# FAMILY NETWORKS AND SCHOOL ENROLMENT: EVIDENCE FROM A RANDOMIZED SOCIAL EXPERIMENT

MANUELA ANGELUCCI  GIACOMO DE GIORGI
MARCOS A.RANGEL  IMRAN RASUL

MAY 2008

# 1 Appendix

## 1.1 Descriptive Evidence on Surnames

Table A1 provides descriptive evidence on each surname type – the paternal and maternal surnames of the head ($F1$, $f1$) and spouse ($F2$, $f2$). For both head and spouse, there are fewer paternal than maternal surnames reported. As Figure 1 shows, this reflects the fact that under a patronymic naming convention, paternal surnames have a greater survival rate across generations. There are 1696 different paternal surnames reported by heads ($F1$), lower than for the other types of surname including those reported as the spouse's paternal surname ($F2$). This is both because the patronymic naming convention implies spouse's paternal surnames have lower survival rates across generations than those of male heads of household, and also be partly due to spouses moving into the 506 villages in the data from villages outside the evaluation sample.

The second row shows that the majority of surnames are mentioned at least twice. For each surname type, the most frequent surname covers around 9% of households, and the half the households have one of the 50 most frequent surnames for each surname type. The third row shows the probability of two randomly matched households having the same surname type is close to zero, and the expected number of households with the same head's paternal surname is 13.3. This is higher than the expected number of households with the same spouse's paternal surname, again suggestive of women moving into *Progresa* villages from other locations.[1]

[1] These population values are calculated as follows for any given surname type. Let $n_i$ denote the number of households with surname $i$ and let $N$ denote the number of households that report some surname of the given type. The probability, without replacement, that two randomly chosen households have surname $i$ is then $P_i =$

The next two rows report the same information but at the village level. The probability (without replacement) of two randomly chosen households in the village having the same surname is orders of magnitude larger than in the population. Hence households are not randomly allocated by surname type into villages. On the other hand, the fact that the expected number of households in the village with the same surname is smaller than in the population implies households do not perfectly sort into villages by surname either.[2]

The final row sheds light on the degree of sorting of households into villages by surname type. This is measured by an odds ratio, defined as the ratio of the probability that two randomly chosen households from the same village have the same surname, divided by the probability that two randomly chosen households from *Progresa* villages have the same surname. This odds ratio suggests households are, for example, 356 times more likely to match within a village on their head's paternal surname than if they were randomly allocated by this surname across villages.[3]

## 1.2 Measurement Error in Extended Family Links

There are a number of potential forms of measurement error in the surnames data that can be checked for. The first arises from the convention that women change their paternal surname to their husband's paternal surname at the time of marriage. To address this concern, we note that the precise wording of the question specifically asks respondents to name the paternal and maternal surname of each household member. Furthermore, in only 5.8% of households is the spouse's maternal surname recorded to be the same as her husband's paternal surname. This provides an upper bound on the extent to which measurement error of this form is occurring.

Second, if the male head is the respondent, he may not recall his wife's maternal surname and simply replace it with her paternal surname. This may occur because his children only inherit his wife's paternal surname. Reassuringly, this problem occurs in only 4.9% of households. A final

$\left(\frac{n_i}{N}\right) \cdot \left(\frac{n_i-1}{N-1}\right)$, and the expected number of households in the population with name $i$ is $E_i = n_i.\left(\frac{N-1}{N}\right)$. The values reported in Table A1 are the averages of $P_i$ and $E_i$ over all surnames $i$.

[2] These village values are calculated as follows for any given surname type. Let $n_{iv}$ denote the number of households with surname $i$ in village $v$ and $n_v$ denote the number of households that report some surname of the type in village $v$. The probability, without replacement, that two randomly chosen households in the village have surname $i$ is then $p_{iv} = \left(\frac{n_{iv}}{n_v}\right) \cdot \left(\frac{n_{iv}-1}{n_v-1}\right)$, and the expected number of households in the village with name $i$ is $e_{iv} = n_{iv}.(\frac{n_v-1}{n_v})$. The values reported in Table A1 are the weighted averages of $p_{iv}$ and $e_{iv}$ over all villages $v$, where the weights are $\frac{n_{iv}}{n_v}$. These weights account for the same name being reported to different extents across villages. The expected number of matches in the village is based on only one surname, and so provides an upper bound on the total number of extended family links our matching algorithm actually defines.

[3] This odds ratio is calculated as follows for any given surname type. We first take the weighted average of $p_{iv}$ over all names $i$ where the weights are $\frac{n_v}{N}$. These weights take account of the fact that if two households are drawn from the population at random, they are more likely to come from a larger village. Denote this weighted probability as $\widetilde{p}_i$. The reported odds ratio is then given by $\frac{\widetilde{p}_i}{P_i}$.

circumspect case is households in which the paternal and maternal surnames of both the head and spouse are all reported to be the same. This occurs for 1.6% of households, although the figure drops to .5% if we exclude households with the most common surname in the data.[4]

Some forms of measurement error however cannot be addressed. The first arises from any remaining typos in surnames. Second, there may be two identical families in the village who share the same paternal and maternal surnames of head and spouse but are genuinely unrelated. The matching algorithm then assigns the number of family links to be double what they actually are. A check for the severity of this problem is based on the following intuition. By definition, household $i$ cannot have parental links to more than two other households (the parent's of the head and the parent's of the spouse), conditional on the parents not being present within the household. This is true for 97% of households using our matching algorithm. Third, consider a scenario in which a women's brother marries someone with the same maternal surname as himself. Then the woman's niece will be identified as her sister and although the households are within the same family network, the strength of their tie may be inferred to be stronger than it actually is.

