

```
1 %<*paper|techreport|present>
2 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
3 %%
4 %% TITLE:    On Fragile Grounds:
5 %%           A replication of "Are Muslim immigrants
6 %%           different in terms of cultural integration?"
7 %% AUTHORS:  Mahmood Arai, Jonas Karlsson and Michael
8 %%           Lundholm
9 %% CONTACT:  mahmood.arai@ne.su.se
10 %%          jonas.karlsson@sofi.su.se
11 %%          michael.lundholm@ne.su.se
12 %%
13 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
14 %%
15 %% This batch file can be used to generate the relevant
16 %% files performing the estimations, presentations and
17 %% documentations of the project by using the R package
18 %% Sweave, the LaTeX DOCSTRIP utility and the standard LaTeX
19 %% family of software using the UN*X commands
20 %%
21 %% echo 'Sweave("araietal_source.Rnw")' | R --vanilla --quiet
22 %% echo 'source("script.R")' | R --vanilla --quiet
23 %%
24 %% or from within R using
25 %%
26 %% Sweave("araietal_source.Rnw")
27 %% source("script.R")
28 %%
29 %% The file araietal_source.Rnw contains all source code
30 %% and run through Sweave it generates the files
31 %% araietal_source.tex, script.R, araietal_source.R,
32 %% araietal.ins and araietal.bib. When the file script.R is
33 %% run first LaTeX's DOCSTRIP utility is run via
34 %% araietal.ins and the files araietal_paper.tex,
35 %% araietal_techreport.tex and araietal_present.tex are
36 %% generated. The script then continues to process these
37 %% files with the LaTeX family of programs, using the
38 %% BibTeX bibliography database file araietal.bib, to
39 %% generate the corresponding PDF-files for paper,
40 %% technical documentation and presentation.
41 %%
42 %% Accordingly the file contains three levels of markup
43 %% 1. noweb markup to define the code chunks evaluated
44 %%    when "araietal_source.Rnw" is run through R via Sweave.
45 %% 2. DOCSTRIP markup to define the conditional LaTeX code
46 %%    to be shipped to the different generated LaTeX files.
47 %%    Conditioning is made on the tags "paper" (which
48 %%    generates araietal_paper.tex and contains our
49 %%    research result in an article, "techreport" (which
50 %%    generates araietal_techreport.tex and contains a
51 %%    complete technical documentation of our research) and
52 %%    "present" (which generates araietal_present.tex and
53 %%    contains a slide presentation of our research).
54 %%    when "latex araietal.ins" is run.
55 %% 3. Standard LaTeX markup to be considered when the
56 %%    generated LaTeX files are run with latex/pdflatex.
57 %%
58 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
59 %</paper|techreport|present>
60 %%
```



```

121 % Defining conditional title pages
122 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
123 %</paper|techreport|present>
124 %<*paper>
125 \title{\LARGE\bfseries On Fragile Grounds:\\
126         A replication of \emph{Are Muslim immigrants
127         different in terms of cultural integration?}}
128 %</paper>
129 %<*techreport>
130 \title{\LARGE\bfseries On Fragile Grounds:\\
131         A replication of \emph{Are Muslim immigrants
132         different in terms of cultural integration?
133         }\\Technical documentation}
134 %</techreport>
135 %<*present>
136 \title{On Fragile Grounds:\\
137         A replication of "Are Muslim immigrants
138         different in terms of cultural integration?"}
139 %</present>
140 %<*paper|techreport>
141 \author{Mahmood Arai,\footnote{Corresponding author.
142         Department of Economics and SULCIS, Stockholm
143         University, SE 106~91 Stockholm, Sweden,
144         \email{mahmood.arai@ne.su.se}.} { }Jonas
145         Karlsson\footnote{The Institute for Social Research and
146         SULCIS, Stockholm University,
147         \email{jonas.karlsson@sofi.su.se}.} { }and Michael
148         Lundholm\footnote{Department of Economics, Stockholm
149         University, \email{michael.lundholm@ne.su.se}.}}
150 %</paper|techreport>
151 %<*present>
152 \author{Mahmood Arai, Jonas Karlsson and Michael Lundholm}
153 %</present>
154 %<*paper|techreport|present>
155 %%
156 %%
157 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
158 %% Starting the LaTeX document and creating title pages
159 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
160 \begin{document}
161 %</paper|techreport|present>
162 %<*present>
163 \sffamily
164 %</present>
165 %<*paper|techreport|present>
166 \maketitle
167 %</paper|techreport|present>
168 %<*paper>
169 \begin{abstract}
170 This study is a replication of ``Are Muslim Immigrants
171 Different in terms of Cultural Integration?' by Alberto
172 Bisin, Eleonora Patacchini, Thierry Verdier and Yves Zenou,
173 published in Journal of European Economic Association, 6,
174 445-456, 2008.
175
176
177 \citet{Bisin08} report that they have 5963 observations in
178 their study.
179 Using their empirical setup, we can only identify
180 \input{crossreference3.tex}

```

```
181 relevant observations in the original data. After removing
182 missing values we are left with
183 \input{crossreference4.tex}
184 observations. We cannot replicate any of their results and
185 our estimations yield no support for their claims.
186
187
188 \end{abstract}
189 %</paper>
190 %<*techreport>
191 \begin{abstract}
192 This is a technical documentation of \citet{Araietal08a}
193 which replicates ``Are Muslim Immigrants Different in terms
194 of Cultural Integration?'' by Alberto Bisin, Eleonora
195 Patacchini, Thierry Verdier and Yves Zenou, published in
196 Journal of European Economic Association, 6, 445-456, 2008.
197
198
199
200 \citet{Bisin08} report that they have 5963 observations in
201 their study. Using their empirical setup,
202 we can only identify
203 \input{crossreference3.tex}
204 relevant observations in the original data. After removing
205 missing values we are left with
206 \input{crossreference4.tex}
207 observations. We cannot replicate any of their results and
208 our estimations yield no support for their
209 claims.
210
211
212 \end{abstract}
213 \newpage
214 \tableofcontents
215 \newpage
216 %</techreport>
217 %<*paper|techreport|present>
218 %%
219 %%
220 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
221 %% Document content starting
222 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
223 %</paper|techreport|present>
224
225
226 %<*present>
227 \begin{slide}{Introduction}
228 \vspace{\stretch{1}}
229 \begin{itemize}
230 \item This study is a replication of ``Are Muslim
231 Immigrants Different in terms of Cultural Integration?'' by
232 Alberto Bisin, Eleonora Patacchini, Thierry Verdier and Yves
233 Zenou, published in Journal of European Economic
234 Association, 6, 445-456, 2008.
235 \end{itemize}
236 \vspace{\stretch{1}}
237 \end{slide}
238
239 %</present>
240
```

```
241
242
243 %<*paper|techreport>
244 \section{Introduction}
245
246 This is a replication of the empirical
247 results reported in \citet{Bisin08}. They use British
248 data and analyse how Muslims and non-Muslims
249 differ in cultural integration measured as (i)
250 \emph{Importance of Religion}, (ii) \emph{Attitude
251 Towards Inter Marriage} and (iii) \emph{Importance of
252 Racial Composition in Schools}.\footnote{To facilitate
253 comparability we use the same labels on the variables as
254 \citet{Bisin08}.}
255 %</paper|techreport>
256
257 %<*present>
258 \begin{slide}{Their claims}
259 \vspace{\stretch{1}}
260 In the abstract of their paper they write:
261 \begin{center}
262 \begin{quote}
263 ``\dots Muslims integrate less and more slowly than
264 non-Muslims. \dots We also find no evidence that
265 segregated neighbourhoods breed intense religious and
266 cultural identities for ethnic minorities, especially for
267 Muslims.''
268
269 ``\dots On the contrary, \dots
270 intense identities in our data are more prominent
271 in relatively mixed neighbourhoods.''
272 (p. 446)
273 \end{quote}
274 \end{center}
275 \vspace{\stretch{1}}
276 \end{slide}
277 %</present>
278
279 %<*paper|techreport>
280 In the abstract of their paper they write:
281 \begin{quote}
282 ``\dots Muslims integrate less and more slowly than
283 non-Muslims. \dots We also find no evidence that
284 segregated neighbourhoods breed intense religious and
285 cultural identities for ethnic minorities, especially for
286 Muslims.'' \citep[p. 245]{Bisin08}
287 \end{quote}
288
289 We wanted to check the robustness
290 of their results when considering the ethnic and religious
291 heterogeneity within both groups, Muslims and non-Muslims.
292 Among other things, we were concerned about the measures of
293 cultural values used in the paper. These measures capture
294 ethnic and religious attributes in different degrees for
295 different groups. For example the variable \emph{Attitudes
296 towards Inter-Marriage} with the majority UK population
297 captures only inter-ethnic marriage for the Christian ethnic
298 minorities but both inter-ethnic and inter-religious
299 marriage for Muslims.
300 %</paper|techreport>
```

```
301
302 %<*present>
303 \begin{slide}[toc=]{Our concerns}
304 \vspace{\stretch{1}}
305 \begin{enumerate}\setcounter{enumi}{1}
306 \item The measures of cultural values capture ethnic and
307 religious attributes in different degrees for different
308 groups.
309 \end{enumerate}
310 \vspace{\stretch{1}}
311 \end{slide}
312 %</present>
313
314 %<*paper|techreport>
315 However, already an initial inspection of data disclosed
316 that the number of observations in \citet{Bisin08}
317 exceeded the total number of observations in the ethnic
318 minority sample. We communicated this
319 to the authors and they answered that there were
320 some coding errors. We have received revised codes and a
321 revised version of their specifications and tables. Their
322 revised codes yield fewer observations than the sample in
323 the published version, but still more than we can identify
324 in the relevant sample of the original data. As far as we
325 can see, a source of the large number of observations
326 in their revised codes is that dummy variable definitions
327 include observations with missing values in the reference
328 categories (defined as zeros). The underlying codes to the
329 published paper were, however, not made available and the
330 exact nature of the original errors are therefore unknown
331 to us.
332
333 \citet{Bisin08} report that they have 5963 observations in
334 their study, whereas the ethnic minority sample in
335 \citet{Berthoud} consists of \input{crossreference1.tex}
336 observations. Implementing their empirical setup, we can
337 only identify \input{crossreference3.tex} relevant
338 observations in the original data. After removing missing
339 values we are left with \input{crossreference4.tex}
340 observations. Using the remaining sample and running their
341 specifications, we find no results that support their
342 claims. Our replication therefore stopped here and we did
343 not perform any sensitivity analysis. The great loss of
344 observations implies that the remaining sample is most
345 likely not representative. Therefore, we hesitate to draw
346 inference from the regressions results.
347 %</paper|techreport>
348
349 %<*present>
350 \begin{slide}{Our results}
351 \vspace{\stretch{1}}
352
353 \begin{itemize}
354 \item In their paper the number of observations
355 is 5963 which is  $\input{crossreference0.tex}$  percent of
356 the number of observations in the relevant sample of
357 the original data ( $\input{crossreference3.tex}$ ).
358 \item Implementing their variable definitions we loose
359  $\input{crossreference2.tex}$  percent of the original
360 sample and have only  $\input{crossreference4.tex}$ 
```

```
361 observations.
362 \item Using the remaining sample and running their
363 specifications, we find no results that support their
364 claims.
365 \end{itemize}
366 \vspace{\stretch{1}}
367 \end{slide}
368
369 %</present>
370
371 %<*paper|techreport>
372
373 In this paper we only document the replication and report
374 and comment results using the variable definitions, the
375 variable names and the specifications used in
376 \citet{Bisin08}. We choose a procedure that makes it easy to
377 reproduce our results. Influenced by \cite{Koenker07}, we
378 use an integrated approach where data management,
379 estimations, and the text that rely on these computations
380 are all integrated in one single file. This strategy has the
381 advantage that it makes is easy to adjust the codes and
382 automatically generate a revised version of the paper.
383 %</paper|techreport>
384
385
386
387
388
389 %<*paper>
390 For details of our analysis of the data and implementation
391 of the variable definitions, see \citet{Araietal08b},
392 which is a technical companion to this paper.
393 %</paper>
394 %<*paper|techreport>
395 All data analysis is made in \proglang{R} \citep{Rcore}
396 and all code files related to this project can be found
397 on \url{http://people.su.se/~lundh/fragile_grounds/}.
398 %</paper|techreport>
399
400
401
402
403 %<*techreport>
404
405 In this technical documentation we present
406 our results in greater detail, but also all our working
407 procedures, variables definitions etc. In addition the
408 central part of our codes are included with typeset
409 comments. This is done as an attempt to implement Literate
410 Statistical Programming.
411
412 The remainder of the paper is organised as follows.
413 The data and variable definitions are described in Section
414 \ref{Data}. Regression results are presented in Section
415 \ref{Regression}. The paper is concluded in Section
416 \ref{Discussion}. Finally the production procedure is
417 described in Section \ref{sec:prodnotes}.
418 %</techreport>
419
420
```

```
421 %<*paper>
422
423 The remainder of the paper is organised as follows.
424 The data are described in Section \ref{Data}. Regression
425 results are presented in Section \ref{Regression} and
426 finally the paper is concluded in Section \ref{Discussion}.
427 %</paper>
428
429
430 %<*present>
431 \begin{slide}[toc=What we do]{What we wanted to do, what we
432 did and what we didn't}
433 \vspace{\stretch{1}}
434
435 \begin{itemize}
436 \item We wanted to:
437     \begin{itemize}
438     \item replicate their study,
439     \item check the robustness of their results
440     \end{itemize}
441 \item What we do
442     \begin{itemize}
443     \item document the replication and report
444     and comment results using their variable definitions
445     etc
446     \end{itemize}
447 \item What we didn't do:
448     \begin{itemize}
449     \item almost no sensitivity checking
450     \end{itemize}
451 \end{itemize}
452 \vspace{\stretch{1}}
453 \end{slide}
454
455
456 %</present>
457
458
459 %<*paper|techreport>
460 \section{Data and variable description}
461 \label{Data}
462 %</paper|techreport>
463
464 %<*present>
465 \begin{slide}{Data}
466 \vspace{\stretch{1}}
467 \begin{itemize}
468 \item Fourth National Survey of Ethnic Minorities 1993-1994
469 \item UK Data Archive (UKDA) via Athens.
470 \end{itemize}
471 \vspace{\stretch{1}}
472 \end{slide}
473 %</present>
474
475
476
477
478
479 %<*techreport>
480 \subsection{Data}
```

```

481 %</techreport>
482 %<*paper|techreport>
483 The data set is the \emph{Fourth National Survey of
484 Ethnic Minorities 1993-1994} (FNSEM); see
485 \citep{Berthoud}.\footnote{The data can be accessed
486 from the UK Data Archive (UKDA) via Athens. UK Data
487 Archive is found at \url{http://www.data-archive.ac.uk/}
488 and Athens at \url{http://www.athens.ac.uk/}.}
489 %</paper|techreport>
490 %<*techreport>
491 It's main objective were:
492 \begin{quote}
493 \begin{itemize}
494 \item ``to describe the social and economic conditions
495 of Britain's main ethnic minority groups, including
496 their health, and to compare these with the social
497 and economic conditions of the white majority
498 \item to assess changes over time through comparisons
499 with other work
500 \item to show how the position of ethnic minority
501 groups is related to the social and ethnic compositions
502 of the areas in which they live
503 \item to explore diversity among different ethnic
504 minority groups
505 \item to describe perceptions and experience of
506 racial discrimination and social harassment''
507 \end{itemize}
508 \citet{Berthoud}
509 \end{quote}
510
511 For our coding we have used \citet{FNSEM}, which contains
512 the project instructions, and \citet{FNSEM93b}, which
513 is the data description file included in the files
514 obtained when the entire data set is downloaded
515 from the UKDA.
516
517 In the following we present how the original data are used
518 to define the data set used in the estimations. We present
519 extracts of our \proglang{R} code \citep{Rcore} with
520 extensive comments and discussions. For details about
521 our working procedures and how we document the
522 research, see section \ref{sec:prodnotes} on page
523 \pageref{sec:prodnotes}. In the code chunks ``$>$''
524 denotes the \proglang{R} prompt and ``$+$''
525 continuation of the previous line.
526
527 \subsection{Reading data and selecting variables}
528 We load package \pkg{foreign} to read STATA data format.
529 Data is read from the unpacked Stata-version of the
530 data and ``\verb+_+' in variable names are converted
531 to ``$.$.``.
532
533 <<Reading data>>=
534 \bblibrary(foreign)
535 FNSEM <- data.frame(read.dta("3685.dta",
536   convert.underscore=TRUE))
537 @
538 <<Selecting variables, echo=FALSE>>=
539 FNSEM<-subset(FNSEM,select=c(
540   alan,ale,a3,a4n,area,ethnic,

