Read and trim files for original 800 HHs

April 11, 2019
17:47

Seiro Ito

Contents

?

This file reads data from a list data_read_in_a_list_with_baseline_patched.rds, merge all non-roster files with admin-roster, and saves in c:/data/GUK/analysis/save/ReadFilesMergeAdminRoster/.

I Read files

I.1 Read from a list

In reading raw files, I added ID information (./ID/ID_Updated_received_from_Abu.dta) to all pages. I further added HH ID information from the admin file.

Define shock variables. FloodInRd1 is reported flood damage in code_1 in round 1.

Description of data:

ad Administrative data: Up to [-24, 48] months after first loan disbursement. This file has not been used in read_cleaned_data.rnw.

ros roster to condition the initial status prior to participation.

sch Schooling panel with attrition. Aged 6-18 in rd1. Enrolled={0,1} is defined for children aged 6-18 in rd1 by referencing to currently_enrolled and age information.

ass Assets. Household assets (houses, durables) and productive assets (machines, tools).

lvo Livestock holding.

lab Labour incomes.

far Farming revenues (no costs reported).

con Household consumption. Food expenditure asks both bought and consumed volumes and prices. We impute consumption values by using median prices. All quantity is set to annualised quantity.

shk Shocks. Merged with all other files.

I.2 Sample selection and treament assignment

I.2.1 Merge admin and roster files

Number of meetings in recorded in admin file.

```
1999
  48
```

Tabulation of arms with ar.0. There are 220 NAs which will be filled in with RCT_village.dta with ar, arAll in the next subsection.

```
traditional       large large grace         cow        <NA>
        485         464         467         487         220
```

Some HHs in admin file are not found in roster. Drop from data.

```
 [1]  9807042003  9807042011  9807042514  9807042706  9807042710  9807054106
 [7]  9807054304  9807054520  9807064605  9807064607  9807064612  9807064617
[13]  9807064619  9807065207  9807065208  9807065212  9807065306  9807065307
[19]  9807065313  9807065315  9807065316  9807065319  9807075702  9807085904
[25]  9807085914  9807086106  9807086107  9807106513  9807106517  9807106518
```

```
[31]  9807126819  9807126820  9807127103  9807127105  9807127106  9807127108
[37]  9807137203  9807137204  9807137206  9807137217  9807137218  9808169816
[43]  9907065108  9907075402  9907075405  9907075406  9907075407  9907075410
[49]  9907075411  9907075413  9907075418  9907075419  9907075420 98081710308
[55] 98081710317 99070210905 99070210906 99070211813 99070310702 99070311402
[61] 99070311403 99070311405 99070311407 99070311408 99070311411 99070311412
[67] 99070311415 99070311416 99070311419 99070311501 99070311502 99070311505
[73] 99070311507 99070311508 99070311509 99070311511 99070311513 99070311517
[79] 99070311520 99070712701 99070712703 99070712704 99070712707 99070712708
[85] 99070712710 99070712713 99070712714 99070712716 99070712720 99071010811
[91] 99071010813 99071010814 99071010819 99081711206 99081711207 99081711208
```

```
 [1]  9807042003  9807042011  9807042514  9807042706  9807042710  9807054106
 [7]  9807054304  9807054520  9807064605  9807064607  9807064612  9807064617
[13]  9807064619  9807065207  9807065208  9807065212  9807065306  9807065307
[19]  9807065313  9807065315  9807065316  9807065319  9807075702  9807085904
[25]  9807085914  9807086106  9807086107  9807106513  9807106517  9807106518
[31]  9807126819  9807126820  9807127103  9807127105  9807127106  9807127108
[37]  9807137203  9807137204  9807137206  9807137217  9807137218  9808169816
[43]  9907065108  9907075402  9907075405  9907075406  9907075407  9907075410
[49]  9907075411  9907075413  9907075418  9907075419  9907075420 98081710308
[55] 98081710317 99070210905 99070210906 99070211813 99070310702 99070311402
[61] 99070311403 99070311405 99070311407 99070311408 99070311411 99070311412
[67] 99070311415 99070311416 99070311419 99070311501 99070311502 99070311505
[73] 99070311507 99070311508 99070311509 99070311511 99070311513 99070311517
[79] 99070311520 99070712701 99070712703 99070712704 99070712707 99070712708
[85] 99070712710 99070712713 99070712714 99070712716 99070712720 99071010811
[91] 99071010813 99071010814 99071010819 99081711206 99081711207 99081711208
```