### 1.3 External Validity of the Extended Family Links: MxFLS Data

To provide external validity to the constructed data on extended family links in the *Progresa* data, we present similar information from an alternative data set that was collected in a comparable economic environment and time period. The *Mexican Family Life Survey* (MxFLS), collected in 2001, provides information on the number of each type of link, by head and spouse, that are still alive in *any* location, not just the same village. This data set therefore provides an upper bound on what should be recorded as family links in the *Progresa* data, in which we only construct links in the same village. In addition, we exploit information from the household roster in the MxFLS to also construct the number of family links inside the household, by each type of family link, and for the head of household and his spouse separately. To make the MxFLS data comparable, we restrict the sample to couple headed households that reside in locations with less than 2500 inhabitants in states that are also covered in the *Progresa* data. There are 580 such households.[5]

Table A2 reports the findings from the MxFLS. The number of family links to parents, children and siblings outside the household and located anywhere, are greater than those we construct using surnames data within the village from the *Progresa* data. The fact that more parents of the spouse are alive is likely to be driven by spouses being younger than their husbands. Moreover, the

[4] There are no differences in the incidence of these potential errors between treatment and control villages.

[5] As discussed in Section 3, one restriction on the matching algorithm used in the *Progresa* data is that we are unable to identify links to parental households if only one of the parents is alive. To ensure the MxFLS data is therefore comparable, we do not include information from couple headed households that report only having a single parent alive in another household. There are no such concerns for parental links defined inside the household.

differences between husbands and spouses in the number of parents and siblings are less dramatic in the MxFLS data, presumably because these statistics refer to family links in any location and so are unaffected by the geographic mobility of women at the time of marriage.

The comparison of family links within the household is also informative. Here the number of each type of family is similar to that found in *Progresa*, although the number of children is slightly lower. This may be driven by differences in the age of respondents in the two data sets – the age of spouses is 40.5 (43.5) in the *Progresa* (MxFLS) data. Heads and spouses are also more educated in the MxFLS data – the mean years of schooling for heads (spouses) in MxFLS is 3.91 (3.46) in comparison to 2.77 (2.27) in *Progresa*. These differences would explain the lower numbers of children in the household in the MxFLS data if more educated couples have lower fertility rates.[6] Moreover, it remains the case that in the MxFLS data as in *Progresa*, the number of family links of the head inside the household are greater than those of the spouse.

## 1.4 Correlates of Extended Family Links

Before exploiting information of extended family networks to explain household behavior, a useful stepping stone is to first establish the correlates of the extended family being present in the same village. This focuses attention on the econometric concerns stemming from the endogenous formation of family networks. Moreover, this analysis also provides further supportive evidence on the accuracy of the constructed extended family links.

A number of mechanisms drive the presence of extended family members, such as the need for insurance and the choice of marriage partners for children [Rosenzweig and Stark 1989], the value of services provided by social networks [Munshi and Rosenzweig 2005], inheritance of land and other household assets [Foster 1993], and the nature of household production [Foster and Rosenzweig 2002]. Our aim here is not to replicate such analyses, but to identify correlations between the presence of extended family ties and three classes of observable characteristics.

First, the age of the head and spouse should be negatively correlated with the likelihood their parents are in close proximity, and positively correlated with the probability of having adult children in the village. These correlations are somewhat mechanical as they depend primarily on the life cycle rather than on economic mechanisms. Second, there can be a positive correlation between wealth or land ownership and the presence of an extended family because – (i) wealthier family dynasties may have higher fertility and lower mortality rates; (ii) landed households both have more need for and can support larger family sizes; (iii) wealthier families may also be more likely to own land – as rural land markets are typically missing, the ability to inherit land, or to

[6] This underestimates the true difference in average years of education of couples between the two data sets because in the MxFLS, years of schooling are top coded at 12.

acquire land specific human capital, may lead adult children to be more likely to remain within the village than otherwise. A third mechanism driving extended family structures is the need to insure against risk. This leads to the formation of networks of related families with negatively correlated shocks, the strategic marriage of daughters into families with less correlated shocks, and migration of some family members to other locations.