```

```

541 f14a,f14b,f15a,f15bn:f15dn,f16a,f16b1n:f16b3n,
542 hh2a.s,hh2c.b,hh2c.c,hh2c.d,hh2c.e,hh2c.f,
543 hh2c.g,hh2c.h,hh2c.i,hh2c.j,hh2c.k,
544 hh2c.l,hh2c.m,hh5d.s,hh40,
545 j1b,j2,j3occ,j55a,j63a,
546 q1,q3,question,
547 s6,s7,s9,s12a,s12b,s12c,s24a,s34a,s34b,s39,
548 v1a,v1b,v1c,v1d,v9a,
549 weightis,wown,wunemp,
550 year,
551 ..q2a1,..q2a2,..q2a3,..q2a4,..q2a5,..q2a6,
552 ..q2a7,..q2a8,..q2a9,..q2a10,..q2a11,
553 ..q2a12,..q2a13,..q2a14,..q2a15,
554 ..q2a16,..q2a17,..q2a18,..q2a19,
555 ..q2a20,..q2a21,..q2a22,..q2a23,
556 ..q2a24,..q2a25,..q2a26,..q2a27,
557 ..q2a28,..q2a29,..q2a30,..q2a31,
558 ..q2a32,..q2a33,
559 ..s12f1,..s12f2, ..s12f3, ..s12f4, ..s12f5,
560 ..s12f6, ..s12f7, ..s12f8, ..s12f9,
561 ..s12f10,..s12f11,..s12f12,..s12f13,
562 ..s12f14,..s12f15,..s12f16,..s12f17,
563 ..s12f18,
564 ..s12g1,..s12g2, ..s12g3, ..s12g4, ..s12g5,
565 ..s12g6, ..s12g7, ..s12g8, ..s12g9,
566 ..s12g10,..s12g11,..s12g12,..s12g13,
567 ..s12g14,..s12g15,..s12g16,..s12g17,
568 ..s12g18,
569 ..s12h1,..s12h2, ..s12h3, ..s12h4, ..s12h5,
570 ..s12h6, ..s12h7, ..s12h8, ..s12h9,
571 ..s12h10,..s12h11,..s12h12,..s12h13,
572 ..s12h14,..s12h15,..s12h16,..s12h17,
573 ..s12h18,
574 ..s12i1,..s12i2, ..s12i3, ..s12i4, ..s12i5,
575 ..s12i6, ..s12i7, ..s12i8, ..s12i9,
576 ..s12i10,..s12i11,..s12i12,..s12i13,
577 ..s12i14,..s12i15,..s12i16,..s12i17,
578 ..s12i18
579 ))
580 @
581 After reading the data we select a subset of variables to
582 be used. This code is in \code{araietal\_source.Rnw} but not
583 shown here. It is also available in
584 \code{araietal\_source.R}.
585
586 %</techreport>
587
588 %<*present>
589 \begin{wideslide}{Sample}
590 \vspace{\stretch{1}}
591 \footnotesize{
592 <<Ethnic, echo=TRUE>>=
593 t(table(FNSEM$ethnic,exclude=c()))
594 @
595 }
596 \begin{itemize}
597 \item Remove Whites
598 \end{itemize}
599 \vspace{\stretch{1}}
600 \end{wideslide}

```

```

601
602
603 \begin{slide}[toc={Sample (cont.)}]
604 \vspace{\stretch{1}}
605 <<Religious, echo=TRUE>>=
606 table(FNSEM$s6,exclude=c())
607 @
608 \begin{itemize}
609 \item Remove non--religious (s6=2)
610 \end{itemize}
611 \vspace{\stretch{1}}
612 \end{slide}
613
614
615 \begin{slide}[toc={Sample (cont.)}]
616 \vspace{\stretch{1}}
617 <<Singles, echo=TRUE>>=
618 table(FNSEM$a1e,exclude=c())
619 @
620
621 \begin{itemize}
622 \item Remove singles (a1e=3)
623 \end{itemize}
624 \vspace{\stretch{1}}
625 \end{slide}
626
627 \begin{slide}[toc={Sample (cont.)}]
628 \vspace{\stretch{1}}
629 \begin{itemize}
630 \item One of three possible questionnaires (green (catageory
631 1), yellow (category 2)
632 and pink (category 3)) were used.
633 \item Questions involved in the study are in the green
634 questionnaire.
635 \end{itemize}
636 <<Questionnaire, echo=TRUE>>=
637 table(FNSEM$question,exclude=c())
638 @
639
640 \begin{itemize}
641 \item Remove pink and yellow (question=2,3)
642 \end{itemize}
643 \vspace{\stretch{1}}
644 \end{slide}
645
646 %</present>
647
648 %<*techreport>
649 \subsection{Defining the relevant (ethnic minority) sample}
650 \label{sec:definingsample}
651 The data consist of two samples, \emph{Ethnic Minorities}
652 and \emph{Whites}. We are only interested in the former
653 and remove all Whites. The variable \code{ethnic} indicates
654 ethnic group of the individual according to the British
655 standard and is used for this purpose.
656
657 One of the three measures of cultural integration in
658 \citet{Bisin08} is \emph{Importance of Religion}.
659 Whether a respondent has a religion or belongs to a
660 church is registered in question \code{s6}. Those who do

```

661 not have a religion or do not belong to a church are coded
662 `2`; we remove these observations from the sample since
663 they cannot be classified in a religious group.
664

665 A variable used in `\citet{Bisin08}` concerns the role of the
666 respondent and his or her parents about choosing the
667 respondent's husband/wife. Since this information is only
668 available for married and previously married persons, the
669 unmarried persons are removed from the sample.
670

671 Furthermore, respondents were faced with one out of three
672 questionnaires (green (category 1), yellow (category 2)
673 and pink (category 3)). The questions involved in the
674 study are only answered by individuals who were faced with
675 the green questionnaire. Therefore we keep only these in the
676 sample.
677

```
678 <<The ethnic minority sample 1,keep.source=FALSE>>=
679 # table(U$ethnic,exclude=c())
680 # table(U$s6,exclude=c())
681 # table(U$a1e,exclude=c())
682 @
683 <<The ethnic minority sample 2,>>=
684 FNSEM <- FNSEM[FNSEM$ethnic!="white" &
685           !is.na(FNSEM$ethnic),]
686 @
687 <<The ethnic minority sample 3,echo=FALSE>>=
688 write(nrow(FNSEM),file="crossreference1.tex")
689 @
690 <<The ethnic minority sample 4,>>=
691 FNSEM <- FNSEM[FNSEM$s6!=2,]
692 FNSEM <- FNSEM[FNSEM$a1e!=3,]
693 U <- FNSEM <- FNSEM[FNSEM$question==1,]
694 @
695 <<The ethnic minority sample 5,keep.source=FALSE,>>=
696 # table(U$ethnic,exclude=c())
697 # table(U$s6,exclude=c())
698 # table(U$a1e,exclude=c())
699 @
700
```

701 An issue where `\citet{Bisin08}` is imprecise is whether the
702 questions they address regard Muslims/non-Muslims or
703 Muslim/non-Muslim immigrants. Different sample selections
704 are possible here. The model specifications
705 in `\citet{Bisin08}` implies that White Muslims
706 are excluded and native ethnic minority Muslims are
707 included. This sample definition does not match
708 `\citet{Bisin08}` writing using the terms `\emph{Muslim and`
709 `non-Muslim immigrants}`, as the sample includes natives.
710

711 `\subsection{Recoding of missing values}`
712

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

```

721
722 <<Recoding of missing values>>=
723 is.na(U) <- U==-1
724 @
725
726 In addition to -1 several other codes ({na}, {7},
727 {8}, {9}, {98}, {99}, {997} and {999}) are occasionally used
728 in the data set to indicate various unknown categories. Some
729 of -1 and other unknown categories are non--available and
730 have to be deleted in estimations. This is done after we
731 have coded all our variables in section {sec:subset}. In
732 questions following a filter question {NA} may have to
733 be set in a category. Some of these other codes used to
734 denote unknowns are genuine
735 {NA}'s and has to be removed. Others will be included
736 in a category. This is done variable by variable below.
737
738 \subsection{Variable definitions}
739
740 Using the same variable names as in {Bisin08}, we
741 define the variables at the precision described by the
742 authors. We here give our interpretation of the variable
743 definitions in {Bisin08}.
744
745 \subsubsection{Religious affiliation}
746 Question {s6} asks whether the respondent belongs to a
747 church or has a religion. For those who answer yes,
748 question {s7} asks which that church or religion is.
749 We define two religious affiliations: {muslim},
750 (which are all who answered category 3 (muslim) on question
751 {s7}) and {non-muslim} (all who did not answer
752 category 3 (muslim) on question {s7}). All
753 non--religious, that is those who answered category 2 on
754 question {s6} are already removed from the sample.
755 Observations containing {na} are recoded to {NA}.
756 {citep[p. 112f]{FNSEM}}
757 <<Religious Affiliation 1,keep.source=FALSE>>=
758 # table(U$s7,exclude=c())
759 @
760 <<Religious Affiliation 2>>=
761 is.na(U$s7) <- U$s7=="na"
762 U$Religion <- ifelse(U$s7=="muslim","muslim","non-muslim")
763 @
764 <<Religious Affiliation 3,keep.source=FALSE>>=
765 # table(U$Religion,exclude=c())
766 @
767
768
769 \subsubsection{Importance of religion}
770 Question {s9} is about the importance of religion. To
771 grade the {Importance of Religion} respondents have
772 to choose between the following categories: 1. not at all
773 important, 2. not very important, 3. fairly important, or
774 4. very important. Following standard coding practice of
775 such questions, {1} and {2} should be one category and 3
776 and 4 another, but {Bisin08} choose to put {1}, {2}
777 and {3} in the same category.
778
779 Among those who have answered the question, very
780 few have chosen the alternative {1} or {2} in their answer,

```

```

781 implying very skewed distribution. Following
782 \citet{Bisin08}, code $4$ ('Very important') as answer on
783 question \code{s9} is coded TRUE; else FALSE. Codes $8$
784 ('Can't say') and $9$ are coded as \code{NA}. \citep[p.
785 112f]{FNSEM}
786
787 <<Importance of religion 1,keep.source=FALSE>>=
788 # table(U$s9,exclude=c())
789 @
790 <<Importance of religion 2>>=
791 is.na(U$s9) <- U$s9==8 | U$s9==9
792 U$Importance.of.Religion <- U$s9 == 4
793 @
794 <<Importance of religion 3,keep.source=FALSE>>=
795 # table(U$Importance.of.Religion,exclude=c())
796 @
797
798 %</techreport>
799
800 %<*present>
801
802 \begin{slide}[toc=Two important variables]{Religious
803 affiliation vs. importance of religion}
804 \vspace{\stretch{1}}
805
806 <<Rowtable function, echo=FALSE>>=
807 RowTable <- function(x1,x2)
      round(cbind(table(x1,x2,exclude=c()),table(x1,exclude=c()))/c(table(x1,
      exclude=c()),2)
808 @
809 \footnotesize{
810 <<Religion vs Importance of Religion, echo=TRUE>>=
811 RowTable(U$s7,U$s9)
812 @
813 }
814 \vspace{\stretch{1}}
815 \end{slide}
816 %</present>
817
818 %<*techreport>
819 \subsubsection{Attitude towards inter-marriage}
820
821 Question \code{s34a} is ``Would you personally mind if a
822 close relative were to marry a white person?'. It serves
823 as a filter question to \code{s34b} ('`Would you mind very
824 much or just a little?') which is asked to those who
825 answered yes (cod $1$) on \code{s34a}. We code those who
826 answer yes on both questions (Mind \& Mind very much) as
827 TRUE. The category FALSE refers then to those who (Do not
828 mind) or (Mind \& Mind Little). Code $8$ (Can't say) on
829 \code{s34a} and code $9$ on \code{s34a} and on \code{s34b}
830 are assigned as \code{NA}. \citep[p. 125]{FNSEM}
831
832 <<Attitude towards inter-marriage 1,keep.source=FALSE>>=
833 # table(U$s34a,exclude=c())
834 # table(U$s34b,exclude=c())
835 @
836 <<Attitude towards inter-marriage 2>>=
837 is.na(U$s34a) <- U$s34a==8 | U$s34a==9
838 is.na(U$s34b) <- U$s34b==8