This results in reduction in observations with 48 meetings.

```
1903
  48
```

Errors in creditstatus. Correct to Yes.

```
      hhid EverSaved TotalRepaid   Mstatus     Mship creditstatus DisDate1
1: 8169303      TRUE       16300 oldMember oldMember           No     <NA>
2: 8169305      TRUE       16800 oldMember oldMember           No     <NA>
3: 8169306      TRUE       16300 oldMember oldMember           No     <NA>
4: 8169316      TRUE       16424 oldMember oldMember           No     <NA>
```

Errors in Mstatus. Correct to oldMember. (No corresponding entry in arAll because it is data only for members.)

```
      hhid CumRepaid CumNetSaving TradGroup       Date   DisDate1 creditstatus
1: 7137220        NA           NA      <NA>       <NA> 2013-11-01          Yes
2: 7137220      1750            0   planned 2014-10-01 2013-11-01          Yes
3: 7137220      4250          385   planned 2015-11-01 2013-11-01          Yes
       Mship    Mstatus
1: oldMember iRejection
2: oldMember iRejection
3: oldMember iRejection
```

There are 22 members (oldMember in Mstatus), 20 members (newGroup in Mstatus), 3 members (iReplacement in Mstatus) who did not borrow but only saved. This is identified by DisDate1 == NA & EverSaved & creditstatus == No (not NAs, because they are offered and declined).

```
          Arm         DisDate1  creditstatus          Mstatus
 traditional:45   Min.   :NA   Yes: 0        gErosion    : 0
 large      : 0   1st Qu.:NA   No :45        gRejection  : 0
```

```
large grace: 0   Median :NA                iRejection  : 0
cow        : 0   Mean   :NA                iReplacement: 3
                 3rd Qu.:NA                newGroup    :20
                 Max.   :NA                oldMember   :22
                 NA's   :45
           Mship
oldMember     :22
newMember     :23
quitMembership: 0
```

There are also members who were offered membership but never took up. This is identified by DisDate1 == NA & !EverSaved & !EverRepaid.

```
    DisDate1   EverSaved       EverRepaid      creditstatus         Mstatus
Min.   :NA    Mode :logical   Mode :logical   Yes:  0      gErosion    : 80
1st Qu.:NA    FALSE:379       FALSE:379       No :379      gRejection  :140
Median :NA                                                 iRejection  :159
Mean   :NA                                                 iReplacement:  0
3rd Qu.:NA                                                 newGroup    :  0
Max.   :NA                                                 oldMember   :  0
NA's   :379
```

Create BorrowerStatus to indicate these guys (DisDate1 == NA & EverSaved & creditstatus == No) as a pure saver. 379 entries with DisDate1 == NA & !EverSaved & !EverRepaid are people who quit so set as quit membership.

```
             BorrowerStatus
Mstatus       borrower pure saver quit membership
  gErosion           0           0              80
  gRejection         0           0             140
  iRejection         0           0             159
  iReplacement     112           3               0
  newGroup         388          20               0
  oldMember       1199          22               0
```

In arAll, nonmembers (gRejection, gErosion) are not included.

```
             Mship
Mstatus       oldMember newMember quitMembership
  gErosion            0         0              0
  gRejection          0         0              0
  iRejection          0         0            159
  iReplacement        0       115              0
  newGroup            0       408              0
  oldMember        1221         0              0
```

Need to merge in 2 steps: Merge admin (time-invariant) with roster with hhid as a key, then merge to admin (time-variant [e.g., OtherRepaid, OtherNetSaving, OtherMisses, CumOtherMisses, CumRepaid, CumEffectiveRepayment, CumNetSaving, CumPlannedInstallment, CumLoanAmount, CumOtherRepaid, CumOtherNetSaving, CumMisses, CumRepaidRate, CumEffectiveRepaidRate, RMOtherNetSaving, RMOtherRepaid]) with hhid, Year, Month as keys. This is because there are 8398 non-matching cases if we merge using Year, Month of IntDate in roster data and Year, Month of Date in admin data. This is inevitable because survey precedes the first meeting of borrowers: The admin data starts from 2013-05-01 while survey data starts from 2011-10-09 and rd 1 ends at 2013-10-12 for oldMembers with the median date 2012-10-20. Below gives Year, Month in roster data in rd 1 with no match in admin data.

```
 2011-October  2011-November   2012-January   2012-October  2012-November
            6              1             19           1146            327
2012-December 2013-September   2013-October   2014-January   2014-October
           79              6             19             12             83
```

```
 2014-November  2014-December  2015-November  2015-December   2016-January
            43             36            111             40             26
  2017-January  2017-February     2017-March     2017-April          NA-NA
            44             97             17             17             21
```

