pdflatex.
%
%
%

```

```

%</paper|techreport|present>
%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Setting document classes for LaTeX conditional on the
% type of document (paper, techreport or present)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%<*paper>
\documentclass[a4paper,11pt]{article}
%</paper>
%<*techreport>
\documentclass[a4paper,11pt]{article}
%</techreport>
%<*present>
\documentclass[style=horatio,mode=present,%
    paper=screen]{powerdot}
\pdsetup{method=direct}
%</present>
%<*paper|techreport|present>
%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Loading LaTeX packages
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
\usepackage[latin1]{inputenc}
\usepackage[T1]{fontenc}
\usepackage[round]{natbib}
\usepackage{Sweave,fancyvrb,color,url,hyperref,multirow,%
    paralist,rotating,}
%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% LaTeX definitions
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
\definecolor{Red}{rgb}{0.5,0,0}
\definecolor{Blue}{rgb}{0,0,0.5}
\hypersetup{%
    breaklinks = {true},
    colorlinks = {true},
    linkcolor = {Blue},
    citecolor = {Blue},
    urlcolor = {Red}
}
\newcommand{\code}[1]{\upshape\mdseries\ttfamily #1}
\newcommand{\proglang}[1]{\upshape\mdseries\sffamily #1}
\newcommand{\pkg}[1]{\upshape\bfseries\rmfamily #1}
\newcommand{\email}[1]{\href{mailto:#1}%
    {\normalfont\texttt{#1}}}
%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Setting options for R/Sweave
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
\setkeys{Gin}{width=\textwidth}
\SweaveOpts{keep.source=TRUE,echo=TRUE}
<<Options,echo=FALSE,quite=TRUE>>=
options(width=90,scipen=3,digits=4)
@

```

```

%%
%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% Defining conditional title pages
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%</paper|techreport|present>
%<*paper>
\title{\LARGE\bfseries On Fragile Grounds:\\
  A replication of \emph{Are Muslim immigrants
  different in terms of cultural integration?}\footnote{
  We are grateful to Lena Nekby and participants at the
  SULCIS seminar, Stockholm Univeristy for valuable comments.}}
%</paper>
%<*techreport>
\title{\LARGE\bfseries On Fragile Grounds:\\
  A replication of \emph{Are Muslim immigrants
  different in terms of cultural integration?
  }\\Technical documentation\footnote{
  We are grateful to Lena Nekby and participants at the
  SULCIS seminar, Stockholm Univeristy for valuable comments.}}
%</techreport>
%<*present>
\title{On Fragile Grounds:\\
  A replication of "Are Muslim immigrants
  different in terms of cultural integration?"}
%</present>
%<*paper|techreport>
\author{Mahmood Arai,\footnote{Corresponding author.
  Department of Economics and SULCIS, Stockholm
  University, SE 106~91 Stockholm, Sweden,
  \email{mahmood.arai@ne.su.se}.} { }Jonas
  Karlsson\footnote{The Institute for Social Research and
  SULCIS, Stockholm University,
  \email{jonas.karlsson@sofi.su.se}.} { }and Michael
  Lundholm\footnote{Department of Economics, Stockholm
  University, \email{michael.lundholm@ne.su.se}.}}
%</paper|techreport>
%<*present>
\author{Mahmood Arai, Jonas Karlsson and Michael Lundholm}
%</present>
%%
%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% Commands for the version to be published in the JEEA
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%<*paper>
\usepackage{setspace}
\usepackage{times}
\addtolength\textwidth{.6cm}
\setlength\parindent{2cm}
%</paper>
%<*paper|techreport|present>
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% Starting the LaTeX document and creating title pages
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
\begin{document}
%</paper|techreport|present>
%<*present>
\sffamily

```

```
%</present>
%<*paper|techreport|present>
\maketitle
%</paper|techreport|present>
%<*paper>
\begin{abstract}
This study is a replication of ``Are Muslim Immigrants
Different in terms of Cultural Integration?'' by Alberto
Bisin, Eleonora Patacchini, Thierry Verdier and Yves Zenou,
published in the Journal of the European Economic Association, 6,
445-456, 2008.
```

```
\citet{Bisin08} report that they have 5963 observations in
their study. Using their empirical setup, we can only
identify \input{crossreference3.tex} relevant observations
in the original data. After removing missing values we are
left with \input{crossreference4.tex} observations. We
cannot replicate any of their results and our estimations
yield no support for their claims.
```

```
\end{abstract}
%</paper>
%<*techreport>
\begin{abstract}
This is a technical documentation of \citet{Araietal08a}
which replicates ``Are Muslim Immigrants Different in terms
of Cultural Integration?'' by Alberto Bisin, Eleonora
Patacchini, Thierry Verdier and Yves Zenou, published in the
Journal of the European Economic Association, 6, 445-456, 2008.
```

```
\citet{Bisin08} report that they have 5963 observations in
their study. Using their empirical setup, we can only
identify \input{crossreference3.tex} relevant observations
in the original data. After removing missing values we are
left with \input{crossreference4.tex} observations. We
cannot replicate any of their results and our estimations
yield no support for their claims.
```

```
\end{abstract}
\newpage
\tableofcontents
\newpage
%</techreport>
%<*paper|techreport|present>
%%
%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% Document content starting
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%</paper|techreport|present>
```

```
%<*present>
\begin{slide}{Introduction}
```

```

\vspace{\stretch{1}}
\begin{itemize}
\item This study is a replication of ``Are Muslim
Immigrants Different in terms of Cultural Integration?' by
Alberto Bisin, Eleonora Patacchini, Thierry Verdier and Yves
Zenou, published in the Journal of the European Economic
Association, 6, 445-456, 2008.
\end{itemize}
\vspace{\stretch{1}}
\end{slide}

```

```
%</present>
```

```

%<*paper>
\newpage
\begin{spacing}{2}
%</paper>

```

```

%<*paper|techreport>
\section{Introduction}

```

```

This is a replication of the empirical
results reported in \citet{Bisin08}. They use British
data and analyse how Muslims and non-Muslims
differ in cultural integration measured as (i)
\emph{Importance of Religion}, (ii) \emph{Attitude
Towards Inter Marriage} and (iii) \emph{Importance of
Racial Composition in Schools}.\footnote{To facilitate
comparability we use the same labels on the variables as
\citet{Bisin08}.}
%</paper|techreport>

```

```

%<*present>
\begin{slide}{Their claims}
\vspace{\stretch{1}}
In the abstract of their paper they write:
\begin{center}
\begin{quote}
``\dots Muslims integrate less and more slowly than
non-Muslims. \dots We also find no evidence that
segregated neighbourhoods breed intense religious and
cultural identities for ethnic minorities, especially for
Muslims.''

``\dots On the contrary, \dots
intense identities in our data are more prominent
in relatively mixed neighbourhoods.''
(p. 446)
\end{quote}
\end{center}
\vspace{\stretch{1}}
\end{slide}
%</present>

```

```

%<*paper|techreport>
In the abstract of their paper they write:
\begin{quote}
``\dots Muslims integrate less and more slowly than

```

non-Muslims. \dots We also find no evidence that segregated neighbourhoods breed intense religious and cultural identities for ethnic minorities, especially for Muslims.' ' \citep[p. 245]{Bisin08}

\end{quote}

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 \emph{Attitudes towards Inter-Marriage} between ethnic minority and the majority population captures only inter-ethnic marriage for the Christian ethnic minorities but both inter-ethnic and inter-religious marriage for Muslims.

%</paper|techreport>

%<*present>

\begin{slide}[toc=]{Our concerns}

\vspace{\stretch{1}}

\begin{enumerate}\setcounter{enumi}{1}

\item The measures of cultural values capture ethnic and religious attributes in different degrees for different groups.

\end{enumerate}

\vspace{\stretch{1}}

\end{slide}

%</present>

%<*paper|techreport>

The data set is the \emph{Fourth National Survey of Ethnic Minorities 1993-1994} (FNSEM) covering two samples.\footnote{The data can be accessed from the UK Data Archive (UKDA) via Athens. UK Data Archive is found at \url{http://www.data-archive.ac.uk/} and Athens at \url{http://www.athens.ac.uk/}.} One sample covers the ethnic minority population and another sample covers the white majority population \citep{Berthoud}.

An initial inspection of data disclosed that the number of observations in \citet{Bisin08} exceeded the total number of observations in the ethnic minority sample.\footnote{We communicated this to the authors and they answered that there were some coding errors. We have received revised codes and new versions of their tables. These codes also contain a revision of definitions of various variables and yielded fewer observations than the published version. Still there are more observations 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 the 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 to us and we therefore do not know the exact nature of the original errors.} \citet{Bisin08} report that

they have 5963 observations in their study, whereas the ethnic minority sample consists of `\input{crossreference1.tex}` observations.

Implementing the empirical setup in `\citet{Bisin08}`, we can only identify `\input{crossreference3.tex}` relevant observations. After removing missing values on all variables of interest we are left with `\input{crossreference4.tex}` observations. Using the remaining sample and running their specifications, we find no results that support their claims. Our replication therefore stopped at this stage and we did not perform any further sensitivity analysis. The great loss of observations implies that the remaining sample is most likely not representative. Therefore, we hesitate to draw inference from the regressions results.

`%</paper|techreport>`

`%<*present>`

`\begin{slide}{Our results}`

`\vspace{\stretch{1}}`

`\begin{itemize}`

`\item` In their paper the number of observations is 5963 which is `$$\input{crossreference0.tex}$$%` of the number of observations in the relevant sample of the original data (`$$\input{crossreference3.tex}$$`).

`\item` Implementing their variable definitions we loose `$$\input{crossreference2.tex}$$%` of the original sample and have only `$$\input{crossreference4.tex}$$` observations.

`\item` Using the remaining sample and running their specifications, we find no results that support their claims.

`\end{itemize}`

`\vspace{\stretch{1}}`

`\end{slide}`

`%</present>`

`%<*paper|techreport>`

In this paper we document the replication and report and comment on results based on the variable definitions, the variable names and the specifications as described in `\citet{Bisin08}`. We choose a procedure that makes it easy to reproduce our results. Influenced by `\cite{Koenker07}`, we use an integrated approach where data management, estimations, and the text that rely on these computations are all integrated into 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.

`%</paper|techreport>`

`%<*paper>`

For details of our analysis of the data and implementation

of the variable definitions, see `\citet{Araietal08b}`, which is a technical companion to this paper.

`%</paper>`

`%<*paper|techreport>`

All data analyses are made in `\proglang{R}` `\citep{Rcore}` and all code files related to this project can be found on `\url{http://people.su.se/~lundh/fragile_grounds/}`.

`%</paper|techreport>`

`%<*techreport>`

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 `\ref{Data}`. Regression results are presented in Section `\ref{Regression}`. The paper is concluded in Section `\ref{Discussion}`. Finally the production procedure is described in Section `\ref{sec:prodnotes}`.

`%</techreport>`

`%<*paper>`

The remainder of the paper is organised as follows. The data are described in Section `\ref{Data}`. Regression results are presented in Section `\ref{Regression}` and finally the paper is concluded in Section `\ref{Discussion}`.

`%</paper>`

`%<*present>`

`\begin{slide}[toc=What we do]{What we wanted to do, what we did and what we didn't}`

`\vspace{\stretch{1}}`

`\begin{itemize}`

`\item We wanted to:`

`\begin{itemize}`

`\item replicate their study,`

`\item check the robustness of their results`

`\end{itemize}`

`\item What we do`

`\begin{itemize}`

`\item document the replication and report`

`and comment results using their variable definitions`

`etc`

`\end{itemize}`

`\item What we didn't do:`

`\begin{itemize}`

`\item almost no sensitivity checking`