```

```

839 U$Attitude.Towards.Inter.Marriage <-
840     U$s34a==1 & U$s34b==1
841 @
842 <<Attitude towards inter-marriage 3,keep.source=FALSE>>=
843 # table(U$Attitude.Towards.Inter.Marriage,exclude=c())
844 @
845
846 %</techreport>
847
848 %<*present>
849
850 \begin{slide}{Attitude towards inter-marriage}
851 \vspace{\stretch{1}}
852 \footnotesize{
853 What you mind very much if \dots
854 <<Attitudes, echo=TRUE>>=
855 RowTable(U$s7,U$Attitude.Towards.Inter.Marriage)
856 @
857 }
858 \vspace{\stretch{1}}
859 \end{slide}
860 %</present>
861
862 %<*techreport>
863
864 \subsubsection{Importance of racial composition in schools}
865
866 Two questions are asked: \code{s23} is
867 \begin{quote}
868 ``If you were choosing a school for an eleven-year old
869 child of yours, would your choice be influenced by how
870 many (RESPONDENT'S ETHNIC ORIGIN) children there were in
871 the school?'' \citep[p. 120]{FNSEM}
872 \end{quote}
873 and \code{s24a} asks that if the available school were
874 similar in other ways would you prefer to send this child
875 to school with fewer than half of the pupils (code $1$),
876 about half of the students (code $2$), more than half
877 (code $3$) were of your own ethnic origin. \code{s23} is
878 not a filter question. \emph{Importance of Racial
879 Composition in Schools} is set to TRUE if \code{s24a} is
880 equal to $3$ and FALSE otherwise. Code $7$ (No preference)
881 is coded as FALSE. Codes $8$ (Can't say) and $9$ are
882 assigned as \code{NA}. \citep[p. 120]{FNSEM}
883
884 <<Importance of racial composition in schools 1,keep.source=FALSE>>=
885 # table(U$s24a,exclude=c())
886 @
887 <<Importance of racial composition in schools 2>>=
888 is.na(U$s24a) <- U$s24a==8 | U$s24a==9
889 U$Importance.of.Racial.Composition.in.Schools <- U$s24a==3
890 @
891 <<Importance of racial composition in schools 3,keep.source=FALSE>>=
892 # table(U$Importance.of.Racial.Composition.in.Schools,
893 #       exclude=c())
894 @
895 %</techreport>
896
897 %<*present>
898

```

```

899 \begin{slide}{Importance of racial composition in schools}
900 \vspace{\stretch{1}}
901 \footnotesize{
902 <<Importance, echo=TRUE>>=
903 RowTable(U$Religion,
904           U$Importance.of.Racial.Composition.in.Schools)
905 @
906 }
907 \vspace{\stretch{1}}
908 \end{slide}
909
910
911
912 \begin{slide}[toc={}]{
913
914 \end{slide}
915
916 \begin{slide}[toc={}]{
917
918 \end{slide}
919
920 \begin{slide}[toc={}]{
921
922 \end{slide}
923
924 \begin{slide}[toc={}]{
925
926 \end{slide}
927
928 \begin{slide}[toc={}]{
929
930 \end{slide}
931
932 %</present>
933
934
935 %<*techreport>
936 \subsubsection{Born in the UK}
937
938 Defines who is born in the United Kingdom (question
939 \code{a3}). Category 16 is Northern Ireland, category 17
940 England and Wales and category 18 Scotland. Code 99 is
941 assigned as \code{NA}. \citep[p. 107]{FNSEM}.
942
943 <<Born in the UK 1,keep.source=FALSE>>=
944 #table(U$a3,exclude=c())
945 @
946 <<Born in the UK 2>>=
947 is.na(U$a3)<- U$a3==99
948 U$Born.in.the.UK <- U$a3==16 | U$a3==17 | U$a3==18
949 @
950 <<Born in the UK 3,keep.source=FALSE>>=
951 #table(U$Born.in.the.UK,exclude=c())
952 @
953
954 \subsubsection{Age at and years since arrival}
955
956 This part defines the variables \emph{Age at arrival} and
957 \emph{Years since arrival} by using information about year
958 of migration (question \code{a4n}), age (question

```

```

959 \code{alan}) and the year (variable \code{year} when the
960 interview is made). The interview is made in \code{93} or
961 \code{94}). The result is that some individuals get the age
962 at arrival  $-1$ , which presumably is due to rounding of
963 years since arrival and the age.
964
965 All born in the UK are coded as  $0$  for
966 \code{Age.at.Arrival} and \code{Years.Since Arrival}. Since
967 these two variables are related to the age of the
968 immigrants, one could also add an interaction variable
969 between \code{Born.in.the.UK} and \code{Age} to account for
970 effect of age for the natives. The effect of age for natives
971 is not represented in the model as specified by
972 \citet{Bisin08}. We experimented with this and results were
973 basically unchanged. The interaction variable is
974 insignificant in all specifications. Code  $99$  is assigned
975 as \code{NA} for \code{year}. \citep[pp. 105, 107]{FNSEM}.
976
977 <<Age at and years since arrival 1,keep.source=FALSE>>=
978 # table(U$a4n,exclude=c())
979 # table(U$alan,exclude=c())
980 # table(U$year,exclude=c())
981 @
982 <<Age at and years since arrival 2>>=
983 is.na(U$a4n) <- U$a4n==98 | U$a4n==99
984 is.na(U$year)<- U$year==99
985 U$Age <- U$alan
986 U$Years.Since.Arrival <- ifelse(U$Born.in.the.UK==TRUE,
987     0, U$year-U$a4n)
988 U$Age.at.Arrival <- U$Age - U$Years.Since.Arrival
989 U$Age.at.Arrival <- replace(U$Age.at.Arrival,
990     U$Born.in.the.UK==TRUE, 0)
991 @
992 <<Age at and years since arrival 3,keep.source=FALSE>>=
993 # table(U$Years.Since.Arrival,exclude=c())
994 # table(U$Age.at.Arrival,exclude=c())
995 @
996
997 \subsubsection{Female}
998 Definition of females via question \code{hh2a.s}. Code
999 \code{na} is coded as \code{NA}. \citep[pp. 318]{FNSEM}.
1000
1001 <<Female 1,keep.source=FALSE>>=
1002 # table(U$hh2a.s,exclude=c())
1003 @
1004 <<Female 2>>=
1005 is.na(U$hh2a.s) <- U$hh2a.s == "na"
1006 U$Female <- U$hh2a=="female"
1007 @
1008 <<Female 3,keep.source=FALSE>>=
1009 # table(U$Female,exclude=c())
1010 @
1011
1012 \subsubsection{Arranged marriage}
1013 \label{sec:Arranged.Marriage}
1014 In question \code{s39} Indian, Pakistani, Bangladeshi and
1015 Chinese respondents who has ever been married were asked a
1016 question about the decision regarding their marriage. The
1017 question ask about the role of the respondent and his or her
1018 parents about choosing the respondent's husband/wife. In

```

```

1019 categories $1$ and $2$ of \code{s39} the respondent's parents
1020 made the final decision and these categories define the
1021 dummy where the respondent is or has been living in an
1022 arranged marriage (code $1$; all other are coded $0$).
1023 Notice that singles and Caribbeans have not received this
1024 question. Singles are already removed from the sample (see
1025 section \ref{sec:definingsample}. Caribbeans are coded $0$.
1026 This means that the Caribbeans do not marry according to the
1027 decision of their parents. Category $8$ (Can't say) and
1028 category $9$ are assigned \code{NA}. \citep[pp. 127]{FNSEM}.
1029 <<Arranged marriage 1,keep.source=FALSE>>=
1030 # table(U$s39,exclude=c())
1031 @
1032 <<Arranged marriage 2>>=
1033 is.na(U$s39) <- U$s39==8 | U$s39==9
1034 U$Arranged.Marriage <- ifelse(
1035     U$s39==3 | U$s39==4 | U$s39==5 |
1036     U$ethnic=="caribbean", FALSE , TRUE)
1037 @
1038 <<Arranged marriage 3,keep.source=FALSE>>=
1039 # table(U$Arranged.Marriage,exclude=c())
1040 @
1041
1042 \subsubsection{Discrimination}
1043
1044 The \emph{discrimination} variable is based on a series of
1045 questions related to discrimination; \code{v1a}-\code{v1d}
1046 about physical attacks, \code{v9a} about insults,
1047 \code{j55a} and \code{j63a} discrimination at work.
1048 Basically. anyone answering that they have been
1049 discriminated for any of these reasons are coded
1050 $1$; else code $0$.
1051
1052 Questions \code{v1a}-\code{v1d} is a series of filter
1053 questions: Question \code{v1a} asks if the respondent have
1054 been attacked (yes or no), question \code{v1b} how many
1055 attacks the respondent has been enduring and question,
1056 \code{v1c} asks those who have been attacked once if they
1057 believe the attack had to do with reasons to do with race or
1058 colour and \code{v1d} asks the same question and
1059 regards those who have been attacked more than once.
1060 Generally code $8$ and code $9$ are assigned \code{NA},
1061 except for question
1062 \code{v1b} where also code $7$ is assigned \code{NA}.
1063 \citep[pp. 154ff, 163, 195 and 199]{FNSEM}.
1064 <<Discrimination 1>>=
1065 vjlist <- c(paste("v1",letters[1:4],sep=""),
1066     "v9a","j55a","j63a")
1067 @
1068
1069 <<Discrimination 3>>=
1070 is.na(U[vjlist]) <- U[vjlist]==8 |
1071     U[vjlist]==9
1072 is.na(U$v1b) <- U$v1b==7 | U$v1b==8 |
1073     U$v1b==9
1074 U$Discrimination <-
1075     (U$v1a==1 & U$v1b==1 & U$v1c==1) |
1076     (U$v1a==1 &
1077     (U$v1b >= 2 & U$v1b <= 6) &
1078     U$v1d==1) | U$v9a==1 | U$j55a==1 |

```

```

1079     U$j63a==1
1080 @
1081 <<Discrimination 4,keep.source=FALSE>>=
1082 # table(U$Discrimination,exclude=c())
1083 @
1084
1085 \subsubsection{Children}
1086
1087 No question about the number of children is asked. Instead
1088 the number of children has to be calculated indirectly via
1089 the number of children not living at home (questions
1090 \code{f16a} and \code{f16b1n}-\code{f16b3n}) and the
1091 relation between the respondent and other persons living in
1092 the household (questions \code{hh2c.b}-\code{hh2.c.m}).
1093
1094 The number of children out of home is calculated in the
1095 following way: If children out of home is TRUE
1096 (\code{f16a=1}), then the number of children equals the sum
1097 of \code{f16b1n}, \code{f16b2n} and \code{f16b3n} (the
1098 number of children not living at home below 5 years, between
1099 5 and 15 years and above 15 years of age). Else, if there
1100 are no children out of home (i.e., if \code{f16a=2}), then
1101 the number of children out of home is set to $0$.
1102 Missing values are coded as below. \citep[p. 57]{FNSEM}.
1103
1104 <<Children 1,keep.source=FALSE>>=
1105 # table(U$f16a,exclude=c())
1106 # table(U$f16b1n,exclude=c())
1107 # table(U$f16b2n,exclude=c())
1108 # table(U$f16b3n,exclude=c())
1109 @
1110 <<Children 2>>=
1111 is.na(U$f16a)     <- U$f16a==8 | U$f16a==9
1112 is.na(U$f16b1n)  <- U$f16b1n==99
1113 is.na(U$f16b2n)  <- U$f16b2n==99
1114 is.na(U$f16b3n)  <- U$f16b3n==98 |
1115     U$f16b3n==99
1116 U$Child.not.at.Home <- ifelse(U$f16a==1,
1117     U$f16b1n+U$f16b2n+U$f16b3n,0)
1118 @
1119
1120 Questions \code{hh2c.b}-\code{hh2.c.m} are about the
1121 relationship between the respondent and other individuals in
1122 the household (person b, c, d etc to person m); category $5$
1123 being child of the respondent. First we check if the
1124 person is a child to the respondent and then all children
1125 are summed over the respondents household adding the
1126 variable measuring number of children not at home.
1127 Generally codes $98$ and $99$ are assigned \code{NA}.
1128 \citep[p. 318]{FNSEM}.
1129
1130 <<Children 3>>=
1131 hhlist <- c(paste("hh2c.", letters[2:13], sep=""))
1132 @
1133
1134
1135 <<Children 4>>=
1136 is.na(U[hhlist]) <- U[hhlist]==98 | U[hhlist]==99
1137 U$Children <-
1138     U$Child.not.at.Home + apply(apply(subset(U,

```

```

1139     select=c(hh2c.b:hh2c.m)), 2,
1140     function(x) x==5),1, function(x) sum(x,
1141     na.rm=TRUE))
1142 @
1143 <<Children 5,keep.source=FALSE>>=
1144 # table(U$Children,exclude=c())
1145 @
1146 \subsubsection{No British education}
1147 Question \code{q1} asks whether the respondent has any
1148 British education. Code $2$ is no. Code $8$ (Can't say) is
1149 kept in the alternative category since these individuals
1150 will answer the question \code{q3} about foreign education.
1151 Code $9$ in \code{q1} is assigned \code{NA}. \citep[p.
1152 96]{FNSEM}.
1153 <<No British Education 1,keep.source=FALSE>>=
1154 # table(U$q1,exclude=c())
1155 @
1156 <<No British Education 2>>=
1157 is.na(U$q1) <- U$q1==9
1158 U$No.British.Education <- U$q1==2
1159 @
1160 <<No British Education 3,keep.source=FALSE>>=
1161 # table(U$No.British.Education,exclude=c())
1162 @
1163
1164 \subsubsection{British basic education}
1165
1166 We could not exactly see how this variable was defined in
1167 \citet{Bisin08}. They define the British high education as
1168 A--level and above. One interpretation is then that 0-level
1169 are educations included in the basic level. This
1170 interpretation is implemented here. \code{NA} is assigned to
1171 all observations for which the filter question
1172 \emph{No British Education} was \code{NA}.
1173 \citep[pp. 96ff]{FNSEM}.
1174
1175 <<British basic education 1>>=
1176 q2alist <- c(paste("..q2a", c(1:8,12:18), sep=""))
1177 @
1178 <<British basic education 2,keep.source=FALSE>>=
1179 # apply(U[q2alist], 2, function(x) table(x,
1180 #     exclude=c()))
1181 @
1182 <<British basic education 3>>=
1183 U$British.Basic.Education <-
1184     apply(apply(U[q2alist] ,2,function(x)
1185     x==1),1, function(x) sum(x, na.rm=TRUE))!=0
1186 U$British.Basic.Education <- ifelse(
1187     is.na(U$No.British.Education),
1188     NA,U$British.Basic.Education)
1189 @
1190 <<British basic education 4,keep.source=FALSE>>=
1191 # table(U$British.Basic.Education,exclude=c())
1192 @
1193
1194 \subsubsection{British higher education}
1195 \citet{Bisin08} explicitly defined British higher education
1196 as A-level. Given the definition of British Basic education,
1197 the reference group will include trade apprenticeships as
1198 well as university educations. \code{NA} is assigned to all