To shed light on these channels we estimate a conditional logit regression where the dependent variable, $L_{jh}$, is a dummy equal to one if extended family link type-$j$ exists for household $h$ in the village, and zero otherwise. We consider the correlates of each of the specific family links that we identify, as well as on whether household $h$ has *any* family connections in the village. For each link type, $L_{jh}$, we control for the ages of the head and spouse, and dummy variables for whether they are literate and speak an indigenous language. At the household level, we control for whether the household owns its home, whether any land is owned, whether any member of the household temporarily migrated in the last year, the eligibility status of the household, the household poverty index, and household size at baseline.[7]

We group the conditional logit regression by village to take account of differences across villages that drive the formation of extended family networks. For example, villages may vary in the riskiness of their economic environments, altering the need for households to insure each other and therefore potentially causing alternative patterns of extended family networks to form. Standard errors are clustered by village, we report log odds ratios so that tests of significance relate to the log odds being significantly different from one. All continuous variables are divided by their standard deviation so the corresponding coefficients can be interpreted as the effect of a one standard deviation change in the continuous variable.

The results, reported in Table A3, highlight the following. First, the mechanical correlations with age are as expected with older heads and spouses being significantly less likely to have their parents outside the household and resident in the village, and significantly more likely to have their adult children in other households in the village. Older heads and spouses are more likely to have brothers present and less likely to have sisters present, presumably because, as highlighted in Section 3, women move village at the time of marriage. Second, literate heads and spouses are more likely to have their parents present. If such correlations persist across generations, then parents that educate their children increase the likelihood their children remain geographically proximate, other things equal.

Third, home and land ownership are positively correlated with the likelihood that children and

[7] We experimented with other specifications before settling on this set of controls. For example, we do not control for years of education because it is highly correlated with literacy – 89% (90%) of heads (spouses) have no formal schooling if they are illiterate. We focus on temporary rather than permanent migration because the proportion of households that report any members permanently migrating in the five years prior to 1997 is only 3.3%.

siblings reside in the same village, other things equal. The coefficients are of similar magnitude for brothers and sisters as well as for the adult children of the head and spouse (not shown). This pattern of coefficients is consistent both with inheritance norms in rural Mexico that do not favor men over women, and with a dynastic wealth effect such that wealthier families accumulate greater assets and have higher rates of fertility. In contrast, households in which at least one member has temporarily migrated in the last year – 18.5% of all households – are not more or less likely to have extended family links present.

Fourth, although there is a slight positive correlation between the household poverty index and the presence of adult children, there is no discontinuous effect of eligibility on the presence of any extended family ties. This aids the interpretation of the econometric evidence we provide. Whether the head and spouse speak an indigenous language also does not predict the presence of extended family ties. This is again reassuring because the unconditional number of extended family ties, for each type of tie, are no different between indigenous and non indigenous households.

Fifth, households that have a greater number of individuals within them are also significantly more likely to have a greater number of extended family members residing within the same village. This may be due to persistent differences in fertility levels within the same family dynasty across generations. Alternatively, the presence of extended family members may reduce the costs of having and raising children because extended family members are able to supply of time, labor, and other resources to the household.

A comparison across the columns is also informative. For example, the controls have a differential effect on the likelihood that the parents of the head or spouse are present. This is in line with the earlier evidence suggesting the process that drives the presence of parents are very different for the head and his spouse. In contrast, most of the controls have similar effects on the likelihood of brothers or sisters of the head and spouse being present.

In summary, the final column shows that connected and isolated households differ on a range of observable characteristics that drive the presence of extended family. In the empirical analysis it will therefore be important to both condition on these observables, and to allow household responses to *Progresa* to also vary with them. Hence the analysis sheds light on whether there exists a differential effect of being embedded within a family network or not, over and above potentially heterogeneous effects of characteristics that predict the existence of family links.

## 1.5 Robustness Checks on the Baseline Estimates

We present a series of robustness checks on the main finding in Table 5 that only households embedded within extended family networks respond to *Progresa*. The first series of checks, presented in Table A4, relate to concerns over the information on surnames and the matching algorithm.

We first address concerns over measurement error in the recorded surnames data that can lead to erroneous inference on the presence of extended family members. Column 1 shows our baseline result to be robust to dropping households with any of three potential types of measurement error in their surnames – (i) the spouse's maternal surname is the same as their husband's paternal surname; (ii) the spouse's paternal and maternal surnames are the same; (iii) the paternal and maternal surnames of both the head and spouse are all very similar.[8]

The second check addresses the concern that in larger villages, the matching algorithm is more likely to spuriously link two households that happen to have the same paternal and maternal surnames of head and spouse but are genuinely unrelated. Column 2 shows the baseline results to be robust to dropping villages in the top quartile of the village size distribution, as measured by the number of households in the village. Moreover the point estimate on the $TTE^1$ is larger than the baseline estimate. This may reflect a downward bias in $TTE^1$ because isolated households we previously being ascribed to be connected.

A third concern with the matching algorithm is that it actually measures the intrinsic value of surnames rather than having anything inherently to do with extended family links. For example, individuals with the most frequent surnames are both most likely to be found to have extended family members present, and may belong to family dynasties that have different unobservables that cause them to respond differently to *Progresa* than households with less common surnames. We address this issue in two ways.

First, we randomly reassign households in our baseline sample to another village within the same municipality and then rerun our matching algorithm based on the surname matches in these neighboring villages. We then explore whether our main results capture the effects of true extended family links that are present in the same village, or merely capture the effects of having more or less frequent surname combinations *per se*. The result in Column 3 shows there are no heterogeneous responses to *Progresa* on the basis of these surname based links in neighboring villages – the difference in the $TTEs$ is not significantly different from zero.[9]

Second, we estimate whether the frequency of paternal surnames predicts fertility levels at baseline. If for example some surnames are more frequent because those dynasties have lower

[8] The results are also robust to dropping households with the most common paternal surname of either the head or spouse in the village.