```

\end{itemize}
\end{itemize}
\vspace{\stretch{1}}
\end{slide}

```

```
%</present>
```

```

%<*paper|techreport>
\section{Data and variable description}
\label{Data}
%</paper|techreport>

```

```

%<*present>
\begin{slide}{Data}
\vspace{\stretch{1}}
\begin{itemize}
\item Fourth National Survey of Ethnic Minorities 1993-1994
\item UK Data Archive (UKDA) via Athens.
\end{itemize}
\vspace{\stretch{1}}
\end{slide}
%</present>

```

```

%<*techreport>
\subsection{Data}

```

The main objective of the the `\emph{Fourth National Survey of Ethnic Minorities 1993-1994}` (FNSEM) was:

```

\begin{quote}
\begin{itemize}
\item ``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
\item to assess changes over time through comparisons
with other work
\item to show how the position of ethnic minority
groups is related to the social and ethnic compositions
of the areas in which they live
\item to explore diversity among different ethnic
minority groups
\item to describe perceptions and experience of
racial discrimination and social harassment''
\end{itemize}
\citet{Berthoud}
\end{quote}

```

For our coding we have used `\citet{FNSEM}`, which contains the project instructions, and `\citet{FNSEM93b}`, 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 `\proglang{R}` code `\citep{Rcore}` with extensive comments and discussions. For details about our working procedures and how we document the research, see Section `\ref{sec:prodnotes}` on page `\pageref{sec:prodnotes}`. In the code chunks ```$>$''` denotes the `\proglang{R}` prompt and ```$+$''` continuation of the previous line.

`\subsection{Reading data and selecting variables}`

We load package `\pkg{foreign}` to read STATA data format. Data is read from the unpacked Stata-version of the data and ```\verb+_+''` in variable names are converted to ```$. $''`.

`<<Reading data>>=`

```
library(foreign)
```

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

```
@
```

`<<Selecting variables, echo=FALSE>>=`

```
FNSEM<-subset(FNSEM,select=c(
  a1an,a1e,a3,a4n,area,ethnic,
  f14a,f14b,f15a,f15bn:f15dn,f16a,f16b1n:f16b3n,
  hh2a.s,hh2c.b,hh2c.c,hh2c.d,hh2c.e,hh2c.f,
    hh2c.g,hh2c.h,hh2c.i,hh2c.j,hh2c.k,
    hh2c.l,hh2c.m,hh5d.s,hh40,
  j1b,j2,j3occ,j55a,j63a,
  q1,q3,question,
  s6,s7,s9,s12a,s12b,s12c,s24a,s34a,s34b,s39,
  v1a,v1b,v1c,v1d,v9a,
  weightis,wown,wunemp,
  year,
  ..q2a1,..q2a2,..q2a3,..q2a4,..q2a5,..q2a6,
    ..q2a7,..q2a8,..q2a9,..q2a10,..q2a11,
    ..q2a12,..q2a13,..q2a14,..q2a15,
    ..q2a16,..q2a17,..q2a18,..q2a19,
    ..q2a20,..q2a21,..q2a22,..q2a23,
    ..q2a24,..q2a25,..q2a26,..q2a27,
    ..q2a28,..q2a29,..q2a30,..q2a31,
    ..q2a32,..q2a33,
  ..s12f1,..s12f2, ..s12f3, ..s12f4, ..s12f5,
    ..s12f6, ..s12f7, ..s12f8, ..s12f9,
    ..s12f10,..s12f11,..s12f12,..s12f13,
    ..s12f14,..s12f15,..s12f16,..s12f17,
    ..s12f18,
  ..s12g1,..s12g2, ..s12g3, ..s12g4, ..s12g5,
    ..s12g6, ..s12g7, ..s12g8, ..s12g9,
    ..s12g10,..s12g11,..s12g12,..s12g13,
    ..s12g14,..s12g15,..s12g16,..s12g17,
    ..s12g18,
  ..s12h1,..s12h2, ..s12h3, ..s12h4, ..s12h5,
    ..s12h6, ..s12h7, ..s12h8, ..s12h9,
    ..s12h10,..s12h11,..s12h12,..s12h13,
    ..s12h14,..s12h15,..s12h16,..s12h17,
    ..s12h18,
  ..s12i1,..s12i2, ..s12i3, ..s12i4, ..s12i5,
    ..s12i6, ..s12i7, ..s12i8, ..s12i9,
```

```

    ..s12i10,..s12i11,..s12i12,..s12i13,
    ..s12i14,..s12i15,..s12i16,..s12i17,
    ..s12i18

```

```
))
```

```
@
```

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

```
%</techreport>
```

```
%<*present>
```

```

\begin{wideslide}{Sample}
\vspace{\stretch{1}}
\footnotesize{
<<Ethnic, echo=TRUE>>=
t(table(FNSEM$ethnic,exclude=c()))
@
}
\begin{itemize}
\item Remove Whites
\end{itemize}
\vspace{\stretch{1}}
\end{wideslide}

```

```

\begin{slide}[toc=]{Sample (cont.)}
\vspace{\stretch{1}}
<<Religious, echo=TRUE>>=
table(FNSEM$s6,exclude=c())
@
\begin{itemize}
\item Remove non--religious (s6=2)
\end{itemize}
\vspace{\stretch{1}}
\end{slide}

```

```

\begin{slide}[toc=]{Sample (cont.)}
\vspace{\stretch{1}}
<<Singles, echo=TRUE>>=
table(FNSEM$a1e,exclude=c())
@
\begin{itemize}
\item Remove singles (a1e=3)
\end{itemize}
\vspace{\stretch{1}}
\end{slide}

```

```

\begin{slide}[toc=]{Sample (cont.)}
\vspace{\stretch{1}}
\begin{itemize}
\item One of three possible questionnaires (green (category 1), yellow (category 2) and pink (category 3)) were used.
\item Questions involved in the study are in the green questionnaire.

```

```

\end{itemize}
<<Questionnaire, echo=TRUE>>=
table(FNSEM$question,exclude=c())
@

\begin{itemize}
\item Remove pink and yellow (question=2,3)
\end{itemize}
\vspace{\stretch{1}}
\end{slide}

%</present>

%<*techreport>
\subsection{Defining the relevant (ethnic minority) sample}
\label{sec:definingsample}
The data consist of two samples, \emph{Ethnic Minorities}
and \emph{Whites}. We are only interested in the former
and remove all Whites. The variable \code{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
\citet{Bisin08} is \emph{Importance of Religion}.
Whether a respondent has a religion or belongs to a
church is registered in question \code{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 \citet{Bisin08} 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
questionnaires (green (category 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.

<<The ethnic minority sample 1,keep.source=FALSE>>=
# table(U$ethnic,exclude=c())
# table(U$s6,exclude=c())
# table(U$a1e,exclude=c())
@
<<The ethnic minority sample 2,>>=
FNSEM <- FNSEM[FNSEM$ethnic!="white" &
!is.na(FNSEM$ethnic),]
@
<<The ethnic minority sample 3,echo=FALSE>>=
write(nrow(FNSEM),file="crossreference1.tex")
@
<<The ethnic minority sample 4,>>=
FNSEM <- FNSEM[FNSEM$s6!=2,]
FNSEM <- FNSEM[FNSEM$a1e!=3,]
U <- FNSEM <- FNSEM[FNSEM$question==1,]

```

```
@
<<The ethnic minority sample 5,keep.source=FALSE,>>=
# table(U$ethnic,exclude=c())
# table(U$s6,exclude=c())
# table(U$ale,exclude=c())
@
```

An issue where \citet{Bisin08} 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 \citet{Bisin08} implies that White Muslims are excluded and native ethnic minority Muslims are included. This sample definition does not match \citet{Bisin08} writing using the terms \emph{Muslim and non-Muslim immigrants}, as the sample includes natives.

\subsection{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 \proglang{R}; i.e. code \code{NA}. In the data set genuine non--availables are generally coded as $\$-1\$$. We set all $\$-1\$$ to \code{NA} in the entire data set:

```
<<Recoding of missing values>>=
is.na(U) <- U==-1
@
```

In addition to $\$-1\$$ several other codes (\code{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 \ref{sec:subset}. In questions following a filter question \code{NA} may have to be set in a category. Some of these other codes used to denote unknowns are genuine \code{NA}'s and has to be removed. Others will be included in a category. This is done variable by variable below.

\subsection{Variable definitions}

Using the same variable names as in \citet{Bisin08}, we define the variables at the precision described by the authors. We here give our interpretation of the variable definitions in \citet{Bisin08}.

\subsubsection{Religious affiliation}

Question \code{s6} asks whether the respondent belongs to a church or has a religion. For those who answer yes, question \code{s7} asks which that church or religion is. We define two religious affiliations: \code{muslim}, (which are all who answered category 3 (muslim) on question \code{s7}) and \code{non-muslim} (all who did not answer category 3 (muslim) on question \code{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`.