```

```

1199 observations for which the filter question
1200 \emph{No British Education} was \code{NA}.
1201 \citep[pp. 96ff]{FNSEM}.
1202
1203 <<British higher education 1,keep.source=FALSE>>=
1204 # table(U$..q2a9,exclude=c())
1205 # table(U$..q2a10,exclude=c())
1206 # table(U$..q2a11,exclude=c())
1207 @
1208 <<British higher education 2>>=
1209 U$British.Higher.Education <- apply(apply(subset(
1210     U,select=c(..q2a9:..q2a11,..q2a19,..q2a20))
1211     ,2,function(x) x==1),1,
1212     function(x) sum(x, na.rm=TRUE))!=0
1213 U$British.Higher.Education <- ifelse(
1214     is.na(U$No.British.Education),
1215     NA,U$British.Higher.Education)
1216 @
1217 <<British higher education 3,keep.source=FALSE>>=
1218 # table(U$British.Higher.Education,exclude=c())
1219 @
1220
1221 \subsubsection{Foreign education}
1222
1223 Foreign educations is question \code{q3} asked to all who
1224 answered `no' or `Can't say' on question \code{q1}.
1225 The answer yes is coded as (code $1$) and no is coded
1226 as (code $2$). Contrary to the above
1227 educational variables \code{NA} is \emph{not} assigned to
1228 all \code{NA} on \emph{No British Education} since some of
1229 them (code $8$) actually was asked the question \code{q3}.
1230 Instead \code{NA} is assigned to all observations for which
1231 the filter question \code{q1} was $1$ or $9$ and to all code
1232 $8$ (Can't say) and code $9$ on question \code{q3}. \citep[p.
1233 99]{FNSEM}.
1234
1235 <<Foreign education 1,keep.source=FALSE>>=
1236 # table(U$q3,exclude=c())
1237 @
1238 <<Foreign education 2>>=
1239 is.na(U$q3) <- U$q3==8 | U$q3==9
1240 U$Foreign.Education <- U$q3==1
1241 @
1242 <<Foreign education 3,keep.source=FALSE>>=
1243 # table(U$Foreign.Education,exclude=c())
1244 @
1245
1246 \subsubsection{Labour market status}
1247
1248 We code the labour market status using \code{j1b} (in paid
1249 work last week or not) and \code{j3occ} (classification of
1250 activity; either last week's activity or potential activity
1251 during the last ten years). The variable \code{j1b} takes
1252 the value $1$ for paid work last week and $2$ otherwise. The
1253 variable \code{j3occ} is coded as follows: \citep[the
1254 former pp. 81f]{FNSEM,FNSEM93b}
1255 \vspace{6pt}
1256 \begin{compactenum}
1257 \item Self-employed (25+ employees)
1258 \item Self-employed (1-24 employees)

```

```

1259 \item Self-employed (no employees)
1260 \item Self-employed (employees not known)
1261 \item Manager (establishment of 25+ employees)
1262 \item Manager (establishment of 1-24 employees)
1263 \item Manager (employees not known)
1264 \item Foreman/supervisor
1265 \item Other employee
1266 \item Employee status unknown
1267 \item Not known/not answered
1268 \end{compactenum}
1269 \vspace{6pt}
1270
1271 \paragraph{Employee}
1272 In order to be classified as an employee the individual has
1273 to have answered yes (value $1$) in \code{j1b} and be
1274 classified as employee in \code{j3occ} (value $9$) and
1275 have the value \code{NotAssigned}.\code{NA} is assigned to
1276 categories $10$ and $11$ in \code{j3occ}.
1277
1278 <<Labour market status 1,keep.source=FALSE>>=
1279 # table(U$j1b,exclude=c())
1280 # table(U$j3occ,exclude=c())
1281 @
1282 <<Labour market status 2>>=
1283 is.na(U$j3occ) <- U$j3occ==10 | U$j3occ==11
1284 U$Labour.Market.Status <- ifelse(U$j1b==1 &
1285     U$j3occ==9 & !is.na(U$j1b==1 &
1286     U$j3occ==9), "Employee", "NotAssigned")
1287 @
1288
1289 \paragraph{Self Employed}
1290
1291 Self--employed are also coded using \code{j1b} and
1292 \code{j3occ}; above. Categories $1-4$ in \code{j3occ} are
1293 defined as self-employed. We also require that \code{j1b} is
1294 equal $1$ (Self--Employed).
1295
1296
1297 <<Labour market status 3>>=
1298 U$Labour.Market.Status <- replace(
1299     U$Labour.Market.Status,
1300     (U$j3occ==1 | U$j3occ==2 |
1301     U$j3occ==3 | U$j3occ==4) &
1302     U$j1b==1, "SelfEmployed")
1303 @
1304
1305
1306 \paragraph{Manager}
1307 Managers are also coded using \code{j1b} and \code{j3occ};
1308 see above. Categories $5-8$ in \code{j3occ} are defined as
1309 managers (including supervisors).
1310
1311 <<Labour market status 4>>=
1312 U$Labour.Market.Status <-
1313     replace(U$Labour.Market.Status,
1314     U$j1b==1 & U$j3occ>4 & U$j3occ<9,
1315     "Manager")
1316 @
1317
1318 \paragraph{Unemployed}

```

1319 The question `{hh5d.s}` describes the respondent's labour
1320 market status. Unemployment is defined via this variable.
1321 `{hh5d.s}` is coded in the following way
1322 `{FNSEM93b}`:
1323 `{6pt}`
1324 `{compactenum}`
1325 `{item}` Full-time education
1326 `{item}` Govt. training programme
1327 `{item}` Full-time paid work
1328 `{item}` Part-time paid work
1329 `{item}` Waiting to take up paid work
1330 `{item}` Registered unemployed
1331 `{item}` Unemployed, not registered
1332 `{item}` Permanently sick or disabled
1333 `{item}` Wholly retired from work
1334 `{item}` Looking after the home
1335 `{item}` Doing something else
1336 `{item}` NA
1337 `{end{compactenum}}`
1338 `{6pt}`
1339 We define unemployed as category `{6$}` and `{7$}`. `{citep[p.}`
1340 `{324}{Bisin08}`. There are few cases where the individual is
1341 classified as Employee according to our definition above and
1342 is reported to be unemployed in `{hh5d.s}`. This can for
1343 example be part-time unemployment. We classify these
1344 individuals as having `{Unclear}` labour market status. These
1345 will be checked later and be removed if they are few in the
1346 final sample.
1347
1348 `<<Labour market status 5>>=`
1349 `U$Labour.Market.Status <-`
1350 `replace(U$Labour.Market.Status,`
1351 `(U$hh5d.s==6 | U$hh5d.s == 7) &`
1352 `U$Labour.Market.Status!="NotAssigned",`
1353 `"Unclear")`
1354 `U$Labour.Market.Status <-`
1355 `replace(U$Labour.Market.Status,`
1356 `(U$hh5d.s==6 | U$hh5d.s == 7) &`
1357 `U$Labour.Market.Status=="NotAssigned",`
1358 `"Unemployed")`
1359 `@`
1360
1361 `{paragraph{Out of labour force}}`
1362
1363 The category out of labour force is defined as those having
1364 values (1,2,5,8,9,10,11) in `{hh5d.s}` or value `{2$}` in
1365 `{j1b}` or `{j2}`.
1366 `<<Labour market status 6>>=`
1367 `U$Labour.Market.Status <-`
1368 `replace(U$Labour.Market.Status,`
1369 `(U$hh5d.s==1 | U$hh5d.s == 2 |`
1370 `U$hh5d.s==5 | U$hh5d.s == 8 |`
1371 `U$hh5d.s==9 | U$hh5d.s == 10|`
1372 `U$hh5d.s==11 | U$j1b == 2 |`
1373 `U$j2 == 2) &`
1374 `U$Labour.Market.Status=="NotAssigned",`
1375 `"OutOfLabourForce")`
1376 `@`
1377
1378 Remaining observations with the value `{NotAssigned}` in

```

1379 \code{Labour Market.Status} will be assigned \code{NA}.
1380 At this point there are
1381 \Sexpr{table(U$Labour.Market.Status)[["Unclear"]]}
1382 observations coded as \code{Unclear}. These are now recoded
1383 as \code{NA}.
1384
1385 <<Labour market status 8>>=
1386 is.na(U$Labour.Market.Status) <-
1387     U$Labour.Market.Status == "NotAssigned"
1388 is.na(U$Labour.Market.Status) <-
1389     U$Labour.Market.Status == "Unclear"
1390 @
1391
1392 We will create dummy variables using this variable before we
1393 run our models.
1394
1395 <<Labour market status 9,keep.source=FALSE>>=
1396 # table(U$Labour.Market.Status,exclude=c())
1397 @
1398
1399 \subsubsection{No parents}
1400
1401 Variable \emph{No parents} means that the respondent is not
1402 living with his or her parents. This variable is coded with
1403 question \code{f14a} (which takes code $1$ for both alive,
1404 code $2$ father alive, code $3$ for mother alive and $8$ for
1405 both dead) and \code{f14b} (which takes code $2$ if both
1406 living parents do not live with the respondent and code $6$
1407 if the only living parent does not live with the respondent;
1408 else it takes one of the values $1$ or $3-5$).
1409 \emph{No parents} should be coded TRUE if either both
1410 parents are dead or the respondent does not live with any
1411 living parents. However,
1412 since we follow \citet{Bisin08}, we code this variable
1413 including only those who have both their parents dead or
1414 both paprents live away from the respondent. This definition
1415 implies that those who have a parent living away and one
1416 parent dead are assigned the value FALSE. Code $8$ and $9$
1417 are assigned \code{NA} \citep[p. 56]{FNSEM}
1418 <<No parents 1,keep.source=FALSE>>=
1419 # table(U$f14a,exclude=c())
1420 # table(U$f14b,exclude=c())
1421 @
1422 <<No parents 2>>=
1423 is.na(U$f14a) <- U$f14a==9
1424 is.na(U$f14b) <- U$f14b==9
1425 U$No.Parents <- U$f14a==8 | U$f14b==2
1426 @
1427 <<No parents 3,keep.source=FALSE>>=
1428 # table(U$No.Parents,exclude=c())
1429 @
1430
1431 \subsubsection{Contacts with parents}
1432
1433 The three variables measuring contacts with parents are
1434 defined via three questions asking about the number of
1435 physical contacts (question \code{f15bn}), the number of
1436 contacts via telephone (question \code{f15cn}) and the
1437 number of contacts via letters (question \code{f15dn}) that
1438 the respondent has had with his or her parents during the

```

```

1439 last four weeks conditional on not both parents being dead.
1440 All three takes the value of the underlying variable if at
1441 least one parent is alive and the value $0$ if both parents
1442 are dead. Code $999$ on all three variables and code $997$
1443 on \code{f15bn} are assigned \code{NA}. \citep[p. 56]{FNSEM}
1444
1445 <<Contacts with parents 1,keep.source=FALSE>>=
1446 # table(U$f15bn,exclude=c())
1447 # table(U$f15cn,exclude=c())
1448 # table(U$f15dn,exclude=c())
1449 @
1450 <<Contacts with parents 2>>=
1451 is.na(U$f15bn) <- U$f15bn==997 | U$f15bn==999
1452 is.na(U$f15cn) <- U$f15cn==999
1453 is.na(U$f15dn) <- U$f15dn==999
1454 U$Parents.Physical.Contacts <- ifelse(
1455     U$f14a!=8, U$f15bn,0)
1456 U$Parents.Telephone.Calls <- ifelse(
1457     U$f14a!=8, U$f15cn,0)
1458 U$Parents.Letters <- ifelse(
1459     U$f14a!=8, U$f15dn,0)
1460 @
1461 <<Contacts with parents 3,keep.source=FALSE>>=
1462 # table(U$Parents.Physical.Contacts,exclude=c())
1463 # table(U$Parents.Telephone.Calls,exclude=c())
1464 # table(U$Parents.Letters,exclude=c())
1465 @
1466
1467
1468 \subsubsection{English language}
1469
1470 There are several language variables measuring whether the
1471 respondent is speaking English with different individuals;
1472 at home with older, at home with younger, at work and with
1473 friends. We construct these variables using question
1474 \code{s12a} which asks whether the respondent regularly
1475 speak to anyone in Britain in any other language than
1476 English and \code{..s12f}, \code{..s12g}, \code{..s12h} and
1477 \code{..s12i} which asks which language is spoken to the
1478 above mentioned categories of individuals. Each of
1479 \code{..s12f}, \code{..s12g}, \code{..s12h} and
1480 \code{..s12i} comes in 18 versions (e.g.,
1481 \code{..s12f1},\dots,\code{..s12f18}) where each question
1482 1--15 is coded \code{yes} if the respondent speaks the
1483 language. Question 16 is ``Never speaks to these people/Not
1484 applicable'', question 17 NA and question 18 ``None of
1485 the above answered positive''. Question 1 is always
1486 regarding English.
1487
1488 The respondent is coded as English speaker if either
1489 \code{s12a} is answered negatively or \code{..s12aX1}, where
1490 \code{X=f,g,h,i}, is answered positively. Only respondents
1491 for which either \code{s12a} is \code{NA} or all of
1492 \code{..s12X1},\dots,\code{..s12X16} are answered negatively
1493 are coded as \code{NA}. Below is the code for \emph{English
1494 Spoken At Home With Older}:
1495 <<English language 1,keep.source=FALSE>>=
1496 # table(U$s12a,exclude=c())
1497 # table(U$..s12f1,exclude=c())
1498 # table(U$..s12g1,exclude=c())