[9] Three points are of note. First, we only reassign the subset of 6227 households in the baseline sample. If we were to reassign all 22,553 households then the newly constructed family ties would be more likely to capture the actual family ties originally used. Second, we reassign households to other villages within the same municipality because as discussed in relation to Table A1, households sort across geographic locations by surnames. Doing so, 59% of households are found to be connected. Randomly reassigning households to any other village in the data would however dramatically reduce the likelihood any household is constructed to have family links. Finally, there are 115 municipalities in the data, the median municipality contains six villages, and we drop municipalities that only contain one village.

mortality rates or are better able to insure against income shocks, then we expect them to have higher fertility, other things equal. We estimate an OLS specification analogous to our baseline specification where the left hand side variable is the number of children aged 0 to 16 in the household at baseline, and we control for the share of households with the same paternal surname of the head and spouse, as well as the previously described controls. The result in Column 4a shows the relative frequency of paternal surnames is uncorrelated to baseline fertility levels. This result is robust to focusing attention to only the most common paternal surnames, namely those shared by at least .5% of the population (Column 4b).

The second series of checks, presented in Table A5, relate to a number of remaining concerns. First, there may be unobserved village level characteristics that drive both the presence of isolated households and their differential response to *Progresa*. We address this in two ways. First, we estimate the specification with interactions and additionally control for village characteristics in $Z_i$. These characteristics are the number of village households, the share of eligible households, the village marginality index, and the share of households that report being affected by any natural shocks from October 1998 to November 1999. Column 2 shows the baseline estimates to be robust to allowing household responses to *Progresa* to also vary by these village characteristics.

An alternative way to address this concern is to randomly reassign each household the family links of another household in the same village. This method leaves the likelihood of any given household being connected in a village to be the same as in the original data, as well as leaving unchanged the overall share of connected households in each village. If our previous measure of being connected or isolated were merely capturing some village level phenomenon, we would not expect the results to differ according to whose family ties any given household were assigned. Reassuring the result in Column 2 shows this not to be the case – when households are assigned to be connected or isolated at random, then both $TTE^1$ and $TTE^0$ are significantly greater than zero, and the difference between them is not different from zero.

The third check relates to time varying household characteristics that drive enrolment. Of particular is the fact that households are subject to economic shocks that cause them to take their children out of school [Jacoby and Skoufias 1997]. To address this we additionally control for whether the household reports being affected been any type of natural shock from October 1998 to November 1999, and allow responses to vary depending on this report. The result in Column 3, shows the results to be robust to the inclusion of such time varying household shocks.

The fourth check relates to the underlying identifying assumption that there are no spillover effects from treatment to control villages. We restrict our sample to households in villages that are below the median distance (5km) from any health facility, as recorded in May 1999. The result in Column 4 shows that within such villages, where concerns over spillover effects are perhaps greatest, the signs, significance, and magnitude of the baseline estimates continue to hold. An alternative subset of villages in which concerns over spillover effects may be particularly acute are those villages in which there are no secondary or middle schools present. Given that children resident in such locations attend secondary schools outside their own village, these children may be particularly likely to be in schools with children from both treatment and control villages present. Reassuringly, the result in Column 5 shows the previous parameter estimates to be robust to restricting the analysis to this subset of villages.

The final check directly addresses the assumption that extended family networks are not endogenously changing over time in response to the program. To address this we use data from the marital history module collected in May 1999 that explicitly asked spouses about whether their parents were present in the village or not. We use this information to reconstruct extended family ties to parents and hence to reconstruct whether households are connected or isolated. The result in Column 6 shows the previous estimates to be robust this redefinition of extended family links.

# References


[1] FOSTER.A (1993) "Household Partition in Rural Bangladesh", *Population Studies* 47: 97-114.

[2] FOSTER.A AND M.ROSENZWEIG (2002) "Household Division and Rural Economic Growth", *Review of Economic Studies* 69: 839-69.

[3] JACOBY.H.G AND E.SKOUFIAS (1997) "Risk, Financial Markets, and Human Capital in a Developing Country," *Review of Economic Studies* 64: 311-35.