```
\citep[p. 112f]{FNSEM}
<<Religious Affiliation 1,keep.source=FALSE>>=
# table(U$s7,exclude=c())
@
<<Religious Affiliation 2>>=
is.na(U$s7) <- U$s7=="na"
U$Religion <- ifelse(U$s7=="muslim","muslim","non-muslim")
@
<<Religious Affiliation 3,keep.source=FALSE>>=
# table(U$Religion,exclude=c())
@
```

\subsubsection{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 `Bisin08` 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 `Bisin08`, 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` [p. 112f]

```
<<Importance of religion 1,keep.source=FALSE>>=
# table(U$s9,exclude=c())
@
<<Importance of religion 2>>=
is.na(U$s9) <- U$s9==8 | U$s9==9
U$Importance.of.Religion <- U$s9 == 4
@
<<Importance of religion 3,keep.source=FALSE>>=
# table(U$Importance.of.Religion,exclude=c())
@
```

`%</techreport>`

`%<*present>`

```
\begin{slide}[toc=Two important variables]{Religious
affiliation vs. importance of religion}
\space{\stretch{1}}
```

```
<<Rowtable function, echo=FALSE>>=
RowTable <- function(x1,x2)
round(cbind(table(x1,x2,exclude=c()),table(x1,exclude=c()))/c(table(x1, exclude=c()),2)
@
\footnotesize{
<<Religion vs Importance of Religion, echo=TRUE>>=
```

```
RowTable(U$s7,U$s9)
```

```
@
}
\vspace{\stretch{1}}
\end{slide}
%</present>
```

```
%<*techreport>
\subsubsection{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 `{s1}`) on `{s34a}`. We code those who answer yes on both questions (Mind `{s}` & Mind very much) as TRUE. The category FALSE refers then to those who (Do not mind) or (Mind `{s}` & Mind Little). Code `{s8}` (Can't say) on `{s34a}` and code `{s9}` on `{s34a}` and on `{s34b}` are assigned as `{NA}`. `{citep[p. 125]{FNSEM}}`

```
<<Attitude towards inter-marriage 1,keep.source=FALSE>>=
```

```
# table(U$s34a,exclude=c())
# table(U$s34b,exclude=c())
```

```
@
<<Attitude towards inter-marriage 2>>=
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
```

```
@
<<Attitude towards inter-marriage 3,keep.source=FALSE>>=
```

```
# table(U$Attitude.Towards.Inter.Marriage,exclude=c())
```

```
@
```

```
%</techreport>
```

```
%<*present>
```

```
\begin{slide}{Attitude towards inter-marriage}
\vspace{\stretch{1}}
\footnotesize{
What you mind very much if \dots
<<Attitudes, echo=TRUE>>=
RowTable(U$s7,U$Attitude.Towards.Inter.Marriage)
@
}
\vspace{\stretch{1}}
\end{slide}
%</present>
```

```
%<*techreport>
```

```
\subsubsection{Importance of racial composition in schools}
```

Two questions are asked: `{s23}` is

```
\begin{quote}
``If you were choosing a school for an eleven-year old
child of yours, would your choice be influenced by how
```

many (RESPONDENT'S ETHNIC ORIGIN) children there were in the school?' ' \citep[p. 120]{FNSEM}

\end{quote}

and \code{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. \code{s23} is not a filter question. \emph{Importance of Racial Composition in Schools} is set to TRUE if \code{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 \code{NA}. \citep[p. 120]{FNSEM}

```
<<Importance of racial composition in schools 1,keep.source=FALSE>>=
```

```
# table(U$s24a,exclude=c())
```

```
@
```

```
<<Importance of racial composition in schools 2>>=
```

```
is.na(U$s24a) <- U$s24a==8 | U$s24a==9
```

```
U$Importance.of.Racial.Composition.in.Schools <- U$s24a==3
```

```
@
```

```
<<Importance of racial composition in schools 3,keep.source=FALSE>>=
```

```
# table(U$Importance.of.Racial.Composition.in.Schools,
```

```
#   exclude=c())
```

```
@
```

```
%</techreport>
```

```
%<*present>
```

```
\begin{slide}{Importance of racial composition in schools}
```

```
\vspace{\stretch{1}}
```

```
\footnotesize{
```

```
<<Importance, echo=TRUE>>=
```

```
RowTable(U$Religion,
```

```
  U$Importance.of.Racial.Composition.in.Schools)
```

```
@
```

```
}
```

```
\vspace{\stretch{1}}
```

```
\end{slide}
```

```
\begin{slide}[toc=]{}>
```

```
\end{slide}
```

```
\begin{slide}[toc=]{}>
```

```
\end{slide}
```

```
\begin{slide}[toc=]{}>
```

```
\end{slide}
```

```
\begin{slide}[toc=]{}>
```

```
\end{slide}
```

```
\begin{slide}[toc=]{}>
```

```
\end{slide}
```

```
%</present>
```

```
%<*techreport>
```

```
\subsubsection{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}`. `{citep}`[p. 107]{FNSEM}.

```
<<Born in the UK 1,keep.source=FALSE>>=
```

```
#table(U$a3,exclude=c())
```

```
@
```

```
<<Born in the UK 2>>=
```

```
is.na(U$a3) <- U$a3==99
```

```
U$Born.in.the.UK <- U$a3==16 | U$a3==17 | U$a3==18
```

```
@
```

```
<<Born in the UK 3,keep.source=FALSE>>=
```

```
#table(U$Born.in.the.UK,exclude=c())
```

```
@
```

```
\subsubsection{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 `{alan}`) 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 `{Bisin08}`. 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}`. `{citep}`[pp. 105, 107]{FNSEM}.

```
<<Age at and years since arrival 1,keep.source=FALSE>>=
```

```
# table(U$a4n,exclude=c())
```

```
# table(U$alan,exclude=c())
```

```
# table(U$year,exclude=c())
```

```
@
```

```
<<Age at and years since arrival 2>>=
```

```
is.na(U$a4n) <- U$a4n==98 | U$a4n==99
```

```
is.na(U$year) <- U$year==99
```

```
U$Age <- U$alan
```

```
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)
@
<<Age at and years since arrival 3,keep.source=FALSE>>=
# table(U$Years.Since.Arrival,exclude=c())
# table(U$Age.at.Arrival,exclude=c())
@

```

\subsubsection{Female}

Definition of females via question `{hh2a.s}`. Code `{na}` is coded as `{NA}`. [\citep\[pp. 318\]{FNSEM}](#).

```

<<Female 1,keep.source=FALSE>>=
# table(U$hh2a.s,exclude=c())
@
<<Female 2>>=
is.na(U$hh2a.s) <- U$hh2a.s == "na"
U$Female <- U$hh2a=="female"
@
<<Female 3,keep.source=FALSE>>=
# table(U$Female,exclude=c())
@

```

\subsubsection{Arranged marriage}

\label{sec: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 \ref{sec:definingsample}](#)). 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}`. [\citep\[pp. 127\]{FNSEM}](#).

```

<<Arranged marriage 1,keep.source=FALSE>>=
# table(U$s39,exclude=c())
@
<<Arranged marriage 2>>=
is.na(U$s39) <- U$s39==8 | U$s39==9
U$Arranged.Marriage <- ifelse(
  U$s39==3 | U$s39==4 | U$s39==5 |
  U$ethnic=="caribbean", FALSE , TRUE)
@
<<Arranged marriage 3,keep.source=FALSE>>=
# table(U$Arranged.Marriage,exclude=c())
@

```

\subsubsection{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 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}`. `{citep[pp. 154ff, 163, 195 and 199]{FNSEM}}`.

```
<<Discrimination 1>>=
```

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

```
@
```

```
<<Discrimination 3>>=
```

```
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
```

```
@
```

```
<<Discrimination 4,keep.source=FALSE>>=
```

```
# table(U$Discrimination,exclude=c())
```

```
@
```

`{subsubsection}{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. `{citep[p. 57]{FNSEM}}`.

```
<<Children 1,keep.source=FALSE>>=
```

```
# table(U$f16a,exclude=c())
```

```
# table(U$f16b1n,exclude=c())
# table(U$f16b2n,exclude=c())
# table(U$f16b3n,exclude=c())
@
<<Children 2>>=
is.na(U$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}`-`{hh2c.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}`. [\citep\[p. 318\]{FNSEM}](#).

```
<<Children 3>>=
hhlist <- c(paste("hh2c.", letters[2:13], sep=""))
@
```

```
<<Children 4>>=
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))
@
```

```
<<Children 5,keep.source=FALSE>>=
# table(U$Children,exclude=c())
@
```

\subsubsection{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}`. [\citep\[p. 96\]{FNSEM}](#).

```
<<No British Education 1,keep.source=FALSE>>=
# table(U$q1,exclude=c())
@
```

<<No British Education 2>>=

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

<<No British Education 3,keep.source=FALSE>>=

```
# table(U$No.British.Education,exclude=c())
@
```

\subsubsection{British basic education}

We could not exactly see how this variable was defined in \citet{Bisin08}. They define the British high education as A-level and above. One interpretation is then that 0-level are educations included in the basic level. This interpretation is implemented here. \code{NA} is assigned to all observations for which the filter question \emph{No British Education} was \code{NA}. \citet[pp. 96ff]{FNSEM}.

```
<<British basic education 1>>=
q2alist <- c(paste("../q2a", c(1:8,12:18), sep=""))
@
<<British basic education 2,keep.source=FALSE>>=
# apply(U[q2alist], 2, function(x) table(x,
#   exclude=c()))
@
<<British basic education 3>>=
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)
@
<<British basic education 4,keep.source=FALSE>>=
# table(U$British.Basic.Education,exclude=c())
@
```

\subsubsection{British higher education}
\citet{Bisin08} 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. \code{NA} is assigned to all observations for which the filter question \emph{No British Education} was \code{NA}. \citet[pp. 96ff]{FNSEM}.

```
<<British higher education 1,keep.source=FALSE>>=
# table(U$..q2a9,exclude=c())
# table(U$..q2a10,exclude=c())
# table(U$..q2a11,exclude=c())
@
<<British higher education 2>>=
U$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$British.Higher.Education)
@
<<British higher education 3,keep.source=FALSE>>=
# table(U$British.Higher.Education,exclude=c())
@
```

\subsubsection{Foreign education}

Foreign educations is question \code{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}`. *[p. 99]{FNSEM}*.

```
<<Foreign education 1,keep.source=FALSE>>=
```

```
# table(U$q3,exclude=c())
```

```
@
```

```
<<Foreign education 2>>=
```

```
is.na(U$q3) <- U$q3==8 | U$q3==9
```

```
U$Foreign.Education <- U$q3==1
```

```
@
```

```
<<Foreign education 3,keep.source=FALSE>>=
```

```
# table(U$Foreign.Education,exclude=c())
```

```
@
```

`{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: *[the former pp. 81f]{FNSEM,FNSEM93b}*

```
\vspace{6pt}
```

```
\begin{compactenum}
```

```
\item Self-employed (25+ employees)
```

```
\item Self-employed (1-24 employees)
```

```
\item Self-employed (no employees)
```

```
\item Self-employed (employees not known)
```

```
\item Manager (establishment of 25+ employees)
```

```
\item Manager (establishment of 1-24 employees)
```

```
\item Manager (employees not known)
```

```
\item Foreman/supervisor
```

```
\item Other employee
```

```
\item Employee status unknown
```

```
\item Not known/not answered
```

```
\end{compactenum}
```

```
\vspace{6pt}
```

`{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}`.