```

```

1499 # table(U$.s12h1,exclude=c())
1500 # table(U$.s12i1,exclude=c())
1501 @
1502 <<English language 2>>=
1503 is.na(U$s12a) <- U$s12a== 8 | U$s12a == 9
1504 s12f1list <- c(paste("..s12f", 2:16, sep=""))
1505 U$oOLD <- apply(U[s12f1list]=="yes", 1, sum)
1506 U$English.Spoken.at.Home.with.Older <-
1507   ((U$s12a==1 | is.na(U$s12a)) &
1508     U$.s12f1=="yes") | U$s12a==2
1509 is.na(U$English.Spoken.at.Home.with.Older) <-
1510   U$oOLD==0 &
1511   U$English.Spoken.at.Home.with.Older==FALSE
1512 U$English.Spoken.at.Home.with.Older <-
1513   replace(U$English.Spoken.at.Home.with.Older,
1514     U$oOLD>0 &
1515     is.na(U$English.Spoken.at.Home.with.Older),FALSE)
1516 U$DO.NOT.SPEAK.WITH.OLDER <- ifelse(U$.s12f16=="no",0,1)
1517 @
1518
1519 The codes for \emph{English Spoken At Home With Younger},
1520 \emph{English Spoken At Work } and \emph{English Spoken With
1521 Friends} are equivalent. These codes are in
1522 \code{araietal\_source.Rnw} but not shown here. It is also
1523 available in \code{araietal\_source.R}.
1524
1525 <<English language 3,echo=FALSE,keep.source=FALSE>>=
1526 s12g1list <- c(paste("..s12g", 2:16, sep=""))
1527 U$oYOUNG <- apply(U[s12g1list]=="yes", 1, sum)
1528 U$English.Spoken.at.Home.with.Younger <-
1529   ((U$s12a==1 | is.na(U$s12a)) &
1530     U$.s12g1=="yes") | U$s12a==2
1531 is.na(U$English.Spoken.at.Home.with.Younger) <-
1532   U$oYOUNG==0 &
1533   U$English.Spoken.at.Home.with.Younger==FALSE
1534 U$English.Spoken.at.Home.with.Younger <-
1535   replace(U$English.Spoken.at.Home.with.Younger,
1536     U$oYOUNG>0 &
1537     is.na(U$English.Spoken.at.Home.with.Younger),
1538     FALSE)
1539 U$DO.NOT.SPEAK.WITH.YOUNGER <- ifelse(U$.s12g16=="no",0,1)
1540 @
1541
1542 <<English language 4,echo=FALSE,keep.source=FALSE>>=
1543 s12h1list <- c(paste("..s12h", 2:16, sep=""))
1544 U$oWORK <- apply(U[s12h1list]=="yes", 1, sum)
1545 U$English.Spoken.at.Work <-
1546   ((U$s12a==1 | is.na(U$s12a)) &
1547     U$.s12h1=="yes") | U$s12a==2
1548 is.na(U$English.Spoken.at.Work) <-
1549   U$oWORK==0 & U$English.Spoken.at.Work==FALSE
1550 U$English.Spoken.at.Work <-
1551   replace(U$English.Spoken.at.Work,
1552     U$oWORK>0 & is.na(U$English.Spoken.at.Work),
1553     FALSE)
1554 U$DO.NOT.SPEAK.AT.WORK <- ifelse(U$.s12h16=="no",0,1)
1555 @
1556 <<English language 5,echo=FALSE,keep.source=FALSE>>=
1557 s12i1list <- c(paste("..s12i", 2:16, sep=""))
1558 U$oFRIENDS <- apply(U[s12i1list]=="yes",1,sum)

```

```

1559 U$English.Spoken.With.Friends <-
1560   ((U$s12a==1 | is.na(U$s12a)) &
1561    U$.s12i1=="yes") | U$s12a==2
1562 is.na(U$English.Spoken.With.Friends) <-
1563   U$oFRIENDS==0 &
1564   U$English.Spoken.With.Friends==FALSE
1565 U$English.Spoken.With.Friends <-
1566   replace(U$English.Spoken.With.Friends,
1567           U$oFRIENDS>0 &
1568           is.na(U$English.Spoken.With.Friends), FALSE)
1569 U$DO.NOT.SPEAK.WITH.FRIENDS <- ifelse(U$.s12i16=="no",0,1)
1570 @
1571 <<English language 6,keep.source=FALSE>>=
1572 # table(U$English.Spoken.at.Home.with.Older,
1573 #       exclude=c())
1574 # table(U$English.Spoken.at.Home.with.Younger,
1575 #       exclude=c())
1576 # table(U$English.Spoken.at.Work,exclude=c())
1577 # table(U$English.Spoken.With.Friends,exclude=c())
1578 @
1579
1580
1581
1582 \subsubsection{Household income}
1583
1584 The question \code{hh40} provides information in which
1585 interval the household income of the respondent's household
1586 is. We assign the midpoints in these intervals as the
1587 household income. For the lowest bracket this income is the
1588 midpoint of  $[0,77]$ $. For the highest bracket we assign the
1589 income which is the lowest income in the bracket plus the
1590 income interval down to the midpint of the second highest
1591 bracket; i.e.,
1592  $789 + \frac{788 - 731}{2} = \text{Sexpr}\{\text{round}(789 + (788 - 731)/2, 1)\}$ $. This
1593 method underestimates the range of the highest bracket but
1594 to a lesser extent than the lower limit  $789$ $.
1595 <<Household income 1,keep.source=FALSE>>=
1596 # table(U$hh40,exclude=c())
1597 @
1598 <<Household income 2>>=
1599 is.na(U$hh40) <- {U$hh40=="refused" |
1600   U$hh40=="can't say" | U$hh40=="na"}
1601 U$Household.Income <- c(38.5,96.5,135.0,
1602   173.5,230.5,260.0,318.0,366.0,414.0,
1603   471.5,529.5,587.0,649.5,702.0,759.5,
1604   817.5)[U$hh40]
1605 @
1606 <<Household income 3,keep.source=FALSE>>=
1607 # table(U$Household.Income,exclude=c())
1608 @
1609
1610 \subsubsection{Ward variables}
1611
1612 Ward density of own ethnic group is measured by the variable
1613 \code{wown} in the original data. This variable is coded
1614  $1-7$  depending on the density of the respondent's own
1615 ethnic group is in the ward of the respondent
1616 \citep{FNSEM93b}:
1617 \vspace{6pt}
1618 \begin{compactenum}

```

```

1619 \item Up to 1.99\%
1620 \item 2-4.99\%
1621 \item 5-9.99\%
1622 \item 10-14.99\%
1623 \item 15-24.99\%
1624 \item 25-32.99\%
1625 \item 33\% or more
1626 \end{compactenum}
1627 \vspace{6pt}
1628 We recode the variable to take the midpoints of the
1629 density intervals in the same fashion as the household
1630 income variable was recoded. This means that if the
1631 respondent is the lowest interval the density is set to be
1632 $1$ etc. In the highest interval we set the density to be
1633 the lowest density in the interval plus the density distance
1634 down to the midpoint of the second highest density interval;
1635 i.e.,  $33 + \frac{33-25}{2} = \text{Sexpr}\{\text{round}(33+(33-25)/2,1)\}$ $.
1636
1637 The variable \code{wunemp} is coded $1-6$ depending on
1638 unemployment rate in the ward of the respondent
1639 \citep{FNSEM93b}:
1640 \vspace{6pt}
1641 \begin{compactenum}
1642 \item Up to 1.99\%
1643 \item 2-4.99\%
1644 \item 5-9.99\%
1645 \item 10-14.99\%
1646 \item 15-20\%
1647 \item 20\% or more
1648 \end{compactenum}
1649 \vspace{6pt}
1650 We recode this variable to instead take the midpoints of the
1651 intervals as we did for the household income.
1652 This means that if the respondent is the lowest interval the
1653 the rate is set to be $1$ etc. In the highest interval we
1654 set the rate to be the lowest rate in the interval plus
1655 the distance down to the midpoint of the second
1656 highest rate interval; i.e.,
1657  $20 + \frac{20-15}{2} = \text{Sexpr}\{\text{round}(20+(20-15)/2,1)\}$ $.
1658
1659 <<Ward variables 1,keep.source=FALSE>>=
1660 # table(U$wown,exclude=c())
1661 @
1662 <<Ward variables 2>>=
1663 U$Ward.Density.Own.Ethnicity <- (c(1.0,3.5,
1664     7.5,12.5,20.0,29.0,37.0)[U$wown])/100
1665 U$Ward.Unemployment.Rate <- c(1.0,3.5,7.5,12.5,
1666     17.5,22.5)[U$wunemp]
1667 @
1668 <<Ward variables 3,keep.source=FALSE>>=
1669 # table(U$Ward.Density.Own.Ethnicity,exclude=c())
1670 # table(U$Ward.Unemployment.Rate,exclude=c())
1671 @
1672
1673 \subsection{Discrimination own ethnicity}
1674
1675 Finally we define a variable describing the discrimination
1676 against the own ethnic group. It is defined as the average
1677 of the variable \emph{Discrimination} over ethnic groups
1678 after the removal of non--availables.

```

```

1679
1680 <<Another variable definition>>=
1681 GroupDiscrimination <-
1682     tapply(U$Discrimination,U$ethnic,
1683           function(x) mean(x, na.rm=TRUE))
1684 U$Discrimination.Own.Ethnicity <-
1685     GroupDiscrimination[U$ethnic]
1686 @
1687
1688 \subsection{Defining the subset}
1689 \label{sec:subset}
1690
1691 We define dummy variables for labour market status to have
1692 the same variable labels as in \citet{Bisin08}.
1693
1694 <<Defining dummies for Labour Market Status>>=
1695 U$Employee <- as.numeric(U$Labour.Market.Status=="Employee")
1696 U$Manager <- as.numeric(U$Labour.Market.Status=="Manager")
1697 U$Self.Employed <-
1698     as.numeric(U$Labour.Market.Status=="SelfEmployed")
1699 U$OUT.OF.LABOUR.FORCE <-
1700     as.numeric(U$Labour.Market.Status=="OutOfLabourForce")
1701 U$Unemployed <-
1702     as.numeric(U$Labour.Market.Status=="Unemployed")
1703 @
1704
1705 We save a data set keeping all observations containing
1706 non--availables:
1707
1708 <<Defining the subset 1>>=
1709 U.Original <- U
1710 @
1711 We then choose the variables to keep in \code{U}.
1712 This code is in \code{araietal\_source.Rnw} but not shown
1713 here. It is also available in \code{araietal\_source.R}.
1714
1715
1716 <<Defining the subset 2,keep.source=FALSE,echo=FALSE>>=
1717 U <- subset(U, select=c(
1718     Religion,
1719     s7,
1720     ethnic,
1721     Importance.of.Religion,
1722     Attitude.Towards.Inter.Marriage,
1723     Importance.of.Racial.Composition.in.Schools,
1724     Age.at.Arrival,
1725     Age,
1726     Female,
1727     Born.in.the.UK,
1728     Arranged.Marriage,
1729     Discrimination,
1730     Children,
1731     Years.Since.Arrival,
1732     No.British.Education,
1733     British.Basic.Education,
1734     British.Higher.Education,
1735     Foreign.Education,
1736     Employee,
1737     Manager,
1738     Self.Employed,

```

```

1739 OUT.OF.LABOUR.FORCE,
1740 Unemployed,
1741 No.Parents,
1742 Parents.Physical.Contacts,
1743 Parents.Telephone.Calls,
1744 Parents.Letters,
1745 English.Spoken.at.Home.with.Older,
1746 DO.NOT.SPEAK.WITH.OLDER,
1747 English.Spoken.at.Home.with.Younger,
1748 DO.NOT.SPEAK.WITH.YOUNGER,
1749 English.Spoken.at.Work,
1750 DO.NOT.SPEAK.AT.WORK,
1751 English.Spoken.With.Friends,
1752 DO.NOT.SPEAK.WITH.FRIENDS,
1753 Household.Income,
1754 Ward.Density.Own.Ethnicity,
1755 Ward.Unemployment.Rate,
1756 Discrimination.Own.Ethnicity,
1757 area,
1758 weightis))
1759 @
1760
1761 Finally we remove all observations containing
1762 non--availables from U:
1763 <<Defining the subset 3>>=
1764 U <- na.omit(U)
1765 @
1766
1767 \subsection{Sample statistics}
1768 %</techreport>
1769 %<*paper>
1770 In our definition of the relevant (ethnic minority) sample
1771 we have excluded (i) the UK majority population (defined as
1772 Whites in the data set), (ii) all who do not have a
1773 religion or do not belong to a church since they cannot be
1774 classified in a religious group, (iii) all singles since
1775 only people who are married or have been married answer the
1776 question about who made the final decision of their marriage
1777 and (iv) those who answered questionnaire ``yellow'' and
1778 ``pink'' since they do not answer relevant questions
1779 involved in the study.
1780 %</paper>
1781 %<*paper|techreport>
1782 Table \ref{ta:diffsample} compares the number of
1783 observations in this sample \emph{before}
1784 %</paper|techreport>
1785 %<*techreport>
1786 (see section \ref{sec:subset}; i.e. data frame
1787 U.Original)
1788 %</techreport>
1789 %<*paper|techreport>
1790 and \emph{after}
1791 %</paper|techreport>
1792 %<*techreport>
1793 (see section \ref{sec:subset}; i.e. data frame U)
1794 %</techreport>
1795 %<*paper|techreport>
1796 non--availables are removed with the numbers of observations
1797 reported by the \citet{Bisin08} study. The number of
1798 observations for various groups in the non--Muslim category

```

1799 are not reported in the \citet{Bisin08} paper. These numbers
 1800 are therefore missing in the table. The category definitions
 1801 are from the original dataset and involves no recoding on
 1802 our part.