**Table A1: Descriptive Statistics on Surnames, by Surname Type**

**Mean, standard errors in parentheses, percentages in brackets**

| | Head's Paternal Surname (F1) | Head's Maternal Surname (f1) | Spouse's Paternal Surname (F2) | Spouse's Maternal Surname (f2) |
|---|---|---|---|---|
| **Number of surnames** | 1696 | 1996 | 1912 | 2025 |
| **Number [percentage] of surnames mentioned more than once** | 1064 [62.7] | 1188 [59.5] | 1088 [56.9] | 1100 [54.3] |
| **Probability of same surname in population** | $9.50 \times 10^{-6}$ | $7.54 \times 10^{-6}$ | $8.60 \times 10^{-6}$ | $8.33 \times 10^{-6}$ |
| | $(5.48 \times 10^{-6})$ | $(4.16 \times 10^{-6})$ | $(4.95 \times 10^{-6})$ | $(4.95 \times 10^{-6})$ |
| **Expected number of same surname matches in population** | 13.3 | 11.2 | 9.92 | 9.26 |
| | (1.66) | (1.36) | (1.25) | (1.19) |
| **Probability of same surname in the village** | .042 | .021 | .022 | .020 |
| | (.0005) | (.0004) | (.0004) | (.0004) |
| **Expected number of same surname matches in the village** | 7.55 | 5.31 | 5.42 | 4.98 |
| | (.039) | (.036) | (.036) | (.040) |
| **Odds ratio** | 355.7 | 344.8 | 345.4 | 353.0 |
| | (8.26) | (7.47) | (7.55) | (8.18) |

**Notes:** For the matching probabilities and expected number of same surname matches in the population, the standard errors are clustered by surname for each surname type. The sample is restricted to those households that can be tracked for the first and third waves of the *Progresa* data, namely in the baseline survey in October 1997 (wave 1) and the first post program survey in October 1998 (wave 3). There are 22553 such households.

**Table A2: The Number of Family Links, by Type of, as Reported in the Mexican Family Life Survey**

**Couple Headed Households**
**Mean, standard error in parentheses clustered by village**

| | Outside of the Household (ANY location) | | | | |
|---|---|---|---|---|---|
| | **Parent** | **Children Aged 0-16** | **Adult Children** | **Siblings** | **All** |
| **From head of household to:** | .476 | - | 1.23 | 3.27 | 4.97 |
| | (.035) | | (.089) | (.116) | (.014) |
| **From spouse of household to:** | .669 | - | 1.23 | 3.50 | 5.39 |
| | (.039) | | (.089) | (.113) | (.148) |

| | Inside of the Household | | | | |
|---|---|---|---|---|---|
| | **Parent** | **Children Aged 0-16** | **Adult Children** | **Siblings** | **All** |
| **From head of household to:** | .047 | 2.02 | .571 | .019 | 2.66 |
| | (.009) | (.079) | (.039) | (.007) | (.084) |
| **From spouse of household to:** | .002 | 2.02 | .571 | .009 | 2.60 |
| | (.002) | (.079) | (.039) | (.005) | (.082) |

**Notes:** The sample is taken from the first wave of the Mexican Family Life Survey, 2001. Standard errors are clustered by village. We restrict this sample to the seven Mexican states that are also covered in the *Progresa* evaluation data, and to couple headed households, in locations with less than 2500 inhabitants. There are 580 such households. By construction, the number of family links to parental households is always conditional on two such family links existing. We do not therefore use information on households that have single parents in any location. By construction, the number of children of the couple inside and outside of the household are identical for the head and the spouse. The number of children outside of the household is restricted to be 17 and older (based on spouses' reports).

**Table A3: Correlates of Extended Family Links**

**Couple Headed Households**

**Conditional logit estimates, grouped on village, standard errors clustered by village, log odds ratios reported**

| | Inter-generational Family Links | | | Intra-generational Family Links | | | | |
|---|---|---|---|---|---|---|---|---|
| **Type of Family Link:** | **Parents of Head** | **Parents of Spouse** | **Adult Child** | **Brothers of Head** | **Sisters of Head** | **Brothers of Spouse** | **Sisters of Spouse** | **Any Link [Connected]** |
| **Head age [years]** | .373*** | .307*** | 1.48*** | .953 | .805*** | .792*** | .841*** | .768*** |
| | (.023) | (.024) | (.079) | (.041) | (.037) | (.034) | (.041) | (.041) |
| **Spouse age [years]** | .755*** | 1.02 | 2.97*** | .872*** | .941 | 1.20*** | .969 | 1.18*** |
| | (.043) | (.072) | (.166) | (.037) | (.041) | (.052) | (.047) | (.063) |
| **Head literate [yes=1]** | 1.45*** | 1.13 | .911* | 1.07 | 1.13*** | .966 | .993 | 1.05 |
| | (.099) | (.095) | (.049) | (.046) | (.056) | (.041) | (.046) | (.052) |
| **Spouse literate [yes=1]** | 1.14** | 1.30*** | .785*** | 1.11** | 1.03 | 1.08 | .956 | 1.13** |
| | (.074) | (.100) | (.039) | (.047) | (.046) | (.048) | (.049) | (.062) |
| **Head speaks indigenous language [yes=1]** | 1.01 | .843 | 1.16 | .963 | 1.02 | .967 | 1.02 | .911 |
| | (.186) | (.138) | (.167) | (.107) | (.125) | (.102) | (.105) | (.115) |
| **Spouse speaks indigenous language [yes=1]** | .854 | 1.05 | 1.02 | .936 | .998 | 1.33 | 1.17 | .998 |
| | (.144) | (.148) | (.145) | (.094) | (.133) | (.175) | (.151) | (.154) |
| **House is owned [yes=1]** | 1.03 | 1.08 | 1.34** | 1.54*** | 1.38*** | 1.26*** | 1.33*** | 1.43*** |
| | (.094) | (.122) | (.200) | (.112) | (.112) | (.107) | (.133) | (.126) |
| **Any land is owned [yes=1]** | .846*** | 1.00 | 1.26*** | 1.14*** | 1.19*** | 1.11*** | 1.14*** | 1.14** |
| | (.046) | (.069) | (.077) | (.050) | (.059) | (.047) | (.056) | (.062) |
| **Any member temporarily migrated in last year [yes=1]** | 1.09 | .993 | 1.15** | .937 | .967 | 1.05 | .993 | 1.01 |
| | (.066) | (.071) | (.071) | (.044) | (.044) | (.051) | (.050) | (.064) |
| **Eligible [yes=1]** | .963 | 1.07 | 1.02 | 1.01 | 1.03 | 1.03 | 1.05 | 1.02 |
| | (.069) | (.091) | (.070) | (.051) | (.056) | (.055) | (.059) | (.064) |
| **Poverty index** | 1.02 | 1.03 | 1.19*** | 1.08** | 1.05 | .975 | 1.02 | 1.00 |
| | (.042) | (.052) | (.048) | (.035) | (.036) | (.031) | (.036) | (.042) |
| **Household size** | .967 | 1.18*** | 1.11*** | 1.10*** | 1.11*** | 1.07*** | 1.11*** | .992 |
| | (.030) | (.040) | (.028) | (.021) | (.022) | (.021) | (.024) | (.023) |
| **Mean of Dependent Variable** | .187 | .101 | .217 | .485 | .332 | .306 | .274 | .807 |
| **Number of Observations** | 18309 | 17046 | 18634 | 18907 | 18686 | 18740 | 17648 | 18611 |