```
<<Labour market status 1,keep.source=FALSE>>=
```

```
# table(U$j1b,exclude=c())
```

```
# table(U$j3occ,exclude=c())
```

```
@
```

```
<<Labour market status 2>>=
```

```
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")
@
```

\paragraph{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).

```
<<Labour market status 3>>=
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")
@
```

\paragraph{Manager}

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

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

\paragraph{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

`{FNSEM93b}`:

\vspace{6pt}

\begin{compactenum}

\item Full-time education

\item Govt. training programme

\item Full-time paid work

\item Part-time paid work

\item Waiting to take up paid work

\item Registered unemployed

\item Unemployed, not registered

\item Permanently sick or disabled

\item Wholly retired from work

\item Looking after the home

\item Doing something else

\item NA

\end{compactenum}

\vspace{6pt}

We define unemployed as category \$6\$ and \$7\$. `{p.324}{Bisin08}`.

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.

```
<<Labour market status 5>>=
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}`.

```
<<Labour market status 6>>=
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

```
\Sexpr{table(U$Labour.Market.Status)[["Unclear"]]}
observations coded as {Unclear}. These are now recoded
as {NA}.
```

```
<<Labour market status 8>>=
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.

```
<<Labour market status 9,keep.source=FALSE>>=
# table(U$Labour.Market.Status,exclude=c())
```

@

`{No parents}`

Variable `\emph{No parents}` means that the respondent is not living with his or her parents. This variable is coded with question `\code{f14a}` (which takes code `1` for both alive, code `2` father alive, code `3` for mother alive and `8` for both dead) and `\code{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 `$3-5$`).

`\emph{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 `\citet{Bisin08}`, we code this variable including only those who have both their parents dead or both parents 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 `\code{NA}` `\citep[p. 56]{FNSEM}`

```
<<No parents 1,keep.source=FALSE>>=
```

```
# table(U$f14a,exclude=c())
```

```
# table(U$f14b,exclude=c())
```

```
@
```

```
<<No parents 2>>=
```

```
is.na(U$f14a) <- U$f14a==9
```

```
is.na(U$f14b) <- U$f14b==9
```

```
U$No.Parents <- U$f14a==8 | U$f14b==2
```

```
@
```

```
<<No parents 3,keep.source=FALSE>>=
```

```
# table(U$No.Parents,exclude=c())
```

```
@
```

`\subsubsection{Contacts with parents}`

The three variables measuring contacts with parents are defined via three questions asking about the number of physical contacts (question `\code{f15bn}`), the number of contacts via telephone (question `\code{f15cn}`) and the number of contacts via letters (question `\code{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 underlying 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 `\code{f15bn}` are assigned `\code{NA}`. `\citep[p. 56]{FNSEM}`

```
<<Contacts with parents 1,keep.source=FALSE>>=
```

```
# table(U$f15bn,exclude=c())
```

```
# table(U$f15cn,exclude=c())
```

```
# table(U$f15dn,exclude=c())
```

```
@
```

```
<<Contacts with parents 2>>=
```

```
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)
```

```
@
<<Contacts with parents 3,keep.source=FALSE>>=
# table(U$Parents.Physical.Contacts,exclude=c())
# table(U$Parents.Telephone.Calls,exclude=c())
# table(U$Parents.Letters,exclude=c())
@
```

\subsubsection{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 these variables using question \code{s12a} which asks whether the respondent regularly speak to anyone in Britain in any other language than English and \code{..s12f}, \code{..s12g}, \code{..s12h} and \code{..s12i} which asks which language is spoken to the above mentioned categories of individuals. Each of \code{..s12f}, \code{..s12g}, \code{..s12h} and \code{..s12i} comes in 18 versions (e.g., \code{..s12f1},\dots,\code{..s12f18}) where each question 1--15 is coded \code{yes} if the respondent speaks the language. Question 16 is ``Never speaks to these people/Not applicable'', 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 \code{s12a} is answered negatively or \code{..s12aX1}, where \code{X=f,g,h,i}, is answered positively. Only respondents for which either \code{s12a} is \code{NA} or all of \code{..s12X1},\dots,\code{..s12X16} are answered negatively are coded as \code{NA}. Below is the code for \code{English Spoken At Home With Older}:

```
<<English language 1,keep.source=FALSE>>=
# table(U$s12a,exclude=c())
# table(U$..s12f1,exclude=c())
# table(U$..s12g1,exclude=c())
# table(U$..s12h1,exclude=c())
# table(U$..s12i1,exclude=c())
@
<<English language 2>>=
is.na(U$s12a) <- U$s12a== 8 | U$s12a == 9
s12f1list <- c(paste("..s12f", 2:16, sep=""))
U$oOLD <- apply(U[s12f1list]=="yes", 1, sum)
U$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 `\emph{English Spoken At Home With Younger}`, `\emph{English Spoken At Work }` and `\emph{English Spoken With Friends}` are equivalent. These codes are in `\code{araietal_source.Rnw}` but not shown here. It is also available in `\code{araietal_source.R}`.

```
<<English language 3,echo=FALSE,keep.source=FALSE>>=
s12glist <- c(paste("..s12g", 2:16, sep=""))
U$oYOUNG <- apply(U[s12glist]=="yes", 1, sum)
U$English.Spoken.at.Home.with.Younger <-
  ((U$s12a==1 | is.na(U$s12a)) &
   U$..s12g1=="yes") | U$s12a==2
is.na(U$English.Spoken.at.Home.with.Younger) <-
  U$oYOUNG==0 &
  U$English.Spoken.at.Home.with.Younger==FALSE
U$English.Spoken.at.Home.with.Younger <-
  replace(U$English.Spoken.at.Home.with.Younger,
  U$oYOUNG>0 &
  is.na(U$English.Spoken.at.Home.with.Younger),
  FALSE)
U$DO.NOT.SPEAK.WITH.YOUNGER <- ifelse(U$..s12g16=="no",0,1)
@

<<English language 4,echo=FALSE,keep.source=FALSE>>=
s12hlist <- c(paste("..s12h", 2:16, sep=""))
U$oWORK <- apply(U[s12hlist]=="yes", 1, sum)
U$English.Spoken.at.Work <-
  ((U$s12a==1 | is.na(U$s12a)) &
   U$..s12h1=="yes") | U$s12a==2
is.na(U$English.Spoken.at.Work) <-
  U$oWORK==0 & U$English.Spoken.at.Work==FALSE
U$English.Spoken.at.Work <-
  replace(U$English.Spoken.at.Work,
  U$oWORK>0 & is.na(U$English.Spoken.at.Work),
  FALSE)
U$DO.NOT.SPEAK.AT.WORK <- ifelse(U$..s12h16=="no",0,1)
@

<<English language 5,echo=FALSE,keep.source=FALSE>>=
s12ilist <- c(paste("..s12i", 2:16, sep=""))
U$oFRIENDS <- apply(U[s12ilist]=="yes", 1, sum)
U$English.Spoken.With.Friends <-
  ((U$s12a==1 | is.na(U$s12a)) &
   U$..s12i1=="yes") | U$s12a==2
is.na(U$English.Spoken.With.Friends) <-
  U$oFRIENDS==0 &
  U$English.Spoken.With.Friends==FALSE
U$English.Spoken.With.Friends <-
  replace(U$English.Spoken.With.Friends,
  U$oFRIENDS>0 &
  is.na(U$English.Spoken.With.Friends), FALSE)
U$DO.NOT.SPEAK.WITH.FRIENDS <- ifelse(U$..s12i16=="no",0,1)
@

<<English language 6,keep.source=FALSE>>=
# table(U$English.Spoken.at.Home.with.Older,
#   exclude=c())
# table(U$English.Spoken.at.Home.with.Younger,
#   exclude=c())
# table(U$English.Spoken.at.Work,exclude=c())
# table(U$English.Spoken.With.Friends,exclude=c())
```

@

\subsubsection{Household income}

The question \code{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 midpoint of the second highest bracket; i.e., $789 + \frac{788-731}{2} = \text{Sexpr}\{\text{round}(789+(788-731)/2,1)\}$. This method underestimates the range of the highest bracket but to a lesser extent than the lower limit 789 .

```
<<Household income 1,keep.source=FALSE>>=
# table(U$hh40,exclude=c())
```

@

```
<<Household income 2>>=
```

```
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]
```

@

```
<<Household income 3,keep.source=FALSE>>=
# table(U$Household.Income,exclude=c())
```

@

\subsubsection{Ward variables}

Ward density of own ethnic group is measured by the variable \code{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

\citep{FNSEM93b}:

\vspace{6pt}

\begin{compactenum}

\item Up to 1.99\%

\item 2-4.99\%

\item 5-9.99\%

\item 10-14.99\%

\item 15-24.99\%

\item 25-32.99\%

\item 33\% or more

\end{compactenum}

\vspace{6pt}

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., $33 + \frac{33-25}{2} = \text{Sexpr}\{\text{round}(33+(33-25)/2,1)\}$.

The variable `\code{wunemp}` is coded `$1-6$` depending on unemployment rate `in` the ward of the respondent

`\citep{FNSEM93b}`:

`\vspace{6pt}`

`\begin{compactenum}`

`\item Up to 1.99\%`

`\item 2-4.99\%`

`\item 5-9.99\%`

`\item 10-14.99\%`

`\item 15-20\%`

`\item 20\%` or more

`\end{compactenum}`

`\vspace{6pt}`

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.,

`$20+\frac{20-15}{2}=\Sexpr{round(20+(20-15)/2,1)}$.`

`<<Ward variables 1,keep.source=FALSE>>=`

`# table(U$wown,exclude=c())`

`@`

`<<Ward variables 2>>=`

`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]`

`@`

`<<Ward variables 3,keep.source=FALSE>>=`

`# table(U$Ward.Density.Own.Ethnicity,exclude=c())`

`# table(U$Ward.Unemployment.Rate,exclude=c())`

`@`

`\subsection{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 `\emph{Discrimination}` over ethnic groups after the removal of non-`--`availables.

`<<Another variable definition>>=`

`GroupDiscrimination <-`

`tapply(U$Discrimination,U$ethnic,`
`function(x) mean(x, na.rm=TRUE))`

`U$Discrimination.Own.Ethnicity <-`
`GroupDiscrimination[U$ethnic]`

`@`

`\subsection{Defining the subset}`

`\label{sec:subset}`

We define dummy variables `for` labour market status to have the same variable labels as `in` `\citet{Bisin08}`.

`<<Defining dummies for Labour Market Status>>=`

`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--available:

```
<<Defining the subset 1>>=
```

```
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`.

```
<<Defining the subset 2,keep.source=FALSE,echo=FALSE>>=
```

```

U <- subset(U, select=c(
  Religion,
  s7,
  ethnic,
  Importance.of.Religion,
  Attitude.Towards.Inter.Marriage,
  Importance.of.Racial.Composition.in.Schools,
  Age.at.Arrival,
  Age,
  Female,
  Born.in.the.UK,
  Arranged.Marriage,
  Discrimination,
  Children,
  Years.Since.Arrival,
  No.British.Education,
  British.Basic.Education,
  British.Higher.Education,
  Foreign.Education,
  Employee,
  Manager,
  Self.Employed,
  OUT.OF.LABOUR.FORCE,
  Unemployed,
  No.Parents,
  Parents.Physical.Contacts,
  Parents.Telephone.Calls,
  Parents.Letters,
  English.Spoken.at.Home.with.Older,
  DO.NOT.SPEAK.WITH.OLDER,
  English.Spoken.at.Home.with.Younger,
  DO.NOT.SPEAK.WITH.YOUNGER,
  English.Spoken.at.Work,
  DO.NOT.SPEAK.AT.WORK,
  English.Spoken.With.Friends,
  DO.NOT.SPEAK.WITH.FRIENDS,
  Household.Income,
  Ward.Density.Own.Ethnicity,

```

```
Ward.Unemployment.Rate,
Discrimination.Own.Ethnicity,
area,
weightis))
```

@

Finally we remove all observations containing

```
non--availables from \code{U}:
```

```
<<Defining the subset 3>>=
```

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

@

```
\subsection{Sample statistics}
```

```
%</techreport>
```

```
%<*paper>
```

In our definition of the **relevant** (ethnic minority) sample of the \emph{Fourth National Survey of Ethnic Minorities 1993-1994} (FNSEM)

we have **excluded** (i) the UK majority **population** (defined as

Whites in the data set), (ii) all who do not have a

religion or do not belong to a church since they cannot be

classified in a religious group, (iii) all singles since

only people who are married or have been married answer the

question about who made the final decision of their marriage

and (iv) those who answered questionnaire ``yellow'' since they do not answer

relevant questions involved in the study.

```
%</paper>
```

```
%<*paper|techreport>
```

Table \ref{ta:diffsample} compares the number of observations in this sample \emph{before}

```
%</paper|techreport>
```

```
%<*techreport>
```

(see Section \ref{sec:subset}; i.e. data frame

```
\code{U.Original})
```

```
%</techreport>
```

```
%<*paper|techreport>
```

and \emph{after}

```
%</paper|techreport>
```

```
%<*techreport>
```

(see Section \ref{sec:subset}; i.e. data frame \code{U})

```
%</techreport>
```

```
%<*paper|techreport>
```

non--availables are removed with the numbers of observations

reported by the \citet{Bisin08} study. The number of

observations for various groups in the non--Muslim category

are not reported in the \citet{Bisin08} 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

```
\ref{ta:diffsample}), we are left with
```

```
\Sexpr{nrow(U[U$Religion=="muslim",])} Muslims and
```

```
\Sexpr{nrow(U[U$Religion=="non-muslim",])} non-Muslims. The
```

sample selections induced by the choice of variables and the

missing values in these variables lead to a loss of

\input{crossreference2.tex}% of the

relevant sample of the original

data.\footnote{See \citet{FNSEM93b}.}

The sample means reported in \citet{Bisin08} seem to be unweighted. Since data instructions says that the data should always be weighted, Tables \ref{tab:descriptive} and \ref{tab:descriptive2} report weighted and unweighted sample means before and after removal of non-availables and the \citet{Bisin08} data.\footnote{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 \citet{Bisin08}.}

Comparing means, the \citet{Bisin08} data seem to be different from the original sample. The variables \emph{Attitude Towards Inter Marriage} and \emph{Importance of Racial Composition in Schools} in \citet{Bisin08} deviate to a large extent from corresponding averages in the original data. The deviation is extreme in the case of \emph{Importance of Racial Composition in Schools}.

The original sample yields a mean for this variable of $\text{100} \times \text{round}(\text{mean}(\text{U.Original}\$ \text{Importance.of.Racial.Composition.in.Schools}[\text{U.Original}\$ \text{religion} == \text{"non-muslim"}], \text{na.rm} = \text{TRUE}), 2)$ for non-Muslims (compared with 33\% in \citet{Bisin08}) and $\text{100} \times \text{round}(\text{mean}(\text{U.Original}\$ \text{Importance.of.Racial.Composition.in.Schools}[\text{U.Original}\$ \text{religion} == \text{"muslim"}], \text{na.rm} = \text{TRUE}), 2)$ for Muslims (compare with 65\% in \citet{Bisin08}). Due to this extremely skewed distribution, it is hardly meaningful to run a regression on this variable.

The distribution of the variable \emph{Importance of Religion} is also extremely skewed using standard coding of this type of variable. Such a coding would imply that religion is important when the respondent answers ``Very Important' and ``Fairly Important', to the question ``How important is religion to the way you live your life?'. \citet{Bisin08} do not follow this practice and their definition of the \emph{Importance of Religion} variable implies that religion is important when the respondents have answered that religion is ``Very Important' only. Religion is consequently not important when respondents have answered that religion is ``Fairly Important', "Not Very Important" or "Not at all Important".

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 in the statistics reported in columns labelled ``Before'' and ``After'' 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 \ref{tab:descriptive} and \ref{tab:descriptive2}.

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.

```
\section{Regression Results}
\label{Regression}
```

```
%</paper|techreport>
<<Defining models for regressions, echo=FALSE>>=
# Defining the regression model for
# Importance of Religion
# DO.NOT.SPEAK.WITH.YOUNGER, removed,
# no observations for Muslims.
# only two observations for non-Muslims
# table(U$DO.NOT.SPEAK.WITH.YOUNGER, U$Religion)

modell <- {formula(Importance.of.Religion ~
  Age.at.Arrival + Female + Born.in.the.UK +
  Arranged.Marriage + Discrimination +
  Children + Years.Since.Arrival +
  No.British.Education + British.Basic.Education +
  British.Higher.Education + Foreign.Education +
  Employee+ Manager + Self.Employed + Unemployed+
  No.Parents + Parents.Physical.Contacts +
  Parents.Telephone.Calls + Parents.Letters +
  English.Spoken.at.Home.with.Older +
  English.Spoken.at.Home.with.Younger +
  English.Spoken.at.Work +
  English.Spoken.With.Friends +
  Household.Income +
  Discrimination.Own.Ethnicity +
  Ward.Density.Own.Ethnicity +
  Ward.Unemployment.Rate + area +
  factor(DO.NOT.SPEAK.WITH.OLDER)+
  factor(DO.NOT.SPEAK.AT.WORK) +
  factor(DO.NOT.SPEAK.WITH.FRIENDS))}

# Defining the model using the same regressors as
# in Model 1 and Attitude.Towards.Inter.Marriage
# as dependent variable
modell2 <- {update(modell,
  Attitude.Towards.Inter.Marriage ~ .)}

# Defining the model using the same regressors as
# in Model 1 and Importance.of.Racial.Composition.in.Schools
# as dependent variable
modell3 <- {update(modell2,
  Importance.of.Racial.Composition.in.Schools ~ .)}

@

<<Regressions, echo=FALSE>>=
library(sandwich, warn.conflicts=FALSE)
library(lmtest, warn.conflicts=FALSE)
Umuslim <- U[U$Religion=="muslim",]
UNmuslim <- U[U$Religion!="muslim",]
```

```
ESTIMATE <- function(x,df)
  {fm <- lm(x, weights=weightis, data=df)
  rbind(coeftest(fm,
    vcov = vcovHC(fm, type = "HC1")),
    summary(fm)$adj.r.squared)}
```

```
result1.muslim <- ESTIMATE(model1, Umuslim)
result2.muslim <- ESTIMATE(model2, Umuslim)
result3.muslim <- ESTIMATE(model3, Umuslim)
```

```
result1.non.muslim <- ESTIMATE(model1, UNmuslim)
result2.non.muslim <- ESTIMATE(model2, UNmuslim)
result3.non.muslim <- ESTIMATE(model3, UNmuslim)
```

@

```
%<*paper|techreport>
```

We use linear probability models (LPM).^{\footnote{\citet{Bisin08} use probit estimations. Our attempts to use probit run into convergence problems. The convergence problems are severe for the model using \emph{Importance of Racial Composition in Schools} as dependent variable. Hence, our choice of LPM.}

Another issue is that \citet{Bisin08} 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 and do not include these dummy variables.

Moreover, \citet{Bisin08} 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.

```
%</paper|techreport>
```

```
%<*techreport>
```

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

```
%</techreport>
```

```
%<*paper|techreport>
```

} Our results are presented in Tables \ref{ta:regressions} and \ref{ta:regressions2}. \citet{Bisin08} write that: \begin{enumerate} \item ``Muslims integrate less and more slowly than non-Muslims.''' (abstract, p. 445) and \item ``\dots there is no evidence that segregated neighbourhoods breed intense religious and cultural identities. On the contrary, \dots

intense identities [in](#) our data are more prominent [in](#) relatively mixed neighbourhoods.''
(p. 446)

\end{enumerate}

The first claim is based on their reported results concerning the variable `\emph{Years Since Arrival}`. In this way `\citet{Bisin08}` 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](#) `\emph{Years Since Arrival}`, but the estimates are smaller [in](#) absolute value [for](#) Muslims than [for](#) non-Muslims. In our case, the coefficients [for](#) `\emph{Years Since Arrival}` reported [in](#) Tables `\ref{ta:regressions}` and `\ref{ta:regressions2}` are insignificant [in](#) all cases except [in](#) the regression [for](#) `\emph{Importance of Religion}` [for](#) Muslims, where it is negative. This means that there are no differences between cohorts of non-Muslims arriving [in](#) different time periods but there are differences between various cohorts of Muslims implying that those who have stayed a longer period [in](#) the UK display weaker values as compared to those who have been a shorter period of time [in](#) the UK. This is opposite to what `\citet{Bisin08}` claim.