1803

1804 After removing observations with missing values on all
 1805 variables of interest ('`After'' in Table
 1806 \ref{ta:diffsample}), we are left with
 1807 \Sexpr{nrow(U[U\$Religion=="muslim",])} Muslims and
 1808 \Sexpr{nrow(U[U\$Religion=="non-muslim",])} non-Muslims. The
 1809 sample selections induced by the choice of variables and the
 1810 missing values in these variables lead to a loss of
 1811 \input{crossreference2.tex} percent of the
 1812 relevant sample of the original
 1813 data.\footnote{The variables written in capital letters
 1814 are created to ensure well--defined reference categories.
 1815 They are included in our regressions, but we cannot say
 1816 whether they are included in the regressions of
 1817 \citet{Bisin08}.}

1818

1819

1820 The sample means reported in \citet{Bisin08} seem to be
 1821 unweighted. Since data instructions says that the
 1822 data should always be weighted, tables \ref{tab:descriptive}
 1823 and \ref{tab:descriptive2} report weighted and unweighted
 1824 sample means before and after removal of non-availables and
 1825 the \citet{Bisin08} data.\footnote{See \citet{FNSEM93b}.}
 1826 Comparing means, the \citet{Bisin08} data seem to be
 1827 different from the original sample. The variables
 1828 \emph{Attitude Towards Inter Marriage} and
 1829 \emph{Importance of Racial Composition in Schools}
 1830 in \citet{Bisin08} data deviate largely from corresponding
 1831 averages in the original data. The deviation is extreme in
 1832 case of \emph{Importance of Racial Composition in Schools}.
 1833 The original sample has a mean for this variable that is
 1834 \Sexpr{100*round(mean(U.Original\$Importance.of.Racial.Composition.in.Schools[U.Orig
 inal\$Religion=="non-muslim"],na.rm=TRUE),2)}
 1835 percent for non-Muslims (compare with 33 percent in
 1836 \citet{Bisin08}) and
 1837 \Sexpr{100*round(mean(U.Original\$Importance.of.Racial.Composition.in.Schools[U.Orig
 inal\$Religion=="muslim"],na.rm=TRUE),2)}
 1838 percent for Muslims (compare with 65 percent in
 1839 \citet{Bisin08}). Due to this extremely skewed distribution.
 1840 it is hardly meaningful to run a regression on this
 1841 variable,

1842

1843 Notice that also the distribution of the variable
 1844 \emph{Importance of Religion} would be extremely skewed
 1845 using standard coding of this type of variables. Such a
 1846 coding would imply that religion is important when the
 1847 respondent answer ``Very Important'' and ``Fairly
 1848 Important'', to the question ``How important is religion to
 1849 the way you live your life?''.

1850

1851

1852 The sample means in our data after removing accumulated
 1853 missing values due to all variables in the estimations
 1854 deviate marginally in general from the original data.
 1855 The similarities here are partly due to the fact that the
 1856 statistics are based exactly on the same variable definition

1857 in our implementation. In some respects, the deviations are
 1858 larger. For further comparisons we refer to Tables
 1859 `\ref{tab:descriptive}` and `\ref{tab:descriptive2}`.

1860

1861 Due to the fact that the large majority of observations
 1862 from the original data are lost, the remaining sample is
 1863 likely to be contaminated with sample selection bias.
 1864 To compare the characteristics of the remaining sample
 1865 with the original sample says something about systematic
 1866 attrition with respect to observables. The sample selection
 1867 bias with respect to unobservables cannot, however,
 1868 be resolved.

1869

1870 `\section{Regression Results}`

1871 `\label{Regression}`

1872

1873

1874 `%</paper|techreport>`

1875 `<<Defining models for regressions, echo=FALSE>>=`

1876 `# Defining the regression model for`

1877 `# Importance of Religion`

1878 `# DO.NOT.SPEAK.WITH.YOUNGER, removed,`

1879 `# no observations for Muslims.`

1880 `# only two observations for non-Muslims`

1881 `# table(U$DO.NOT.SPEAK.WITH.YOUNGER, U$Religion)`

1882

```
1883 model1 <- {formula(Importance.of.Religion ~
1884   Age.at.Arrival + Female + Born.in.the.UK +
1885   Arranged.Marriage + Discrimination +
1886   Children + Years.Since.Arrival +
1887   No.British.Education + British.Basic.Education +
1888   British.Higher.Education + Foreign.Education +
1889   Employee+ Manager + Self.Employed + Unemployed+
1890   No.Parents + Parents.Physical.Contacts +
1891   Parents.Telephone.Calls + Parents.Letters +
1892   English.Spoken.at.Home.with.Older +
1893   English.Spoken.at.Home.with.Younger +
1894   English.Spoken.at.Work +
1895   English.Spoken.With.Friends +
1896   Household.Income +
1897   Discrimination.Own.Ethnicity +
1898   Ward.Density.Own.Ethnicity +
1899   Ward.Unemployment.Rate + area +
1900   factor(DO.NOT.SPEAK.WITH.OLDER)+
1901   factor(DO.NOT.SPEAK.AT.WORK) +
1902   factor(DO.NOT.SPEAK.WITH.FRIENDS))}
```

1903

1904

1905 `# Defining the model using the same regressors as`

1906 `# in Model 1 and Attitude.Towards.Inter.Marriage`

1907 `# as dependent variable`

1908 `model2 <- {update(model1,`

1909 `Attitude.Towards.Inter.Marriage ~ .)}`

1910

1911 `# Defining the model using the same regressors as`

1912 `# in Model 1 and Importance.of.Racial.Composition.in.Schools`

1913 `# as dependent variable`

1914 `model3 <- {update(model2,`

1915 `Importance.of.Racial.Composition.in.Schools ~ .)}`

1916 `@`

```

1917
1918 <<Regressions, echo=FALSE>>=
1919 library(sandwich,warn.conflicts=FALSE)
1920 library(lmtest,warn.conflicts=FALSE)
1921 Umuslim <- U[U$Religion=="muslim",]
1922 UNmuslim <- U[U$Religion!="muslim",]
1923
1924 ESTIMATE <- function(x,df)
1925     {fm <- lm(x, weights=weightis, data=df)
1926     rbind(coeftest(fm,
1927           vcov = vcovHC(fm, type ="HC1")),
1928           summary(fm)$adj.r.squared)}
1929
1930 result1.muslim <- ESTIMATE(model1, Umuslim)
1931 result2.muslim <- ESTIMATE(model2, Umuslim)
1932 result3.muslim <- ESTIMATE(model3, Umuslim)
1933
1934
1935 result1.non.muslim <- ESTIMATE(model1, UNmuslim)
1936 result2.non.muslim <- ESTIMATE(model2, UNmuslim)
1937 result3.non.muslim <- ESTIMATE(model3, UNmuslim)
1938
1939 @
1940
1941 %<*paper|techreport>
1942 We use linear probability models
1943 (LPM).\footnote{\citet{Bisin08} use probit estimations. Our
1944 attempts to use probit run into convergence problems. The
1945 convergence problems are severe for the model using
1946 \emph{Importance of Racial Composition in Schools} as
1947 dependent variable. Hence, our choice of LPM.
1948
1949 Another issue is that \citet{Bisin08} should have included
1950 dummy variables indicating religious affiliation:
1951 Christians, Sikhs and others in the non-Muslim category to
1952 check similarities and differences among non-Muslims as
1953 well. In this respect we follow their model specification.
1954
1955 Moreover, \citet{Bisin08} should have adjusted for within
1956 ward correlations. This might matter for their standard
1957 errors, which might be underestimated. In our case, with
1958 almost no significant results, this would not matter much.
1959 The variable is not available in the data set and we did not
1960 make much effort to obtain it.
1961 %</paper|techreport>
1962
1963
1964
1965 %<*techreport>
1966 All estimated models include 7 UK-region dummies, and
1967 the variables \code{DO.NOT.} \code{SPEAK.WITH.OLDER},
1968 \code{DO.NOT.SPEAK.AT.WORK}, and
1969 \code{DO.NOT.SPEAK.WITH.FRIENDS}. It turned out that the
1970 variable \code{DO.NOT.SPEAK.WITH.YOUNGER} is TRUE for few
1971 observations and cannot be included in the model.
1972 %</techreport>
1973 %<*paper|techreport>
1974 } Our results are
1975 presented in Tables \ref{ta:regressions} and
1976 \ref{ta:regressions2}. \citet{Bisin08} write that:

```

1977 `\begin{enumerate}`
1978 `\item` ``Muslims integrate less and more slowly than
1979 non-Muslims.'' (abstract, p. 445) and `\item` ``\dots there
1980 is no evidence that segregated neighbourhoods breed intense
1981 religious and cultural identities. On the contrary, \dots
1982 intense identities in our data are more prominent
1983 in relatively mixed neighbourhoods.'' (p. 446)
1984 (p. 446)
1985 `\end{enumerate}`
1986
1987 The first claim is based on their reported results
1988 concerning the variable `\emph{Years Since Arrival}`. In
1989 this way `\citet{Bisin08}` compare cohorts of Muslims and non
1990 Muslims and attempt to say something about the evolution of
1991 values over time. They do not follow individuals over time
1992 but nonetheless call these cohort differences ``Integration
1993 over time''. They report negative coefficients for
1994 `\emph{Years Since Arrival}`, but the estimates are smaller
1995 in absolute value for Muslims than for non-Muslims. In our
1996 case, the coefficients for `\emph{Years Since Arrival}`
1997 reported in Tables `\ref{ta:regressions}` and
1998 `\ref{ta:regressions2}` are insignificant in all cases except
1999 in the regression for `\emph{Importance of Religion}` for
2000 Muslims, where it is negative. This is opposite to what
2001 `\citet{Bisin08}` claim.
2002
2003 The second claim is based on their reported results
2004 concerning the variable `\emph{Ward Density Own Ethnicity}`.
2005 `\citet{Bisin08}` report negative and significant estimates
2006 for `\emph{Ward Density Own Ethnicity}` in all six
2007 specifications. Their negative coefficient for this variable
2008 would imply that ethnic minorities put more weight on
2009 religion, mind more about inter-ethnic marriage and have
2010 stronger taste for ethnically profiled schools, as we move
2011 from neighbourhoods (Wards) with high density of their own
2012 ethnicity to neighbourhoods where people from their own
2013 ethnicity are scarce. This is not at all what we find in our
2014 replication.
2015
2016 In our estimations, the estimated coefficients for this
2017 variable are all positive but far from significant. The
2018 P-values are
2019 `\Sexpr{round(result1.muslim["Ward.Density.Own.Ethnicity",4],2)}`,
2020 `\Sexpr{round(result2.muslim["Ward.Density.Own.Ethnicity",4],2)}`
2021 and
2022 `\Sexpr{round(result3.muslim["Ward.Density.Own.Ethnicity",4],2)}`
2023 for Muslims and
2024 `\Sexpr{round(result1.non.muslim["Ward.Density.Own.Ethnicity",4],2)}`,
2025 `\Sexpr{round(result2.non.muslim["Ward.Density.Own.Ethnicity",4],2)}`
2026 and
2027 `\Sexpr{round(result3.non.muslim["Ward.Density.Own.Ethnicity",4],2)}`
2028 for non-Muslims contradicting the `\citet{Bisin08}` results.
2029
2030 Inspecting the results presented in Tables
2031 `\ref{ta:regressions}` and `\ref{ta:regressions2}`, there are
2032 many similarities and few differences in the estimated
2033 coefficients for Muslims and non-Muslims. Our results are
2034 generally very different from results reported by
2035 `\citet{Bisin08}`. We are, however, doubtful whether it is
2036 possible to draw any reliable inference from these results

```
2037 due to great loss of observations and possible sample
2038 selection bias, together with the problem of endogeneity
2039 (also mentioned by \citet{Bisin08}).
2040
2041
2042
2043 \section{Concluding remarks}
2044 \label{Discussion}
2045
2046 The \citet{Bisin08} paper rests on fragile grounds. Our
2047 examination of the data using their variable definitions and
2048 the same set-up indicates that their claims about
2049 differences between Muslims and non-Muslims, and their
2050 conclusion that strong Religious/Ethnic identities are found
2051 in mixed neighbourhoods does not hold. There is no
2052 systematic relation between ethnic minorities' views on
2053 religion, inter-ethnic marriage or ethnic profile of schools
2054 and the density of their own ethnic minority in their
2055 neighbourhood. However, we hesitate to draw inference from
2056 these results since the great loss of observations
2057 ( $\input{crossreference2.tex}$  percent) implies that the
2058 remaining sample is most likely not representative.
2059
2060
2061
2062 %</paper|techreport>
2063
2064 %<*present>
2065 \begin{slide}{Conclusions}
2066 \vspace{\stretch{1}}
2067 \begin{itemize}
2068 \item The \citet{Bisin08} paper rests on fragile grounds.
2069 \item Their claims about differences between Muslims and
2070 non-Muslims, and their conclusion that strong
2071 Religious/Ethnic identities are found in mixed
2072 neighbourhoods does not hold.
2073 \item There is no
2074 systematic relation between ethnic minorities' views on
2075 religion, inter-ethnic marriage or ethnic profile of schools
2076 and the density of their own ethnic minority in their
2077 neighbourhood.
2078 \item However, we hesitate to draw inference from
2079 these results since the great loss of observations
2080 ( $\input{crossreference2.tex}$  percent) implies that the
2081 remaining sample is most likely not representative.
2082 \end{itemize}
2083 \vspace{\stretch{1}}
2084 \end{slide}
2085
2086 %</present>
2087
2088
2089
2090 %<*techreport>
2091 \section{Production notes}
2092 \label{sec:prodnotes}
2093
2094 To facilitate reproducibility and save others timely
2095 interpretations of what is done in this paper, we attempt
2096 to follow Literate Statistical Programming procedures. For
```



```

2157 %% Code for Table 1: Religious affiliation ...
2158 %%
2159 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
2160 <<Table 1,results=tex,keep.source=FALSE,echo=FALSE>>=
2161 # Defining variables to use in the table
2162 OnonMuslim      <- table(FNSEM$s7!="muslim")[2]
2163 names(OnonMuslim) <- "nonMuslim"
2164 SnonMuslim      <- table(U$Religion!="muslim")[2]
2165 names(SnonMuslim) <- "nonMuslim"
2166 # Defining the table header
2167 Header <- paste("Religious affiliation",
2168 "&\\multicolumn{1}{c}{(1)}",
2169   "&\\multicolumn{1}{c}{(2)}",
2170   "&\\multicolumn{1}{c}{(3)}",
2171   "&\\multicolumn{1}{c}{(4)}",
2172   "&\\multicolumn{1}{c}{(5)}",
2173   "&\\multicolumn{1}{c}{(6)}",
2174   "\\hline",
2175   "&\\multicolumn{2}{c}{Before}",
2176   "&\\multicolumn{2}{c}{After}",
2177   "&\\multicolumn{2}{c}{Bisin et al.}",
2178   "\\|",
2179   "&\\multicolumn{2}{c}{n=",length(U.Original$s7),"$"},
2180   "&\\multicolumn{2}{c}{n=",length(U$s7),"$"},
2181   "&\\multicolumn{2}{c}{n=5963$"},
2182   "\\cline{2-7}",
2183   "&\\multicolumn{1}{r}{\\#}&\\multicolumn{1}{r}{\\%}",
2184   "&\\multicolumn{1}{r}{\\#}&\\multicolumn{1}{r}{\\%}",
2185   "&\\multicolumn{1}{r}{\\#}&\\multicolumn{1}{r}{\\%}\\|",
2186   sep="")
2187 # Defining table footer
2188 Footer <- paste("\\hline All non-Muslims",
2189   "&",OnonMuslim,"&",
2190   round(100*OnonMuslim/table(FNSEM$s7=="na")[[1]],2),
2191   "&",SnonMuslim,"&",
2192   round(100*SnonMuslim/length(U$Religion),2),
2193   "&3594&",round(100*3594/(3594+2369),2),
2194   "\\hline",
2195   "\\multicolumn{7}{l}{\\multirow{2}{10.5cm}",
2196   "{\\footnotesize NOTE: The row names shows exactly ",
2197   "how the original data is coded, so that e.g., ",
2198   "`NA's' are true missing values whereas `na' is ",
2199   "coded as religious affiliation `na'. On the last ",
2200   "line non--Muslims are calculated excluding na and ",
2201   "NA.}}\\|",
2202   sep="")
2203 addtorow <- list()
2204 addtorow$pos <- list()
2205 addtorow$pos[[1]] <- 0
2206 addtorow$pos[[2]] <- 13
2207 addtorow$command <- c(Header,Footer)
2208 # Defining the data to use in the table
2209 DiffSample <- cbind(
2210   summary(FNSEM$s7),
2211   100*summary(U.Original$s7)/length(U.Original$s7),
2212   c(summary(U$s7),
2213     length(U$s7)-table(is.na(U$s7))[[1]]),
2214   c(100*summary(U$s7)/length(U$s7),
2215     100*(length(U$s7)-table(is.na(U$s7))[[1]])),
2216   matrix(list(c(),c()),round(2369,0),c(),c(),c(),