**Notes:** *** denotes that the odds ratio is significantly different from one at 1%, ** at 5%, and * at 10%. In each column a conditional logit specification is estimated, grouped on village, where the standard errors are clustered by village, and the log odds ratios are reported. All continuous variables are divided by their standard deviation so that the corresponding coefficients can be interpreted as the effect of a one standard deviation change in the continuous variable. The underlying sample is restricted to couple headed households that can be tracked over the first and third *Progresa* waves. The sample varies across the columns because villages in which all or no households have the given type of family link are dropped when the conditional logit regression is estimated. All characteristics are measured in the third wave (October 1998) except household size which is measured at baseline. A higher household poverty index implies the household has a higher level of permanent income and so is less poor.

## Table A4: Matching Algorithm and Surnames Based Robustness Checks

**Dependent Variable (Columns 1 to 3): Change in Household's Secondary School Enrolment Rate (November 1999 - October 1997)**
**OLS regression estimates, standard errors are clustered by village**

| | Omit Households With Potential Measurement Error in Surnames | Drop Villages in Top Quartile of Village Size | Random Reassignment of Households To Another Village in the Same Municipality | Number of Children Aged 0-16 in Household At Baseline | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4a) OLS | (4b) Most Frequent Names |
| **TTE [connected]** | .096*** | .131*** | .069*** | | |
| | (.019) | (.023) | (.026) | | |
| **TTE [isolated]** | .005 | .016 | .072*** | | |
| | (.033) | (.040) | (.020) | | |
| **ITE [connected]** | -.015 | .004 | .016 | | |
| | (.031) | (.040) | (.043) | | |
| **ITE [isolated]** | -.010 | -.082 | -.047 | | |
| | (.062) | (.070) | (.040) | | |
| **Share of households with same head's paternal surname** | | | | -.293 | -.054 |
| | | | | (.444) | (.483) |
| **Share of households with same spouse's paternal surname** | | | | .043 | .329 |
| | | | | (.473) | (.489) |
| **ΔTTE** | .091*** | .115** | -.003 | | |
| | (.037) | (.047) | (.029) | | |
| **ΔITE** | -.005 | .086 | .063 | | |
| | (.065) | (.074) | (.058) | | |
| **Observations** | 5490 | 3121 | 5447 | 6227 | 3954 |

**Notes:** *** denotes significance at 1%, ** at 5%, and * at 10%. The sample is restricted to couple headed households that can be tracked over the first and third *Progresa* waves. Standard errors are clustered by village in each column. A household's secondary school enrolment rate is defined to be the fraction of children aged 11 to 16 resident in the household that are full-time enrolled in school at the time of the survey. The link variable is defined to be equal to one if household h has any family links of type j in the village, and zero otherwise. The specifications in all columns except 4a and 4b also control for the following - the husband's age, years of schooling, literacy, whether he speaks an indigenous language, the spouse's age, years of schooling, literacy, whether she speaks an indigenous language, the household poverty index, the number of individuals in the household at baseline, the number of households in the village, the share of households in the village that are eligible, the marginality index for the village, and the village level enrolment rate at baseline among eligible and non eligible households. In all columns except 4a and 4b the effects of the following controls are also allowed to vary with eligibility status, *Progresa* , and the interaction of the two - whether the head's (spouse's) age is above or below the median among couple headed households, whether the head (spouse) is literate, whether the household owns land, whether the household size at baseline is above or below the median among couple headed households, and the village level enrolment rates at the baseline among eligible and non-eligible households. In Column 1 we drop from the sample households in which the - (i) spouse's maternal surname is the same as her husband's; (ii) wife's paternal and maternal surnames are the same; (iii) paternal and maternal surnames of both the head and spouse are the same. In Column 2 the sample is restricted to villages with less than 57 households in them. In Column 3 we randomly reassign each household in our baseline sample to another village within the same municipality and recalculate their extended family links if they actually lived in that alternative village. In Columns 4a and 4b the dependent variable is the number of children aged 0-16 in the household at baseline and control for the following - the husband's age, years of schooling, literacy, whether he speaks an indigenous language, the spouse's age, years of schooling, literacy, whether she speaks an indigenous language, the household poverty index, the number of individuals in the household at baseline, and village fixed effects. In Column 4b the sample is restricted to households in which either the head's or spouse's paternal surname is shared by .5% of the population.