The second claim is based on their reported results concerning the variable `\emph{Ward Density Own Ethnicity}`. `\citet{Bisin08}` report negative and significant estimates [for](#) `\emph{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. In our estimations, the estimated coefficients [for](#) this variable are all positive but far from significant. The P-values are

```
\Sexpr{round(result1.muslim["Ward.Density.Own.Ethnicity",4],2)},
\Sexpr{round(result2.muslim["Ward.Density.Own.Ethnicity",4],2)}
and
\Sexpr{round(result3.muslim["Ward.Density.Own.Ethnicity",4],2)}
for Muslims and
\Sexpr{round(result1.non.muslim["Ward.Density.Own.Ethnicity",4],2)},
\Sexpr{round(result2.non.muslim["Ward.Density.Own.Ethnicity",4],2)}
and
\Sexpr{round(result3.non.muslim["Ward.Density.Own.Ethnicity",4],2)}
for non-Muslims contradicting the \citet{Bisin08} results.
```

Inspecting the results presented [in](#) Tables `\ref{ta:regressions}` and `\ref{ta:regressions2}`, 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 `\citet{Bisin08}`. 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 \citet{Bisin08}).

```
\section{Concluding remarks}
\label{Discussion}
```

The \citet{Bisin08} 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 (\input{crossreference2.tex}\$\%) implies that the remaining sample is most likely not representative.

```
%</paper|techreport>
```

```
%<*present>
```

```
\begin{slide}{Conclusions}
```

```
\vspace{\stretch{1}}
```

```
\begin{itemize}
```

```
\item The \citet{Bisin08} paper rests on fragile grounds.
```

```
\item 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.
```

```
\item 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.
```

```
\item However, we hesitate to draw inference from
these results since the great loss of observations
(\input{crossreference2.tex}$\%) implies that the
remaining sample is most likely not representative.
```

```
\end{itemize}
```

```
\vspace{\stretch{1}}
```

```
\end{slide}
```

```
%</present>
```

```
%<*techreport>
```

```
\section{Production notes}
```

```
\label{sec:prodnotes}
```

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 `\pkg{Sweave}` by [\citet{Leisch02}](#) [in](#) combination with the `\LaTeX` family of programs using the packages `\code{inputenc}`, `\code{fontenc}`, `\code{natbib}`, `\code{Sweave}`, `\code{fancyvrb}`, `\code{color}`, `\code{url}`, `\code{hyperref}` and `\code{multirow}`.

All **code** (`\proglang{R}` code, `\LaTeX` code and Bib`\TeX` data base code) used to do the econometric estimations, to produce this technical documentation including all tables and to produce the companion paper [\citet{Araietal08a}](#) is contained [in](#) the file `\code{araietal_source.Rnw}`.

The estimations and the documentation can be reproduced [in](#) by following the instructions at the top of the file `\url{http://people.su.se/~lundh/fragile_grounds/araietal_source.Rnw}`

Our results were obtained on a `\Sexpr{sessionInfo()}$R.version$platform` platform using `\proglang \Sexpr{sessionInfo()}$R.version$version.string` `\citep{Rcore}` with packages `\pkg{\Sexpr{sessionInfo()}$otherPkgs$lmtest$Package}` `\Sexpr{sessionInfo()}$otherPkgs$lmtest$Version` (`\Sexpr{sessionInfo()}$otherPkgs$lmtest$Date`), `\pkg{\Sexpr{sessionInfo()}$otherPkgs$sandwich$Package}` `\Sexpr{sessionInfo()}$otherPkgs$sandwich$Version` (`\Sexpr{sessionInfo()}$otherPkgs$sandwich$Date`), `\pkg{\Sexpr{sessionInfo()}$otherPkgs$zoo$Package}` `\Sexpr{sessionInfo()}$otherPkgs$zoo$Version` (`\Sexpr{sessionInfo()}$otherPkgs$zoo$Date`), `\pkg{\Sexpr{sessionInfo()}$otherPkgs$foreign$Package}` `\Sexpr{sessionInfo()}$otherPkgs$foreign$Version` (`\Sexpr{sessionInfo()}$otherPkgs$foreign$Date`) and `\pkg{\Sexpr{sessionInfo()}$otherPkgs$xtable$Package}` `\Sexpr{sessionInfo()}$otherPkgs$xtable$Version` (`\Sexpr{sessionInfo()}$otherPkgs$xtable$Date`).

`%</techreport>`

`%<*paper>`
`\end{spacing}`
`%</paper>`

`%<*paper|techreport>`

`\bibliography{araietal}`

`\bibliographystyle{plainnat}`

`\newpage`

%%
 %% Comment: We use a combination of `xtable()` and `print()`
 %% functionality [in](#) order to have R produce the desired
 %% tables directly i LaTeX-code.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
<<Loading xtable,echo=FALSE>>=
```

```
library(xtable)
```

```
@
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
%%
```

```
%% Code for Table 1: Religious affiliation.
```

```
%%
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
<<Table 1,results=tex,keep.source=FALSE,echo=FALSE>>=
```

```
# Defining variables to use in the table
```

```
OnonMuslim      <- table(FNSEM$s7!="muslim")[2]
```

```
names(OnonMuslim) <- "nonMuslim"
```

```
SnonMuslim      <- table(U$Religion!="muslim")[2]
```

```
names(SnonMuslim) <- "nonMuslim"
```

```
# Defining the table header
```

```
Header <- paste("Religious affiliation",
```

```
"&\\multicolumn{1}{c}{(1)}",
```

```
"&\\multicolumn{1}{c}{(2)}",
```

```
"&\\multicolumn{1}{c}{(3)}",
```

```
"&\\multicolumn{1}{c}{(4)}",
```

```
"&\\multicolumn{1}{c}{(5)}",
```

```
"&\\multicolumn{1}{c}{(6)}",
```

```
"\\hline",
```

```
"&\\multicolumn{2}{c}{Before}",
```

```
"&\\multicolumn{2}{c}{After}",
```

```
"&\\multicolumn{2}{c}{Bisin et al.}",
```

```
"\\",
```

```
"&\\multicolumn{2}{c}{n=",length(U.Original$s7),"$}",
```

```
"&\\multicolumn{2}{c}{n=",length(U$s7),"$}",
```

```
"&\\multicolumn{2}{c}{n=5963$}",
```

```
"\\cline{2-7}",
```

```
"&\\multicolumn{1}{r}{\\#}&\\multicolumn{1}{r}{\\%}",
```

```
"&\\multicolumn{1}{r}{\\#}&\\multicolumn{1}{r}{\\%}",
```

```
"&\\multicolumn{1}{r}{\\#}&\\multicolumn{1}{r}{\\%}\\",
```

```
sep="")
```

```
# Defining table footer
```

```
Footer <- paste("\\hline All non-Muslims",
```

```
"&",OnonMuslim,"&",
```

```
round(100*OnonMuslim/table(FNSEM$s7=="na")[[1]],2),
```

```
"&",SnonMuslim,"&",
```

```
round(100*SnonMuslim/length(U$Religion),2),
```

```
"&3594&",round(100*3594/(3594+2369),2),
```

```
"\\hline",
```

```
"\\multicolumn{7}{l}{\\multirow{2}{10.5cm}}",
```

```
"{\\footnotesize 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.}\\",
```

```
sep="")
```

```
addtorow <- list()
```

```
addtorow$pos <- list()
```

```
addtorow$pos[[1]] <- 0
```

```
addtorow$pos[[2]] <- 13
```

```
addtorow$command <- c(Header,Footer)
```

```
# Defining the data to use in the table
```

```

DiffSample <- cbind(
  summary(FNSEM$s7),
  100*summary(U.Original$s7)/length(U.Original$s7),
  c(summary(U$s7),
    length(U$s7)-table(is.na(U$s7))[[1]]),
  c(100*summary(U$s7)/length(U$s7),
    100*(length(U$s7)-table(is.na(U$s7))[[1]])),
  matrix(list(c(),c()),round(2369,0),c(),c(),c(),
    c(),c(),c(),c(),c(),c(),c(),c()),
  matrix(list(c(),c(),39.73,c(),c(),c(),c(),c(),c(),
    c(),c(),c(),c(),c(),c()))
)
# Taking dimnames from original data
dimnames(DiffSample)[[1]] <- names(summary(U.Original$s7))
# Generate the table
TabDiffSample <- print(xtable(DiffSample,align=c("lrrrrrr"),
  caption=c("Religious affiliation (absolute (\\#) and
  relative (\\%) numbers), before (columns 1 and 2) and
  after (columns 3 and 4) removal of \\code{NA} compared
  with \\citet{Bisin08} (columns 5 and 6)."),
  label="ta:difffsample",
  digits=c(0,0,2,0,2,0,2)),
  table.placement = "h",
  caption.placement="top",
  include.colnames=FALSE,
  add.to.row=addtorow,
  hline.after=c(-1,0,11))
@

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%
%% Code for Table 2-3: Weighted and weighted means.
%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

<<Table 2-3,results=tex,keep.source=FALSE,echo=FALSE>>=
# Generate sample with all NA removed
SampleU <- subset(U,select=c(
  Importance.of.Religion,
  Attitude.Towards.Inter.Marriage,
  Importance.of.Racial.Composition.in.Schools,
  Age.at.Arrival,
  Age,
  Female,
  Born.in.the.UK,
  Arranged.Marriage,
  Discrimination,
  Children,
  Years.Since.Arrival,
  No.British.Education,
  British.Basic.Education,
  British.Higher.Education,
  Foreign.Education,
  Employee,
  Manager,
  Self.Employed,
  OUT.OF.LABOUR.FORCE,
  Unemployed,
  No.Parents,

```

```

Parents.Physical.Contacts,
Parents.Telephone.Calls,
Parents.Letters,
English.Spoken.at.Home.with.Older,
DO.NOT.SPEAK.WITH.OLDER,
English.Spoken.at.Home.with.Younger,
DO.NOT.SPEAK.WITH.YOUNGER,
English.Spoken.at.Work,
DO.NOT.SPEAK.AT.WORK,
English.Spoken.With.Friends,
DO.NOT.SPEAK.WITH.FRIENDS,
Household.Income,
Ward.Density.Own.Ethnicity,
Ward.Unemployment.Rate,
Discrimination.Own.Ethnicity
))
# Generate sample with no NA removed
SampleU.Original <- subset(U.Original,select=c(
  Importance.of.Religion,
  Attitude.Towards.Inter.Marriage,
  Importance.of.Racial.Composition.in.Schools,
  Age.at.Arrival,
  Age,
  Female,
  Born.in.the.UK,
  Arranged.Marriage,
  Discrimination,
  Children,
  Years.Since.Arrival,
  No.British.Education,
  British.Basic.Education,
  British.Higher.Education,
  Foreign.Education,
  Employee,
  Manager,
  Self.Employed,
  OUT.OF.LABOUR.FORCE,
  Unemployed,
  No.Parents,
  Parents.Physical.Contacts,
  Parents.Telephone.Calls,
  Parents.Letters,
  English.Spoken.at.Home.with.Older,
  DO.NOT.SPEAK.WITH.OLDER,
  English.Spoken.at.Home.with.Younger,
  DO.NOT.SPEAK.WITH.YOUNGER,
  English.Spoken.at.Work,
  DO.NOT.SPEAK.AT.WORK,
  English.Spoken.With.Friends,
  DO.NOT.SPEAK.WITH.FRIENDS,
  Household.Income,
  Ward.Density.Own.Ethnicity,
  Ward.Unemployment.Rate,
  Discrimination.Own.Ethnicity
))
# Set Age.at.Arrival and Years.Since.Arrival
# for those Born.in.the.UK to NA. This is only for the
# sample statistics.

```