After 2014, it is mostly drop out members who do not match with admin data because they do not attend the meeting.

```
table0(ar00[is.na(MonthsElapsed) & MemNum == 1 & Year ≥ 2014,
  Mgroup])
```

```
   continued    drop outs    new group replacements
          78           381           58             9
```

No additional match if matching only with Year.

```
                FALSE TRUE
YearMonthMatch  2055 5958
YearMatch       2055 5958
```

In roster + admin (base: roster): Tabulate hhid observations by survey round and RArm before supplementing with AssignOriginal and VArm. Note: 220 observations with NA are also pointed in read_cleaned_data.rnw and are going to be dealt with in the next subsection.

```
      RArm
survey traditional large large grace cow <NA>
     1         485   464         467 487  220
     2         472   445         447 446  173
     3         472   448         452 453  168
     4         465   444         447 444  114
```

I.2.2 Merge village level info

ar: ar.1 + vr (RCT_village.dta)

I use arm VArm from village level information. Tabulation of AssignOriginal against VArm shows complementarity so I can use one variable to fill in NAs in another.

```
              VArm
AssignOriginal traditional large large grace  cow <NA>
   traditional        1244     0           0    0  650
   large                 0  1423           0    0  378
   large grace           0     0        1437    0  376
   cow                   0     0           0 1631  199
   <NA>                418   158          40   59    0
```

Tabulation of RArm after supplementing with AssignOriginal and VArm.

```
ar[is.na(RArm) & !is.na(AssignOriginal), RArm := AssignOriginal]
ar[is.na(RArm) & !is.na(VArm), RArm := VArm]
```

```
      RArm
survey traditional large large grace cow
     1         605   504         507 507
     2         585   485         447 466
     3         582   487         452 472
     4         540   483         447 444
```

Below is what is supplemented from VArm of village level information to the 220 NAs.

```
                 RArm
BorrowerStatus    traditional large large grace cow
  borrower                  0     0           0   0
  pure saver                0     0           0   0
  quit membership         120    40          40  20
```

arA: arAll (admin data as base + roster) + vr (village randomisation)

```

1903
  48
```

Tabulation of BorrowerStatus in arA at round 1.

```
                traditional large large grace cow total
borrower                387   452         445 415  1699
pure saver               45     0           0   0    45
quit membership          53    12          22  72   159
total                   485   464         467 487  1903
```

Tabulation of Mstatus in arA at round 1.

```
             traditional large large grace cow total
gErosion               0     0           0   0     0
gRejection             0     0           0   0     0
iRejection            53    12          22  72   159
iReplacement          39     8          11  57   115
newGroup             166    96          96  50   408
oldMember            227   348         338 308  1221
total                485   464         467 487  1903
```

Tabulation of Mstatus in ar at round 1.

```
             traditional large large grace cow total
gErosion              40     0          20  20    80
gRejection            80    40          20   0   140
iRejection            53    12          22  72   159
iReplacement          39     8          11  57   115
newGroup             166    96          96  50   408
oldMember            227   348         338 308  1221
total                605   504         507 507  2123
```

adw3 idfu[adw2]: admin data adw2 + idfu (arm information)

ad0 Selected columns of adw3.

Base: roster.

ar.0 adbase[ros]: ros (33223, 37) + invariant portion of admin data ad0 (1999, 8).

ar.1 adrest[ar.0]: ar.0 (33223, 43)+ variable portion of admin data ad0 (95952, 42).

ar vr[ar.1]: ar.1 (33223, 85) + vr (RCT_village.dta) (80, 4), resulting in (33223, 89). Number of individuals: 2123.

Base: admin. This has a smaller number of individuals because admin data do not include individuals who left the group.

ar.00 ros.00W[ad0]: ad0 (95952, 49) + ros.00W (survey round info) (2123, 5).

arAll ros.0[ar.00]: ar.00 (admin data with survey round info) (95952, 50) + ros.0 (roster only with first observed round) (2123, 11).

arA vr[arAll]: arAll (admin data as base + roster) (91344, 65) + vr (village randomisation) (80, 4), resulting in (91344, 70). Number of individuals: 1903.