## Table A5: Village and Household Level Characteristics Based Robustness Checks

### Dependent Variable: Change in Household's Secondary School Enrolment Rate (November 1999 - October 1997)
### OLS regression estimates, standard errors are clustered by village

| | Village Interactions | Random Reassignment To Another Household in the Same Village | Household Shocks | Close to Any Health Facility (Less Than 5km) | No Secondary or Middle School in the Village | Marital History Module Defined Links (May 1999) |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **TTE [connected]** | .095*** | .074*** | .095*** | .111*** | .114*** | .095*** |
| | (.017) | (.017) | (.018) | (.031) | (.021) | (.018) |
| **TTE [isolated]** | .013 | .098*** | .010 | .011 | .020 | -.023 |
| | (.033) | (.029) | (.032) | (.048) | (.038) | (.033) |
| **ITE [connected]** | -.004 | -.031 | -.001 | -.077 | -.011 | -.020 |
| | (.039) | (.032) | (.029) | (.061) | (.040) | (.030) |
| **ITE [isolated]** | .010 | .011 | .001 | -.069 | -.012 | -.055 |
| | (.062) | (.052) | (.060) | (.090) | (.077) | (.064) |
| **ΔTTE** | .083** | -.024 | .086** | .100* | .094** | .118*** |
| | (.036) | (.030) | (.035) | (.053) | (.042) | (.037) |
| **ΔITE** | -.014 | -.041 | -.002 | -.008 | -.001 | .035 |
| | (.065) | (.059) | (.064) | (.107) | (.083) | (.066) |
| **Observations** | 6227 | 6227 | 6227 | 2118 | 4347 | 6227 |

**Notes:** *** denotes significance at 1%, ** at 5%, and * at 10%. The sample is restricted to couple headed households that can be tracked over the first and third *Progresa* waves. Standard errors are clustered by village in each column. A household's secondary school enrolment rate is defined to be the fraction of children aged 11 to 16 resident in the household that are full-time enrolled in school at the time of the survey. The link variable is defined to be equal to one if household h has any family links of type j in the village, and zero otherwise. The specifications in all columns also control for the following - the husband's age, years of schooling, literacy, whether he speaks an indigenous language, the spouse's age, years of schooling, literacy, whether she speaks an indigenous language, the household poverty index, the number of individuals in the household at baseline, the number of households in the village, the share of households in the village that are eligible, the marginality index for the village, regional fixed effects, and the village level enrolment rate at baseline among eligible and non eligible households. In all columns the effects of the following controls are also allowed to vary with eligibility status, *Progresa* , and the interaction of the two - whether the head's (spouse's) age is above or below the median among couple headed households, whether the head (spouse) is literate, whether the household owns land, whether the household size at baseline is above or below the median among couple headed households, and the village level enrolment rates at the baseline among eligible and non-eligible households. In Column 1 the effects of the following village characteristics are also allowed to vary with eligibility status, *Progresa* , and the interaction of the two – the number of households in the village, the share of households in the village that are eligible, the marginality index for the village, and the share of households that report being affected by any natural shocks from October 1998 to November 1999. These shocks include being affected by droughts, floods, frosts, fires, pests, earthquakes, or hurricanes. In Column 2 we randomly reassign each household in our baseline sample the family links of another household in the same village. In Column 3 we allow household responses to *Progresa* to vary by whether they themselves have been affected by any shock from October 1998 to November 1999. In Column 4 the sample is restricted to villages that are less than the median distance (5km) from any health facility as measured in November 1999 (wave 5). In Column 5 the sample is restricted to the 410 villages in which there is no secondary or middle school present. In Column 6 we redefine the family links based on information on the presence of parental links in the village collected in the marital history module in May 1999 (wave 4).