```

is.na(SampleU.Original$Age.at.Arrival)      <-
      SampleU.Original$Born.in.the.UK==1
is.na(SampleU.Original$Years.Since.Arrival) <-
      SampleU.Original$Born.in.the.UK==1

# A function to compute weighted mean
# for a variable in a data set (x), with
# weights (w) and column (n).
WMEAN <- function(x,W,n) {
  weighted.mean(x[!is.na(x[,n]),n], W[!is.na(x[, n])])}

# Weights to use in computing weighted means for Muslims
Muslim.weight <-
  U$weightis[U$Religion=="muslim"]
Orig.Muslim.weight <-
  U.Original$weightis[U.Original$Religion=="muslim"]
Non.Muslim.weight <-
  U$weightis[U$Religion=="non-muslim"]
Orig.Non.Muslim.weight <-
  U.Original$weightis[U.Original$Religion=="non-muslim"]

# Create a matrix to gather means
# Columns 1:5 for Muslims:
# 1:2 weighted means, 3:4 unweighted, 5 Bisin et al.
# Columns 6:10 for non-Muslims:
# 6:7 weighted means, 8:9 unweighted, 10 Bisin et al.

Means.Table <- matrix(,ncol(SampleU),10)
rownames(Means.Table) <- colnames(SampleU)

# Compute weighted means in columns 1:36
for (i in 1:ncol(SampleU.Original)) {Means.Table[i,1]<-
  WMEAN(SampleU.Original[U.Original$Religion=="muslim",],
        Orig.Muslim.weight, i)}

# Compute weighted means in columns 1:36
for (i in 1:ncol(SampleU)) {Means.Table[i,2]<-
  WMEAN(SampleU[U$Religion=="muslim",] , Muslim.weight, i)}

# Unweighted Means for Muslims
Means.Table[,3] <-
  mean(SampleU.Original[U.Original$Religion=="muslim",],
      na.rm=TRUE)
Means.Table[,4] <-
  mean(SampleU[U$Religion=="muslim",], na.rm=TRUE)

# Bisin et al. Muslims
Means.Table[,5] <-
c(0.79,0.70,0.65,39.18, NA, 0.47,0.21,0.22,0.17,2.17,
  26.43,0.81,0.06,0.08,0.25,0.38,0.02,0.09,NA,0.19,
  0.34,3.05,3.38,0.67,0.03,NA,0.20,NA,0.19,NA,0.22,

```

```
NA,200.74,0.15,16.57,0.21)
```

```
# Compute weighted means SampleU.Original
for (i in 1:ncol(SampleU.Original)) {Means.Table[i,6]<-
  WMEAN(SampleU.Original[U.Original$Religion=="non-muslim",],
  Orig.Non.Muslim.weight, i)}
```

```
# Compute weighted means SampleU
for (i in 1:ncol(SampleU)) {Means.Table[i,7]<-
  WMEAN(SampleU[U$Religion=="non-muslim",],
  Non.Muslim.weight, i)}
```

```
# Unweighted Means for non-Muslims
Means.Table[,8] <-
  mean(SampleU.Original[U.Original$Religion=="non-muslim",],
  na.rm=TRUE)
Means.Table[,9] <-
  mean(SampleU[U$Religion=="non-muslim",],na.rm=TRUE)
```

```
# Bisin et al. non-Muslims
Means.Table[,10] <-
c(0.42,0.37,0.33,42.57,NA,0.48,0.28,0.12,0.19,1.68,
  27.08,0.52,0.13,0.16,0.29,0.59,0.04,0.14,NA,0.08,
  0.32,3.87,4.74,0.37,0.08,NA,0.25,NA,0.27,NA,0.27,
  NA,330.26,0.11,12.60,0.18)
```

```
# Replace "." with " " in variable names to appear in the table
rownames(Means.Table) <- gsub("\\.", " ", rownames(Means.Table))
```

```
# Creating the table i LaTeX code
print(xtable(Means.Table[1:17,],align=c("lrrrrrrrrrr"),
  caption="Weighted and Unweighted Means for Muslims and
  non--Muslims before and after removal of \\code{NA}
  compared with \\citet{Bisin08}.",
  label="tab:descriptive"),
  table.placement = "p",
  floating.environment = "sidewaystable",
  caption.placement="top",
  add.to.row=list(pos=list(0),
  command=paste(
  "&\\multicolumn{5}{c}{Muslim}",
  "&\\multicolumn{5}{c}{Non-Muslim}\\\\\\\\\\\\\\\\hline",
  "&\\multicolumn{2}{c}{W e i g h t e d}",
  "&\\multicolumn{3}{c}{U n w e i g h t e d}",
  "&\\multicolumn{2}{c}{W e i g h t e d}",
  "&\\multicolumn{3}{c}{U n w e i g h t e d}\\\\\\\\\\\\",
  "&\\multicolumn{1}{c}{Before}",
  "&\\multicolumn{1}{c}{After}",
  "&\\multicolumn{1}{c}{Before}",
  "&\\multicolumn{1}{c}{After}",
  "&\\multicolumn{1}{c}{Bisin}"),
```

```

"&\multicolumn{1}{c}{Before}",
"&\multicolumn{1}{c}{After}",
"&\multicolumn{1}{c}{Before}",
"&\multicolumn{1}{c}{After}",
"&\multicolumn{1}{c}{Bisin}\\"",
"&&&&Set al.&&&&Set al.\\"hline",
sep=""),
include.colnames=FALSE,
hline.after=c(-1,0,NULL))
# Table 3: Continuation of Table 2
print(xtable(Means.Table[18:nrow(Means.Table)],
  align=c("lrrrrrrrrrr"),
  caption="Table 2 continued. Weighted and Unweighted
Means for Muslims and non--Muslims before and after
removal of \\code{NA} compared with \\citet{Bisin08}.",
  label="tab:descriptive2"),
  table.placement = "p",
  floating.environment = "sidewaystable",
  caption.placement="top",
  add.to.row=list(pos=list(0),
  command=paste(
"&\multicolumn{5}{c}{Muslim}",
"&\multicolumn{5}{c}{Non-Muslim}\\"hline",
"&\multicolumn{2}{c}{W e i g h t e d}",
"&\multicolumn{3}{c}{U n w e i g h t e d}",
"&\multicolumn{2}{c}{W e i g h t e d}",
"&\multicolumn{3}{c}{U n w e i g h t e d}\\"",
"&\multicolumn{1}{c}{Before}",
"&\multicolumn{1}{c}{After}",
"&\multicolumn{1}{c}{Before}",
"&\multicolumn{1}{c}{After}",
"&\multicolumn{1}{c}{Bisin}",
"&\multicolumn{1}{c}{Before}",
"&\multicolumn{1}{c}{After}",
"&\multicolumn{1}{c}{Before}",
"&\multicolumn{1}{c}{After}",
"&\multicolumn{1}{c}{Bisin}\\"",
"&&&&Set al.&&&&Set al.\\"hline",
sep="")),
include.colnames=FALSE,
hline.after=c(-1,0,NULL))
@
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%
%% Code for Table 4-5: Regression results.
%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
<<Table 4 prepare, echo=FALSE>>=
MakeTable <- function(x, n=2) {
  # take coef() objects and makes table
  Stars <- ifelse(x[,4]<0.05, paste("*"), paste(""))
  names(Stars) <- NULL
  Estimates <- format(round(x[,1],n))
  Std.errors <- format(round(x[,2],n))
  x <- cbind(Estimates,Std.errors, Stars)
  Table <- matrix(,2*nrow(x),1)
  Table[seq(1,2*nrow(x)-1,2),1] <-
  paste(x[,1],x[,3], sep="")

```

```

Table[seq(2,2*nrow(x),2 ),1] <-
paste("(",x[,2],")",sep="")
Xname <- unlist(strsplit(unlist(strsplit(rownames(x),
"TRUE")), "yes") )
Xname <- gsub("\\.", " ", Xname)
Xnames <- matrix(,2*nrow(x),1)
Xnames[seq(1,2*nrow(x)-1,2),1] <- Xname
Xnames[seq(2,2*nrow(x) ,2),1] <- ""
strsplit(Xnames, "TRUE")
strsplit(Xnames, "yes")
rownames(Table) <- Xnames
Table
}

```

```

Table.Reg <-
cbind(
rbind(MakeTable(result1.muslim [1:28,]),
format(round(result1.muslim[39,1],2), nsmall=2)),
rbind(MakeTable(result1.non.muslim[1:28,]),
round(result1.non.muslim[40,1],2)),
rbind(MakeTable(result2.muslim [1:28,]),
round(result2.muslim[39,1],2)),
rbind(MakeTable(result2.non.muslim[1:28,]),
round(result2.non.muslim[40,1],2)),
rbind(MakeTable(result3.muslim [1:28,]),
round(result3.muslim[39,1],2)),
rbind(MakeTable(result3.non.muslim[1:28,]),
round(result3.non.muslim[40,1],2))
)

```

```

@
<<Print Table 4, results=tex, echo=FALSE>>=

```

```

rownames(Table.Reg)[nrow(Table.Reg)] <- "Adjusted R-square"
varnames <- rownames(Table.Reg)
rownames(Table.Reg) <- NULL
Table.Reg.dat <- data.frame(varnames, Table.Reg)
print(xtable(Table.Reg.dat[1:28,],align=c("llccccc"),
caption=paste("Regression Results for Muslims",
nrow(Umuslim),"and non-Muslims", nrow(UNmuslim),
"to be compared with Table 2 in \\citet{Bisin08}. ",
"Heteroskedasticity corrected (HCl) Standard Errors ",
"are in parentheses. P-values $< 0.05$ are ",
"marked with *.",sep=" "),
label="ta:regressions"),
table.placement = "p",
floating.environment = "sidewaystable",
caption.placement="top",
include.rownames=FALSE,
include.colnames=FALSE,
add.to.row=list(pos=list(0),
command=paste(
"&\\multicolumn{2}{c}{Importance of}",
"&\\multicolumn{2}{c}{Inter Ethnic }",
"&\\multicolumn{2}{c}{Ethnic Composition}\\\\\\\\",
"&\\multicolumn{2}{c}{Religion}",
"&\\multicolumn{2}{c}{Marriage}",

```