```

```

2217     c(),c(),c(),c(),c(),c(),c()),
2218     matrix(list(c(),c(),39.73,c(),c(),c(),c(),c(),
2219     c(),c(),c(),c(),c()))
2220 )
2221 # Taking dimnames from original data
2222 dimnames(DiffSample)[[1]] <- names(summary(U.Original$s7))
2223 # Generate the table
2224 TabDiffSample <- print(xtable(DiffSample,align=c("lrrrrrr"),
2225     caption=c("Religious affiliation (absolute (\\#) and
2226     relative (\\%) numbers), before (columns 1 and 2) and
2227     after (columns 3 and 4) removal of \\code{NA} compared
2228     with \\citet{Bisin08} (columns 5 and 6)."),
2229     label="ta:difffsample",
2230     digits=c(0,0,2,0,2,0,2)),
2231     table.placement = "h",
2232     caption.placement="top",
2233     include.colnames=FALSE,
2234     add.to.row=addtorow,
2235     hline.after=c(-1,0,11))
2236 @
2237
2238 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
2239 %%
2240 %% Code for Table 2-3: Weighted and weighted means ...
2241 %%
2242 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
2243
2244 <<Table 2-3,results=tex,keep.source=FALSE,echo=FALSE>>=
2245 # Generate sample with all NA removed
2246 SampleU <- subset(U,select=c(
2247     Importance.of.Religion,
2248     Attitude.Towards.Inter.Marriage,
2249     Importance.of.Racial.Composition.in.Schools,
2250     Age.at.Arrival,
2251     Age,
2252     Female,
2253     Born.in.the.UK,
2254     Arranged.Marriage,
2255     Discrimination,
2256     Children,
2257     Years.Since.Arrival,
2258     No.British.Education,
2259     British.Basic.Education,
2260     British.Higher.Education,
2261     Foreign.Education,
2262     Employee,
2263     Manager,
2264     Self.Employed,
2265     OUT.OF.LABOUR.FORCE,
2266     Unemployed,
2267     No.Parents,
2268     Parents.Physical.Contacts,
2269     Parents.Telephone.Calls,
2270     Parents.Letters,
2271     English.Spoken.at.Home.with.Older,
2272     DO.NOT.SPEAK.WITH.OLDER,
2273     English.Spoken.at.Home.with.Younger,
2274     DO.NOT.SPEAK.WITH.YOUNGER,
2275     English.Spoken.at.Work,
2276     DO.NOT.SPEAK.AT.WORK,

```

```
2277     English.Spoken.With.Friends,
2278     DO.NOT.SPEAK.WITH.FRIENDS,
2279     Household.Income,
2280     Ward.Density.Own.Ethnicity,
2281     Ward.Unemployment.Rate,
2282     Discrimination.Own.Ethnicity
2283 ))
2284 # Generate sample with no NA removed
2285 SampleU.Original <- subset(U.Original,select=c(
2286     Importance.of.Religion,
2287     Attitude.Towards.Inter.Marriage,
2288     Importance.of.Racial.Composition.in.Schools,
2289     Age.at.Arrival,
2290     Age,
2291     Female,
2292     Born.in.the.UK,
2293     Arranged.Marriage,
2294     Discrimination,
2295     Children,
2296     Years.Since.Arrival,
2297     No.British.Education,
2298     British.Basic.Education,
2299     British.Higher.Education,
2300     Foreign.Education,
2301     Employee,
2302     Manager,
2303     Self.Employed,
2304     OUT.OF.LABOUR.FORCE,
2305     Unemployed,
2306     No.Parents,
2307     Parents.Physical.Contacts,
2308     Parents.Telephone.Calls,
2309     Parents.Letters,
2310     English.Spoken.at.Home.with.Older,
2311     DO.NOT.SPEAK.WITH.OLDER,
2312     English.Spoken.at.Home.with.Younger,
2313     DO.NOT.SPEAK.WITH.YOUNGER,
2314     English.Spoken.at.Work,
2315     DO.NOT.SPEAK.AT.WORK,
2316     English.Spoken.With.Friends,
2317     DO.NOT.SPEAK.WITH.FRIENDS,
2318     Household.Income,
2319     Ward.Density.Own.Ethnicity,
2320     Ward.Unemployment.Rate,
2321     Discrimination.Own.Ethnicity
2322 ))
2323
2324 # Set Age.at.Arrival and Years.Since.Arrival
2325 # for those Born.in.the.UK to NA. This is only for the
2326 # sample statistics.
2327
2328
2329 is.na(SampleU.Original$Age.at.Arrival) <-
2330     SampleU.Original$Born.in.the.UK==1
2331 is.na(SampleU.Original$Years.Since.Arrival) <-
2332     SampleU.Original$Born.in.the.UK==1
2333
2334
2335 # A function to compute weighted mean
2336 # for a variable in a data set (x), with
```

```

2337 # weights (w) and column (n).
2338 WMEAN <- function (x,W,n) {
2339     weighted.mean(x[!is.na(x[,n]),n], W[!is.na(x[, n])])}
2340
2341
2342 # Weights to use in computing weighted means for Muslims
2343 Muslim.weight <-
2344     U$weightis[U$Religion=="muslim"]
2345 Orig.Muslim.weight <-
2346     U.Original$weightis[U.Original$Religion=="muslim"]
2347 Non.Muslim.weight <-
2348     U$weightis[U$Religion=="non-muslim"]
2349 Orig.Non.Muslim.weight <-
2350     U.Original$weightis[U.Original$Religion=="non-muslim"]
2351
2352
2353 # Create a matrix to gather means
2354 # Columns 1:5 for Muslims:
2355 # 1:2 weighted means, 3:4 unweighted, 5 Bisin et al.
2356 # Columns 6:10 for non-Muslims:
2357 # 6:7 weighted means, 8:9 unweighted, 10 Bisin et al.
2358
2359 Means.Table <- matrix(,ncol(SampleU),10)
2360 rownames(Means.Table) <- colnames(SampleU)
2361
2362 # Compute weighted means in columns 1:36
2363 for (i in 1:ncol(SampleU.Original)) {Means.Table[i,1]<-
2364     WMEAN(SampleU.Original[U.Original$Religion=="muslim",],
2365           Orig.Muslim.weight, i)}
2366
2367 # Compute weighted means in columns 1:36
2368 for (i in 1:ncol(SampleU)) {Means.Table[i,2]<-
2369     WMEAN(SampleU[U$Religion=="muslim",] , Muslim.weight, i)}
2370
2371
2372
2373 # Unweighted Means for Muslims
2374 Means.Table[,3] <-
2375     mean(SampleU.Original[U.Original$Religion=="muslim",],
2376         na.rm=TRUE)
2377 Means.Table[,4] <-
2378     mean(SampleU[U$Religion=="muslim",], na.rm=TRUE)
2379
2380
2381 # Bisin et al. Muslims
2382 Means.Table[,5] <-
2383     c(0.79,0.70,0.65,39.18, NA, 0.47,0.21,0.22,0.17,2.17,
2384       26.43,0.81,0.06,0.08,0.25,0.38,0.02,0.09,NA,0.19,
2385       0.34,3.05,3.38,0.67,0.03,NA,0.20,NA,0.19,NA,0.22,
2386       NA,200.74,0.15,16.57,0.21)
2387
2388
2389 # Compute weighted means SampleU.Original
2390 for (i in 1:ncol(SampleU.Original)) {Means.Table[i,6]<-
2391     WMEAN(SampleU.Original[U.Original$Religion=="non-muslim",],
2392           Orig.Non.Muslim.weight, i)}
2393
2394 # Compute weighted means SampleU
2395 for (i in 1:ncol(SampleU)) {Means.Table[i,7]<-
2396     WMEAN(SampleU[U$Religion=="non-muslim",],

```

```

2397     Non.Muslim.weight, i)}
2398
2399
2400 # Unweighted Means for non-Muslims
2401 Means.Table[,8] <-
2402     mean(SampleU.Original[U.Original$Religion=="non-muslim"],
2403          na.rm=TRUE)
2404 Means.Table[,9] <-
2405     mean(SampleU[U$Religion=="non-muslim"],na.rm=TRUE)
2406
2407
2408
2409 # Bisin et al. non-Muslims
2410 Means.Table[,10] <-
2411 c(0.42,0.37,0.33,42.57,NA,0.48,0.28,0.12,0.19,1.68,
2412    27.08,0.52,0.13,0.16,0.29,0.59,0.04,0.14,NA,0.08,
2413    0.32,3.87,4.74,0.37,0.08,NA,0.25,NA,0.27,NA,0.27,
2414    NA,330.26,0.11,12.60,0.18)
2415
2416
2417
2418 # Replace "." with " " in variable names to appear in the table
2419 rownames(Means.Table) <- gsub("\\.", " ", rownames(Means.Table))
2420
2421
2422
2423 # Creating the table i LaTeX code
2424 print(xtable(Means.Table[1:17,],align=c("lrrrrrrrrrr"),
2425            caption="Weighted and Unweighted Means for Muslims and
2426            non--Muslims before and after removal of \\code{NA}
2427            compared with \\citet{Bisin08}.",
2428            label="tab:descriptive"),
2429            table.placement = "p",
2430            floating.environment = "sidewaystable",
2431            caption.placement="top",
2432            add.to.row=list(pos=list(0),
2433            command=paste(
2434            "&\\multicolumn{5}{c}{Muslim}",
2435            "&\\multicolumn{5}{c}{Non-Muslim}\\\\\\\\\\\\\\\\hline",
2436            "&\\multicolumn{2}{c}{W e i g h t e d}",
2437            "&\\multicolumn{3}{c}{U n w e i g h t e d}",
2438            "&\\multicolumn{2}{c}{W e i g h t e d}",
2439            "&\\multicolumn{3}{c}{U n w e i g h t e d}\\\\\\\\",
2440            "&\\multicolumn{1}{c}{Before}",
2441            "&\\multicolumn{1}{c}{After}",
2442            "&\\multicolumn{1}{c}{Before}",
2443            "&\\multicolumn{1}{c}{After}",
2444            "&\\multicolumn{1}{c}{Bisin}",
2445            "&\\multicolumn{1}{c}{Before}",
2446            "&\\multicolumn{1}{c}{After}",
2447            "&\\multicolumn{1}{c}{Before}",
2448            "&\\multicolumn{1}{c}{After}",
2449            "&\\multicolumn{1}{c}{Bisin}\\\\\\\\",
2450            "&&&&et al.&&&&et al.\\\\\\\\\\\\\\\\hline",
2451            sep="")),
2452            include.colnames=FALSE,
2453            hline.after=c(-1,0,NULL))
2454 # Table 3: Continuation of Table 2
2455 print(xtable(Means.Table[18:nrow(Means.Table),],
2456            align=c("lrrrrrrrrrr"),

```

```

2457     caption="Table 2 continued. Weighted and Unweighted
2458     Means for Muslims and non-Muslims before and after
2459     removal of \code{NA} compared with \citet{Bisin08}.",
2460     label="tab:descriptive2"),
2461     table.placement = "p",
2462     floating.environment = "sidewaystable",
2463     caption.placement="top",
2464     add.to.row=list(pos=list(0),
2465     command=paste(
2466     "&\multicolumn{5}{c}{Muslim}",
2467     "&\multicolumn{5}{c}{Non-Muslim}\hline",
2468     "&\multicolumn{2}{c}{Weighted}",
2469     "&\multicolumn{3}{c}{Unweighted}",
2470     "&\multicolumn{2}{c}{Weighted}",
2471     "&\multicolumn{3}{c}{Unweighted}\hline",
2472     "&\multicolumn{1}{c}{Before}",
2473     "&\multicolumn{1}{c}{After}",
2474     "&\multicolumn{1}{c}{Before}",
2475     "&\multicolumn{1}{c}{After}",
2476     "&\multicolumn{1}{c}{Bisin}",
2477     "&\multicolumn{1}{c}{Before}",
2478     "&\multicolumn{1}{c}{After}",
2479     "&\multicolumn{1}{c}{Before}",
2480     "&\multicolumn{1}{c}{After}",
2481     "&\multicolumn{1}{c}{Bisin}\hline",
2482     "&&&&et al.&&&&et al.\hline",
2483     sep="")),
2484     include.colnames=FALSE,
2485     hline.after=c(-1,0,NULL))
2486 @
2487 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
2488 %%
2489 %% Code for Table 4-5: Regression results ...
2490 %%
2491 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
2492
2493 <<Table 4 prepare, echo=FALSE>>=
2494 MakeTable <- function(x, n=2) {
2495     # take coef() objects and makes table
2496     Stars <- ifelse(x[,4]<0.05, paste("*"), paste(""))
2497     names(Stars) <- NULL
2498     Estimates <- format(round(x[,1],n))
2499     Std.errors <- format(round(x[,2],n))
2500     x <- cbind(Estimates,Std.errors, Stars)
2501     Table <- matrix(,2*nrow(x),1)
2502     Table[seq(1,2*nrow(x)-1,2),1] <-
2503     paste(x[,1],x[,3], sep="")
2504     Table[seq(2,2*nrow(x),2),1] <-
2505     paste("(",x[,2],")",sep="")
2506     Xname <- unlist(strsplit(unlist(strsplit(rownames(x),
2507     "TRUE")), "yes"))
2508     Xname <- gsub("\\.", " ", Xname)
2509     Xnames <- matrix(,2*nrow(x),1)
2510     Xnames[seq(1,2*nrow(x)-1,2),1] <- Xname
2511     Xnames[seq(2,2*nrow(x),2),1] <- ""
2512     strsplit(Xnames, "TRUE")
2513     strsplit(Xnames, "yes")
2514     rownames(Table) <- Xnames
2515     Table
2516 }