I.3 Merge admin-roster with other files

I.3.1 Choosing sample in admin-roster

In roster + admin 2: Keep if Mstatus includes strings old, iRej, gEro, gRej, & TradGroup does not include strings tw (relaxing DisDate1 is before 2015-01-01). This the data used in this note. This also shows a lower attrition rate for large arm.

	traditional	large	large	grace	cow	total
1	400	400		400	400	1600
2	327	384		342	366	1419
3	324	386		348	366	1424
4	287	382		343	342	1354

Create o1600 to indicate the original 1600 HHs. Tabulation of total observations in roster by o1600 and survey.

o1600 \ survey	1	2	3	4
0	2101	2510	2543	2457
1	6532	5817	5843	5420

Tabulation of total observations in roster by o1600 and survey after restricting to 1 obs per HH.

o1600 \ survey	1	2	3	4
0	523	611	616	607
1	1600	1372	1377	1307

Tabulation for arA. It has survey == 5 which are meetings after the rd 4 interview. arA has fewer observations per meeting than ar when only using 1 obs per rd,

	traditional	large	large	grace	cow	total
1	12	21		30	49	112
2	167	343		342	346	1198
3	165	341		338	335	1179
4	165	343		342	342	1192
5	128	265		235	269	897

but more observations per round because there are multiple meetings per round.

	traditional	large	large	grace	cow	total
1	445	845		967	1886	4143
2	3054	6197		6221	6156	21628
3	2220	4650		4607	4596	16073
4	2379	5074		5073	5095	17621
5	302	514		412	507	1735

Pick 800 o800 by referring to JDS data.

Tabulation of o800 for ar.

	traditional	large	large	grace	cow	total
1	200	200		200	200	800
2	190	191		172	190	743
3	188	193		174	190	745
4	168	192		171	177	708

Tabulation of o800 for arA.

	traditional	large	large grace	cow	total
1	9	10	14	18	51
2	134	171	172	180	657
3	133	170	169	175	647
4	132	173	171	177	653
5	102	133	120	138	493

	traditional	large	large grace	cow	total
gErosion	0	0	0	0	0
gRejection	0	0	0	0	0
iRejection	31	9	13	37	90
iReplacement	0	0	0	0	0
newGroup	0	0	0	0	0
oldMember	109	171	167	153	600

arA is used in saving and repayment regressions.

In ar, there are 114 cases of group rejections in GroupStatus classified as individual rejections in Mstatus. Overwrite Mstatus with GroupStatus in these cases, which results in the below:

	GroupStatus			
Mstatus	accepted	erosion	group	rejection
gErosion	0	80		0
gRejection	0	0		140
iRejection	159	0		0
iReplacement	115	0		0
newGroup	408	0		0
oldMember	1221	0		0

	traditional	large	large grace	cow
accepted	485	464	467	487
erosion	40	0	20	20
group rejection	80	40	20	0
total	605	504	507	507

In ar, as one can see below, gRejection is more frequent in traditional and large, while there is none in cow. traditional, cow have more frequent iRejection. So traditional was disliked both at group and individual levels, large was disliked as a group, cow was disliked at an individual level, and large grace were well received at both group and individual levels. This indicates attractiveness of a grace period at least at the group level, and a large cash form (over small cash or in-kind) at the individual level.

	traditional	large	large grace	cow
gErosion	40	0	20	20
gRejection	80	40	20	0
iRejection	53	12	22	72
iReplacement	39	8	11	57
newGroup	166	96	96	50
oldMember	227	348	338	308
total	605	504	507	507

	traditional	large	large grace	cow
gErosion	0.07	0.00	0.04	0.04
gRejection	0.13	0.08	0.04	0.00
iRejection	0.09	0.02	0.04	0.14
iReplacement	0.06	0.02	0.02	0.11
newGroup	0.27	0.19	0.19	0.10
oldMember	0.38	0.69	0.67	0.61
total	1.00	1.00	1.00	1.00

In ar, for o800 we have:

```
             traditional large large grace cow
gErosion              20     0          10  10
gRejection            40    20          10   0
iRejection            31     9          13  37
iReplacement           0     0           0   0
newGroup               0     0           0   0
oldMember            109   171         167 153
total                200   200         200 200
```

Contrast it with arA:

```
Error: <text>:1:15: unexpected ';'
1: arA[, tee := 1;
                  ^
```

Save roster-admin data to c:/data/GUK/analysis/save/ReadFilesMergeAdminRoster/.

```
saveRDS(ar, paste0(pathsaveHere, "RosterAdminData.rds"))
saveRDS(arA, paste0(pathsaveHere, "AllMeetingsRosterAdminData.rds"))
fwrite(ar, paste0(pathsaveHere, "RosterAdminData.prn"), sep = "\t", quote = F)
fwrite(arA, paste0(pathsaveHere, "AllMeetingsRosterAdminData.prn"), sep = "\t", quote = F)
```

Schooling.