```

"&\multicolumn{2}{c}{of Schools}",
"\\\\\\cline{2-7}\\\\\\",
"&\multicolumn{1}{c}{Muslims}",
"&\multicolumn{1}{c}{non--Muslims}",
"&\multicolumn{1}{c}{Muslims}",
"&\multicolumn{1}{c}{non--Muslims}",
"&\multicolumn{1}{c}{Muslims}",
"&\multicolumn{1}{c}{non--Muslims}\\\\\\hline",
sep="")
, hline.after=c(-1,0,NULL))
# Table 5: Continuation of Table 4
Footer <- paste("\\hline",
  "\\multicolumn{7}{l}{\\multirow{3}{20cm}",
  "{\\footnotesize NOTE: All estimated models include ",
  "7 UK-region dummies, and ",
  "the variables \\code{DO.NOT.SPEAK.WITH.OLDER}, ",
  "\\code{DO.NOT.SPEAK.AT.WORK}, and ",
  "\\code{DO.NOT.SPEAK.WITH.FRIENDS}. It turned out ",
  "that the variable \\code{DO.NOT.SPEAK.WITH.YOUNGER} ",
  "is TRUE for few observations and cannot be ",
  "included in the model.}\\",
  sep="")

Header <- paste(
  "&\multicolumn{2}{c}{Importance of}",
  "&\multicolumn{2}{c}{Inter Ethnic }",
  "&\multicolumn{2}{c}{Ethnic Composition}\\\\\\",
  "&\multicolumn{2}{c}{Religion}",
  "&\multicolumn{2}{c}{Marriage}",
  "&\multicolumn{2}{c}{of Schools}",
  "\\\\\\cline{2-7}\\\\\\",
  "&\multicolumn{1}{c}{Muslims}",
  "&\multicolumn{1}{c}{non--Muslims}",
  "&\multicolumn{1}{c}{Muslims}",
  "&\multicolumn{1}{c}{non--Muslims}",
  "&\multicolumn{1}{c}{Muslims}",
  "&\multicolumn{1}{c}{non--Muslims}\\\\\\hline",
  sep="")

print(xtable(Table.Reg.dat[29:57,],align=c("llcccccc"),
  caption=paste("Table 4 continued. Regression Results ",
  "for Muslims", nrow(Umuslim)," and non-Muslims",
  nrow(UNmuslim), "to be compared with Table 2 in ",
  "\\citet{Bisin08}. Heteroskedasticity corrected ",
  "(HC1) Standard Errors are in parentheses. P-values ",
  "$< 0.05$ are marked with *.", sep=" "),
  label="ta:regressions2"),
  table.placement = "p",
  floating.environment = "sidewaystable",
  caption.placement="top",
  include.rownames=FALSE,
  include.colnames=FALSE,
  add.to.row=list(pos=list(0,29),
  command=c(Header,Footer))
  ,hline.after=c(-1,0,NULL))

```

@

```
%</paper|techreport>
```

```
%<*paper|techreport|present>
```

```
\end{document}
```

```
%</paper|techreport|present>
```

Here we create the statistics that are cross--references to in the abstract and the introduction. It is the relative number of observations that remains after missing values are removed and its complement:

```
<<Cross-references,echo=FALSE,keep.source=FALSE>>=
RelativeSampleSize <- round(100*nrow(U)/nrow(FNSEM),
  digits=0)
BisinRelativeSampleSize <- round(100*5963/nrow(FNSEM),
  digits=0)
write(BisinRelativeSampleSize,file="crossreference0.tex")
write(100-RelativeSampleSize,file="crossreference2.tex")
write(nrow(U.Original),file="crossreference3.tex")
write(nrow(U),file="crossreference4.tex")
```

```
@
```

The following code writes the DOCSTRIP commands to the file araietal.ins:

```
<<LaTeX,echo=FALSE>>=
write(paste(
  "\\input docstrip.tex",
  "\\keepsilent",
  "\\askforoverwritefalse",
  "\\nopreamble",
  "\\nopostamble",
  "\\generate{",
  "\\file{araietal_paper_bis.tex}{",
  "  \\from{araietal_source_bis.tex}{paper}}",
  "\\file{araietal_techreport_bis.tex}{",
  "  \\from{araietal_source_bis.tex}{techreport}}",
  "\\file{araietal_present_bis.tex}{",
  "  \\from{araietal_source_bis.tex}{present}}",
  "}",
  "\\endbatchfile",
  "\\n",sep="\\n"),file="araietal_bis.ins")
```

```
@
```

```
<<BibTeX,echo=FALSE>>=
write(paste(
"@Manual{FNSEM,",
"author      = {FNSEM}," ,
"title       = {P1312 Fourth National Study of Ethnic Minorities. {P}roject
instructions}," ,
"note        = {Project 3685}," ,
"organization = {Social and Community Planning Research}, ",
"address     = {London}," ,
"year        = {1993}," ,
"url         = {http://www.data-archive.ac.uk/doc/3685/mrdoc/pdf/a3685uab.pdf}," ,
"}",
"\\n",
"@Manual{FNSEM93b,",
```

```

"author      = {FNSEM},"",
"title       = {UK Data Archive Data Dictionary}, ",
"note       = {An RTF file called
\\url{UKDA-3685-tab/mrdoc/allissue/3685_UKDA_Data_Dictionary.rtf} available when the
FNSEM data set is downloaded from the UK Data Archive}, ",
"organization = {Social andCommunity Planning Research}, ",
"address     = {London},"",
"year       = {1993}",
"}",
"\n",
"@Article{Bisin08,"",
"author      = {Alberto Bisin and Eleonora Patacchini and Thierry Verdier and Yves
Zenou},"",
"title       = {Are Muslim Immigrants Different in Terms of Cultural Integration?},"",
"journal     = {Journal of the European Economic Association},"",
"pages       = {445--456},"",
"volume     = {6},"",
"year       = {2008}",
"}",
"\n",
"@Manual{Rcore,"",
"title       = {R: A Language and Environment for Statistical Computing},"",
"author      = {{R Development Core Team}},"",
"organization = {R Foundation for Statistical Computing},"",
"address     = {Vienna, Austria},"",
"year       = {2008},"",
"note       = {{ISBN} 3-900051-07-0},"",
"url        = {http://www.R-project.org}",
"}",
"\n",
"@Unpublished{Koenker07,"",
"author      = {Koenker, Roger and Zeileis, Achim},"",
"year       = {2007},"",
"title       = {Reproducible Econometric Research (A Critical Review of the State of the
Art)},"",
"note       = {Report 60, Department of Statistics and Mathematics,
Wirtschaftsuniversit t Wien, Research Report Series},"",
"url        =
{http://epub.wu-wien.ac.at/dyn/virlib/wp/eng/mediate/epub-wu-01_c75.pdf?ID=epub-wu-01_c75}"
,
"}",
"\n",
"@Unpublished{Araietal08a,"",
"author      = {Arai, Mahmood and Karlsson, Jonas and Lundholm, Michael},"",
"year       = {2009},"",
"title       = {On Fragile Grounds: A replication of \\emph{Are Muslim immigrants
different in terms of cultural
integration?}},"",
"note       = {Under publication in the Journal of the European Economic Association}",
"}",
"\n",
"@Unpublished{Araietal08b,"",
"author      = {Arai, Mahmood and Karlsson, Jonas and Lundholm, Michael},"",
"year       = {2009},"",
"title       = {On Fragile Grounds: {A} replication of \\emph{Are Muslim immigrants
different in terms of cultural integration? {T}echnical documentation}},"",
"note       = {Unpublished manuscript, Department of Economics and SULCIS, Stockholm
University},"",
"url        = {http://people.su.se/~lundh/fragile_grounds/},

```

```

"}",
"\n",
"@Unpublished{Leisch02,",
"author      = {Friedrich Leisch},"",
"title       = {Sweave User Manual},"",
"year       = {2002},"",
"url        = {http://www.ci.tuwien.ac.at/~leisch/Sweave/},"",
"}",
"\n",
"@Manual{Berthoud,",
"title       = {Fourth National Survey of Ethnic Minorities, 1993-1994 [computer
file]},"",
"author     = {Berthoud, R.G. and Modood, T. and Smith, P. and Prior, G.},"",
"publisher  = {UK Data Archive [distributor]},"",
"address    = {Colchester, Essex: UK},"",
"year       = {1997},"",
"note      = {SN: 3685},"",
"url        =
{http://www.data-archive.ac.uk/doc/3685%5Cmrdoc%5CUKDA%5CUKDA_Study_3685_Information.htm}
",
"}",
sep="\n"),file="araietal.bib")
@

```

```

<<Extracting R code,results=hide>>=
Stangle("araietal_source_bis.Rnw")
@

```

```

<<Script to make LaTeX files,echo=FALSE,silent=TRUE,results=hide>>=
write(paste(
"system(\"latex araietal_bis.ins\")",
"system(\"pdflatex araietal_techreport_bis.tex\")",
"system(\"bibtex araietal_techreport_bis\")",
"system(\"pdflatex araietal_techreport_bis.tex\")",
"system(\"pdflatex araietal_techreport_bis.tex\")",
"system(\"latex araietal_paper_bis.tex\")",
"system(\"bibtex araietal_paper_bis\")",
"system(\"latex araietal_paper_bis.tex\")",
"system(\"latex araietal_paper_bis.tex\")",
"system(\"dvips araietal_paper_bis.dvi\")",
"system(\"ps2pdf araietal_paper_bis.ps\")",
"system(\"latex araietal_present_bis.tex\")",
"system(\"dvips -o araietal_present1.ps -pp 1:14 araietal_present_bis.dvi\")",
"system(\"dvips -o araietal_present2.ps -pp 11:15 -x 1095 araietal_paper_bis.dvi\")",
"system(\"dvips -o araietal_present3.ps -p 20 araietal_present_bis.dvi\")",
"system(\"gs -q -r2400 -dNOPAUSE -dBATCH -sDEVICE=pdfwrite
-sOutputFile=araietal_present_bis.pdf -f araietal_present1.ps araietal_present2.ps
araietal_present3.ps\")",
"system(\"pdflatex araietal_paper_bis.tex\")",
sep="\n"),file="script.R")
@

```