```

```

2517
2518
2519 Table.Reg <-
2520   cbind(
2521     rbind(MakeTable(result1.muslim [1:28,]),
2522           format(round(result1.muslim[39,1],2), nsmall=2)),
2523     rbind(MakeTable(result1.non.muslim[1:28,]),
2524           round(result1.non.muslim[40,1],2)),
2525     rbind(MakeTable(result2.muslim [1:28,]),
2526           round(result2.muslim[39,1],2)),
2527     rbind(MakeTable(result2.non.muslim[1:28,]),
2528           round(result2.non.muslim[40,1],2)),
2529     rbind(MakeTable(result3.muslim [1:28,]),
2530           round(result3.muslim[39,1],2)),
2531     rbind(MakeTable(result3.non.muslim[1:28,]),
2532           round(result3.non.muslim[40,1],2))
2533   )
2534
2535 @
2536 <<Print Table 4, results=tex, echo=FALSE>>=
2537
2538
2539 rownames(Table.Reg)[nrow(Table.Reg)] <- "Adjusted R-square"
2540 varnames <- rownames(Table.Reg)
2541 rownames(Table.Reg) <- NULL
2542 Table.Reg.dat <- data.frame(varnames, Table.Reg)
2543 print(xtable(Table.Reg.dat[1:28,],align=c("llccccc"),
2544             caption=paste("Regression Results for Muslims",
2545                             nrow(Umuslim),"and non-Muslims", nrow(UNmuslim),
2546                             "to be compared with Table 2 in \\citet{Bisin08}. ",
2547                             "Heteroskedasticity corrected (HC1) Standard Errors ",
2548                             "are in parentheses. P-values $< 0.05$ are ",
2549                             "marked with *.",sep=" "),
2550             label="ta:regressions"),
2551             table.placement = "p",
2552             floating.environment = "sidewaystable",
2553             caption.placement="top",
2554             include.rownames=FALSE,
2555             include.colnames=FALSE,
2556             add.to.row=list(pos=list(0),
2557                             command=paste(
2558                               "&\\multicolumn{2}{c}{Importance of}",
2559                               "&\\multicolumn{2}{c}{Inter Ethnic }",
2560                               "&\\multicolumn{2}{c}{Ethnic Composition}\\\\\\\\",
2561                               "&\\multicolumn{2}{c}{Religion}",
2562                               "&\\multicolumn{2}{c}{Marriage}",
2563                               "&\\multicolumn{2}{c}{of Schools}",
2564                               "\\\\\\\\\\cline{2-7}\\\\\\\\",
2565                               "&\\multicolumn{1}{c}{Muslims}",
2566                               "&\\multicolumn{1}{c}{non--Muslims}",
2567                               "&\\multicolumn{1}{c}{Muslims}",
2568                               "&\\multicolumn{1}{c}{non--Muslims}",
2569                               "&\\multicolumn{1}{c}{Muslims}",
2570                               "&\\multicolumn{1}{c}{non--Muslims}\\\\\\\\\\\\\\hline",
2571                               sep=""))
2572             ,hline.after=c(-1,0,NULL))
2573 # Table 5: Continuation of Table 4
2574 Footer <- paste("\\hline",
2575                "\\multicolumn{7}{l}{\\multirow{3}{20cm}",
2576                "\\footnotesize NOTE: All estimated models include ",

```

```

2577     "7 UK-region dummies, and ",
2578     "the variables \\code{DO.NOT.SPEAK.WITH.OLDER}, ",
2579     "\\code{DO.NOT.SPEAK.AT.WORK}, and ",
2580     "\\code{DO.NOT.SPEAK.WITH.FRIENDS}. It turned out ",
2581     "that the variable \\code{DO.NOT.SPEAK.WITH.YOUNGER} ",
2582     "is TRUE for few observations and cannot be ",
2583     "included in the model.}}\\\\",
2584     sep="")
2585
2586
2587 Header <- paste(
2588     "&\\\\multicolumn{2}{c}{Importance of}",
2589     "&\\\\multicolumn{2}{c}{Inter Ethnic }",
2590     "&\\\\multicolumn{2}{c}{Ethnic Composition}\\\\",
2591     "&\\\\multicolumn{2}{c}{Religion}",
2592     "&\\\\multicolumn{2}{c}{Marriage}",
2593     "&\\\\multicolumn{2}{c}{of Schools}",
2594     "\\\\\\\\\\cline{2-7}\\\\",
2595     "&\\\\multicolumn{1}{c}{Muslims}",
2596     "&\\\\multicolumn{1}{c}{non--Muslims}",
2597     "&\\\\multicolumn{1}{c}{Muslims}",
2598     "&\\\\multicolumn{1}{c}{non--Muslims}",
2599     "&\\\\multicolumn{1}{c}{Muslims}",
2600     "&\\\\multicolumn{1}{c}{non--Muslims}\\\\\\\\\\hline",
2601     sep="")
2602
2603
2604 print(xtable(Table.Reg.dat[29:57,],align=c("llccccc"),
2605     caption=paste("Table 4 continued. Regression Results ",
2606     "for Muslims", nrow(Umuslim)," and non-Muslims",
2607     nrow(UNmuslim), "to be compared with Table 2 in ",
2608     "\\cite{Bisin08}. Heteroskedasticity corrected ",
2609     "(HC1) Standard Errors are in parentheses. P-values ",
2610     "$< 0.05$ are marked with *.", sep=" "),
2611     label="ta:regressions2"),
2612     table.placement = "p",
2613     floating.environment = "sidewaystable",
2614     caption.placement="top",
2615     include.rownames=FALSE,
2616     include.colnames=FALSE,
2617     add.to.row=list(pos=list(0,29),
2618     command=c(Header,Footer))
2619     ,hline.after=c(-1,0,NULL))
2620 @
2621
2622 %</paper|techreport>
2623
2624 %<*paper|techreport|present>
2625 \\end{document}
2626 %</paper|techreport|present>
2627 Here we create the statistics that are cross--references to
2628 in the abstract and the introduction. It is the relative
2629 number of observations that remains after missing values
2630 are removed and its complement:
2631
2632 <<Cross-references,echo=FALSE,keep.source=FALSE>>=
2633 RelativeSampleSize <- round(100*nrow(U)/nrow(FNSEM),
2634     digits=0)
2635 BisinRelativeSampleSize <- round(100*5963/nrow(FNSEM),
2636     digits=0)

```

```

2637 write(BisinRelativeSampleSize,file="crossreference0.tex")
2638 write(100-RelativeSampleSize,file="crossreference2.tex")
2639 write(nrow(U.Original),file="crossreference3.tex")
2640 write(nrow(U),file="crossreference4.tex")
2641
2642 @
2643
2644
2645 The following code writes the DOCSTRIP commands to the file
2646 araietal.ins:
2647
2648 <<LaTeX, echo=FALSE>>=
2649 write(paste(
2650     "\\input docstrip.tex",
2651     "\\keepsilent",
2652     "\\askforoverwritefalse",
2653     "\\nopreamble",
2654     "\\nopostamble",
2655     "\\generate{",
2656     "\\file{araietal_paper.tex}{",
2657     "    \\from{araietal_source.tex}{paper}}",
2658     "\\file{araietal_techreport.tex}{",
2659     "    \\from{araietal_source.tex}{techreport}}",
2660     "\\file{araietal_present.tex}{",
2661     "    \\from{araietal_source.tex}{present}}",
2662     "}",
2663     "\\endbatchfile",
2664     "\\n",sep="\\n"),file="araietal.ins")
2665 @
2666
2667 <<BibTex, echo=FALSE>>=
2668 write(paste(
2669 "@Manual{FNSEM,",
2670 "author      = {FNSEM},"",
2671 "title       = {P1312 Fourth National Study of Ethnic Minorities. {P}project
instructions},"",
2672 "note        = {Project 3685},"",
2673 "organization = {Social and Community Planning Research}, ",
2674 "address      = {London},"",
2675 "year         = {1993},"",
2676 "url          = {http://www.data-archive.ac.uk/doc/3685/mrdoc/pdf/a3685uab.pdf},"",
2677 "},"",
2678 "\\n",
2679 "@Manual{FNSEM93b,",
2680 "author       = {FNSEM},"",
2681 "title        = {UK Data Archive Data Dictionary}, ",
2682 "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}, ",
2683 "organization = {Social andCommunity Planning Research}, ",
2684 "address       = {London},"",
2685 "year          = {1993},"",
2686 "},"",
2687 "\\n",
2688 "@Article{Bisin08,",
2689 "author        = {Alberto Bisin and Eleonora Patacchini and Thierry Verdier and
Yves Zenou},"",
2690 "title         = {Are Muslim Immigrants Different in Terms of Cultural
Integration?},"",
2691 "journal       = {Journal of the European Economic Association},"",

```

```
2692 "pages      = {445--456}," ,
2693 "volume     = {6}," ,
2694 "year       = {2008}" ,
2695 "}" ,
2696 "\n" ,
2697 "@Manual{Rcore," ,
2698 "title      = {R: A Language and Environment for Statistical Computing}," ,
2699 "author     = {{R Development Core Team}}}," ,
2700 "organization = {R Foundation for Statistical Computing}," ,
2701 "address    = {Vienna, Austria}," ,
2702 "year       = {2008}," ,
2703 "note       = {{ISBN} 3-900051-07-0}," ,
2704 "url        = {http://www.R-project.org}" ,
2705 "}" ,
2706 "\n" ,
2707 "@Unpublished{Koenker07," ,
2708 "author     = {Koenker, Roger and Zeileis, Achim}," ,
2709 "year       = {2007}," ,
2710 "title      = {Reproducible Econometric Research (A Critical Review of the State
of the Art)}," ,
2711 "note       = {Report 60, Department of Statistics and Mathematics,
Wirtschaftsuniversität Wien, Research Report Series}," ,
2712 "url        =
{http://epub.wu-wien.ac.at/dyn/virlib/wp/eng/mediate/epub-wu-01_c75.pdf?ID=epub-wu-01
_c75}" ,
2713 "}" ,
2714 "\n" ,
2715 "@Unpublished{Araietal08a," ,
2716 "author     = {Arai, Mahmood and Karlsson, Jonas and Lundholm, Michael}," ,
2717 "year       = {2008}," ,
2718 "title      = {On Fragile Grounds: A replication of \emph{Are Muslim immigrants
different in terms of cultural
2719 integration?}}}," ,
2720 "note       = {Unpublished manuscript}" ,
2721 "}" ,
2722 "\n" ,
2723 "@Unpublished{Araietal08b," ,
2724 "author     = {Arai, Mahmood and Karlsson, Jonas and Lundholm, Michael}," ,
2725 "year       = {2008}," ,
2726 "title      = {On Fragile Grounds: {A} replication of \emph{Are Muslim
immigrants different in terms of cultural integration? {T}echnical
documentation}}}," ,
2727 "note       = {Unpublished manuscript}" ,
2728 "}" ,
2729 "\n" ,
2730 "@Unpublished{Leisch02," ,
2731 "author     = {Friedrich Leisch}," ,
2732 "title      = {Sweave User Manual}," ,
2733 "year       = {2002}," ,
2734 "url        = {http://www.ci.tuwien.ac.at/~leisch/Sweave/}" ,
2735 "}" ,
2736 "\n" ,
2737 "@Manual{Berthoud," ,
2738 "title      = {Fourth National Survey of Ethnic Minorities, 1993-1994 [computer
file]}," ,
2739 "author     = {Berthoud, R.G. and Modood, T. and Smith, P. and Prior, G.}," ,
2740 "publisher  = {UK Data Archive [distributor]}," ,
2741 "address    = {Colchester, Essex: UK}," ,
2742 "year       = {1997}," ,
2743 "note       = {SN: 3685}," ,
```

```
2744 "url          =
      {http://www.data-archive.ac.uk/doc/3685%5Cmrdoc%5CUKDA%5CUKDA_Study_3685_Informatio
      n.htm}",
2745 "}",
2746 sep="\n"),file="araietal.bib")
2747 @
2748
2749 <<Extracting R code,results=hide>>=
2750 Stangle("araietal_source.Rnw")
2751 @
2752
2753 <<Script to make LaTeX files,echo=FALSE,silent=TRUE,results=hide>>=
2754 write(paste(
2755 "system(\"latex araietal.ins\")",
2756 "system(\"pdflatex araietal_techreport.tex\")",
2757 "system(\"bibtex araietal_techreport\")",
2758 "system(\"pdflatex araietal_techreport.tex\")",
2759 "system(\"pdflatex araietal_techreport.tex\")",
2760 "system(\"latex araietal_paper.tex\")",
2761 "system(\"bibtex araietal_paper\")",
2762 "system(\"latex araietal_paper.tex\")",
2763 "system(\"latex araietal_paper.tex\")",
2764 "system(\"dvips araietal_paper.dvi\")",
2765 "system(\"ps2pdf araietal_paper.ps\")",
2766 "system(\"latex araietal_present.tex\")",
2767 "system(\"dvips -o araietal_present1.ps -pp 1:14 araietal_present.dvi\")",
2768 "system(\"dvips -o araietal_present2.ps -pp 7:11 -x 1095 araietal_paper.dvi\")",
2769 "system(\"dvips -o araietal_present3.ps -p 20 araietal_present.dvi\")",
2770 "system(\"gs -q -r2400 -dNOPAUSE -dBATCH -sDEVICE=pdfwrite
      -sOutputFile=araietal_present.pdf -f araietal_present1.ps araietal_present2.ps
      araietal_present3.ps\")",
2771 "system(\"pdflatex araietal_paper.tex\")",
2772 sep="\n"),file="script.R")
2773 @
2774
```