I.3.2 Attach variables from admin-roster to other files

Attach RArm, Arm, TradGroup, Mem, ObPattern, AttritIn, o1600, o800, Mstatus, BorrowerStatus, creditstatus, povertystatus, RMvalue.repay, RMvalue.NetSaving, RMOtherNetSaving, RMOtherRepaid, HHsize, HeadLiteracy, IntDate, DisDate1 from ar.

```
vartoattach ← c("RArm", "Arm", "TradGroup", "Mem",
  "ObPattern", "AttritIn", "o1600", "o800",
  "Mstatus", "BorrowerStatus",
  "creditstatus", "povertystatus", "RMvalue.repay",
  "RMvalue.NetSaving", "RMOtherNetSaving", "RMOtherRepaid",
  "HHsize", "HeadLiteracy", "IntDate", "DisDate1")
dfiles ← c("ass", "s1", "lvo", "lvp", "lab", "far", "con", "shk")
for (j in 1:length(dfiles)) {
  dd ← get(dfiles[j])
  dd[, groupid := as.integer(as.numeric(as.character(gid)))]
  dd[, gid := NULL]
  dd[, Year :=  as.numeric(format(as.Date(IntDate), "%Y"))]
  dd[, Month := as.character(format(as.Date(IntDate), "%B"))]
  dd[Year ≤ 2010, Year := Year + 10]
  # drop all variables in each page before copying from ar0
  dd[, (vartoattach) := NULL]
  setorder(dd, groupid, hhid, survey, Year, Month)
  setkey(dd, groupid, hhid, survey)
  if (j < length(dfiles)) dd ← ar0[dd]
  assign(dfiles[j], dd)
}
```

Check number of HHs in assets by o1600:

```
table(ass[, .(creditstatus, survey, o1600)])
```

```
, , o1600 = 0

            survey
```

```
creditstatus    1    2    3    4
         Yes  478  588  593  586
         No    23   23   23   21

, , o1600 = 1

             survey
creditstatus    1    2    3    4
         Yes 1196 1051 1058 1043
         No   399  319  319  264
```

```
tb ← table(ass[o800 == 1, .(survey, creditstatus)])
cbind(tb, total = apply(tb, 1, sum))
```

```
  Yes  No total
1 597 199   796
2 584 158   742
3 586 159   745
4 578 130   708
```

```
#table0(ass[o1600 == 0L, .(creditstatus, survey)])
```

Check number of HHs in schooling by o1600:

```
table(s1[, .(Schooling, survey, o1600)])
```

```
, , o1600 = 0

             survey
Schooling       1    2    3    4
  primary0512 457  427  361  202
  junior1315  114  129  140  204
  high1618     80   94   93  111

, , o1600 = 1

             survey
Schooling       1    2    3    4
  primary0512 1389  911  659  322
  junior1315   326  279  427  499
  high1618     216  198  179  225
```

```
tb ← table(s1[o800 == 1, .(survey, Schooling)])
cbind(tb, total = apply(tb, 1, sum))
```

```
  primary0512 junior1315 high1618 total
1         695         159      110   964
2         483         147      105   735
3         344         230       90   664
4         165         264      115   544
```

Check number of o800 HHs in ar:

```
ar[, tee := as.integer(1:.N), by = .(hhid, survey)]
tb ← table0(ar[tee == 1 & o800 == 1L, .(survey, RArm)])
cbind(tb, total = apply(tb, 1, sum))
```

```
  traditional large large grace cow total
1         200   200         200 200   800
2         190   191         172 190   743
3         188   193         174 190   745
4         168   192         171 177   708
```

Check number of o800 HHs in arA:

```
arA[, tee := as.integer(1:.N), by = .(hhid, survey)]
tb ← table0(arA[tee == 1 & o800 == 1L, .(survey, RArm)])
cbind(tb, total = apply(tb, 1, sum))
```

```
  traditional large large grace cow total
1           9    10          14  18    51
2         134   171         172 180   657
3         133   170         169 175   647
4         132   173         171 177   653
5         102   133         120 138   493
```