**Figure A1: Propensity Scores For Connected and Isolated Households, by Eligibility Status**

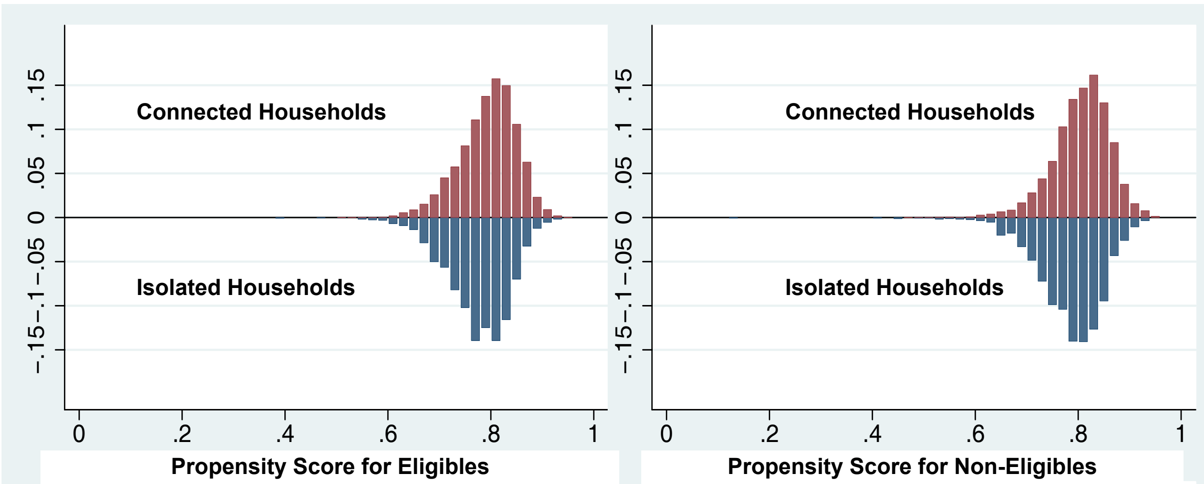
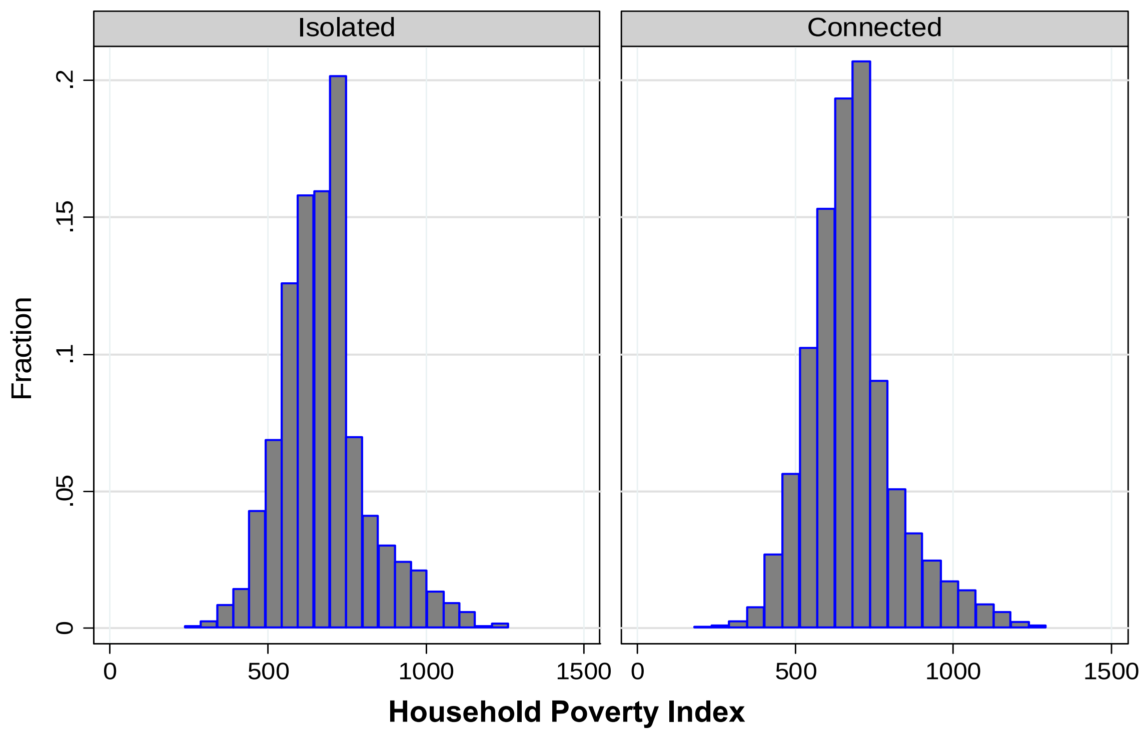



**Figure A2: The Household Poverty Index, for Connected and Isolated Households**




**Notes:** In Figure A1, the propensity score is based on the following observable characteristics - the ages of the head and spouse, dummy variables for whether they are working, literate, and speak an indigenous language, whether the household owns its home, whether any land is owned, whether any member of the household temporarily migrated in the last year, the household poverty index, the number of male (females) aged 0 to 5, 6 to 9, 10 to 12, 13 to 16, 17 to 29, 40 to 55, 56 and older, and village fixed effects. For Figure A2, at baseline, households were classified as either being eligible (poor) or non-eligible (not poor) for Progresa transfers according to a household poverty index. This index is a weighted average of household income (excluding children), household size, durables, land and livestock, education, and other physical characteristics of the dwelling. The index is designed to give relatively greater weight to correlates of permanent income rather than current income.