Number of observations differ between ar and arA because the latter does not include rejecters.

```
             traditional large large grace cow traditional large large grace
gErosion              20     0          10  10           0     0           0
gRejection            40    20          10   0           0     0           0
iRejection            31     9          13  37          31     9          13
iReplacement           0     0           0   0           0     0           0
newGroup               0     0           0   0           0     0           0
oldMember            109   171         167 153         109   171         167
             cow
gErosion       0
gRejection     0
iRejection    37
iReplacement   0
newGroup       0
oldMember    153
```

Original 800 households in arA (members only).

```
[1] TRUE
```

```
             EverRepaid
Mstatus        TRUE
  gErosion        0
  gRejection      0
  iRejection      0
  iReplacement    0
  newGroup        0
  oldMember     600
```

What is relevant in estimation is observations by LoanYear, total of 600.

```
        RArm
LoanYear traditional large large grace cow
       1         109   171         167 153
       2         109   171         167 153
       3         109   171         167 153
       4         109   171         167 153
```

If we restrict to planned in TradGroup, number of observation becomes 576.

```
        RArm
LoanYear traditional large large grace cow
       1          85   171         167 153
       2          85   171         167 153
       3          85   171         167 153
       4          85   171         167 153
```

ObPattern in original 800.

```
          RArm
ObPattern traditional large large grace cow
     0111           0     2           3   1
     1000           1     5           1   1
     1010           0     1           0   0
     1011           0     0           0   0
     1100           0     1           3   2
     1110           1     0           3   3
     1111         107   162         157 146
```

BorrowerStatus pattern in original 800.

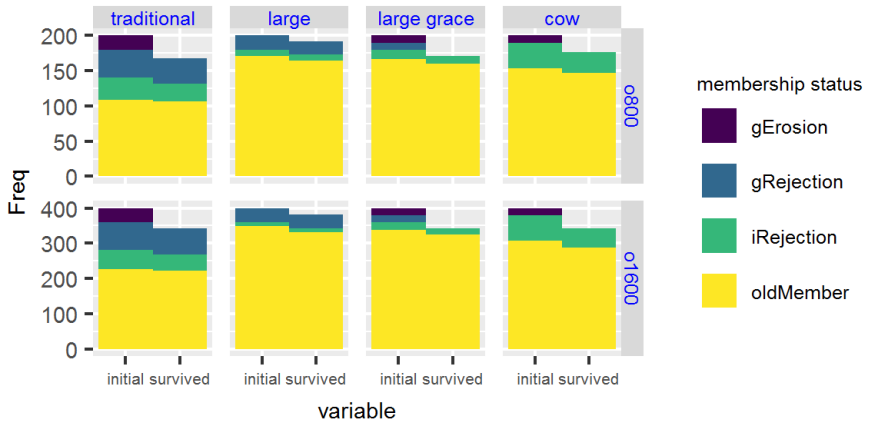
```
                RArm
BorrowerStatus   traditional large large grace cow
  borrower                 109   171         167 153
  pure saver                 0     0           0   0
  quit membership            0     0           0   0
```

BorrowerStatus pattern in original 800 with only planned in TradGroup.

```
                RArm
BorrowerStatus   traditional large large grace cow
  borrower                  85   171         167 153
  pure saver                 0     0           0   0
  quit membership            0     0           0   0
```

Below tabulates attrition pattern in ar for 800 and 1600 households.

FIGURE 1: ATTRITION AND MEMBERSHIP STATUS AMONG ORIGINAL 800 AND 1600 HOUSEHOLDS



Source: Survey data.

Note: Top panel: Membership status and respective non-attrition in o800. Bottom panel: Membership status and respective non-attrition in o1600.

Save all data in c:/data/GUK/analysis/save/ReadFilesMergeAdminRoster/.

```
fwrite(s1, paste0(pathsaveHere, "RosterAdminSchoolingData.prn"), sep = "\t", quote = F)
fwrite(ass, paste0(pathsaveHere, "AssetAdminData.prn"), sep = "\t", quote = F)
fwrite(lvo, paste0(pathsaveHere, "LivestockAdminData.prn"), sep = "\t", quote = F)
fwrite(lvp, paste0(pathsaveHere, "LivestockProductsAdminData.prn"), sep = "\t", quote = F)
fwrite(lab, paste0(pathsaveHere, "LabourIncomeAdminData.prn"), sep = "\t", quote = F)
fwrite(far, paste0(pathsaveHere, "FarmRevenueAdminData.prn"), sep = "\t", quote = F)
fwrite(con, paste0(pathsaveHere, "ConsumptionAdminData.prn"), sep = "\t", quote = F)
fwrite(shk, paste0(pathsaveHere, "Shocks.prn"), sep = "\t", quote = F)
```