I Z A

IZA DP No. 4498

Extended Family Networks in Rural Mexico: A Descriptive Analysis

Manuela Angelucci
Giacomo De Giorgi
Marcos A. Rangel
Imran Rasul

October 2009

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Extended Family Networks in
Rural Mexico: A Descriptive Analysis

Manuela Angelucci
University of Arizona and IZA

Giacomo De Giorgi
Stanford University

Marcos A. Rangel
University of Chicago and USP, Brazil

Imran Rasul
University College London and IZA

ABSTRACT

Extended Family Networks in Rural Mexico:
A Descriptive Analysis[*]

We provide descriptive evidence on the characteristics of a household's extended family network using data from the *Progresa* social assistance program in rural Mexico. We exploit information on the paternal and maternal surnames of household heads and their spouses and the patronymic naming convention to identify the inter and intra generational family links of each household to others in the village. This provides an almost complete mapping of extended family networks structures across 506 Mexican villages, covering 22,000 households and over 130,000 individuals. We then provide evidence on – (i) whether husbands and wives differ in the extent to which members of their extended family are located in geographic proximity; (ii) the characteristics that predict the existence of extended family links; (iii) the similarity of households within the same family network in terms of their poverty, and how this differs within and between generations of the extended family.

JEL Classification: J12, O12

Keywords: extended family network, Progresa

Corresponding author:

Manuela Angelucci
Department of Economics
University of Arizona
McClelland Hall 401AA
Tucson, AZ 85721
USA
E-mail: angelucm@eller.arizona.edu

1   Introduction

Economists usually focus on the household as the key unit of analysis from which to study household behavior. There are good reasons for this. Theories of household behavior, such as the unitary model [Becker 1981], bargaining models [Manser and Brown 1980, McElroy and Horney 1981], collective choice models [Chiappori 1988], and non-cooperative models [Chen and Woolley 2001] all emphasize the interplay between the preferences and resources of individuals within the household in shaping household outcomes. In addition, household data is typically constrained in the sense that it is not possible to identify the familial ties between surveyed households.

However, every household is actually embedded within a wider network of its extended family members, and these related households may exert important influences that determine the behavior of any given household within the family network. Indeed, if the institution of the extended family shapes the objectives and constraints relevant for each household, then analyzing household behavior in isolation from the presence and characteristics of its extended family members may lead to an incomplete understanding of the forces driving household choices. The divergence between what may be optimal for a single household acting in isolation and what is actually optimal for the family network as whole can be expected to increase in economic environments characterized by missing markets, correlated income shocks, the prevalence of informal institutions that enforce contracts, and widespread policy interventions that affect many households in the local economy.

This paper begins to fill this void in the literature by providing descriptive evidence from household data used to evaluate the *Progresa* social assistance program in rural Mexico on the presence and characteristics of each member of a household's extended family in the same village. This environment has all of these features described above in which there may be significant influences of the extended family on household behavior.

There are two key aspects of our data and context that lie at the heart of this analysis. First, to build data on the existence of extended family links between households, we exploit information in the *Progresa* data on the paternal and maternal surnames of heads of households and their spouses. Second, we combine this information with the patronymic naming convention in Mexico to build two types of extended family link outside of the household but within the village – (i) inter-generational family links, such as those from the head of a household to his parents, from the spouse of the head of the household to her parents, and from the head and spouse of the household to their adult sons and daughters; (ii) intra-generational family links, such as those from the head of the household to his brothers and sisters, and similarly those for his spouse.

We also use information from the household roster to build equivalent measures of extended family links that co-reside within the household. Taken together, these two sources of information provide an almost complete mapping of extended family structures across 506 villages in rural

Mexico, covering around 22,000 households and over 130,000 individuals.

In this paper we then use this information to provide evidence on – (i) whether husbands and wives differ in the extent to which members of their extended family are located in geographic proximity; (ii) the characteristics that predict the existence of each type of extended family link; (iii) the similarity of households within the same family network in terms of their poverty, and how this differs within and between generations of the extended family.

Family networks may influence household behavior through a number of mechanisms. First, there is scope for insurance or redistribution within families because, as highlighted in the descriptive analysis, the correlation in poverty of households within the same family network vary considerably between inter-generationally and intra-generationally linked households. Much of the existing literature has focused on the determinants of transfers along such inter-generational links, and whether they are crowded in or crowding out by publicly provided transfers [Altonji *et al* 1992, Altonji *et al* 1997, Jensen 2003, Albarran and Attanasio 2004, Raut and Tran 2005]. Understanding how flows of private transfers across households respond to the provision of public transfers can ultimately reveal whether private transfers are motivated by altruistic concerns [Becker 1981, Cox and Jakubson 1995] or the exchange of goods and services [Bernheim *et al* 1985].

Second, households within the same family network may learn about, or revise their expectations of, outcomes on the basis of each others experiences. Relatedly, there may be peer pressure from parents, and rivalry, cooperation, conformity, or imitation among siblings. Evidence from similar rural settings shows the quantitative importance of learning from others in agriculture [Foster and Rosenzweig 1995, Conley and Udry 2005, and Bandiera and Rasul 2006]. However, little is known about learning or peer effects in general, within the same extended family.

Third, inside and outside options in marital bargaining between husbands and wives may be shaped by the presence of each partners' extended family. Much of the literature has explored the effects of the resources controlled by husbands and wives, and other factors shaping spouses' outside options, on marital bargaining [Schultz 1990, Thomas 1990, Lundberg *at el* 1997, Attanasio and Lechene 2002, Rubalcava *et al* 2004, Rangel 2005, Behrman and Rosenzweig 2006, Qian 2006]. A more recent literature has begun to explore the allocation of resources within the household when there are more than two relevant actors present, such as the parents of one of the spouses [Rangel 2006]. However, little is known about the presence of extended family members outside of the household driving within household bargaining.

We view the construction of extended family networks as being a first step in a broader research agenda in which the challenge lies in understanding whether and how such networks shape household behavior. The experimental research design of the *Progresa* evaluation data also opens up the possibility of overcoming some of the econometric concerns that have affected earlier stud-

3

ies, and thus identify the causal effect of extended family links on household behavior. In the conclusion we discuss particular issues that the constructed data on extended family networks can be used to explore.

The emphasis throughout this research agenda is on the influence of the extended family on household behavior, and not the influence of others, such as friends of the family or those that are similar on observable characteristics *per se*. While undoubtedly other households outside of the extended family network also influence behavior, there are a number of theoretical, empirical, and sociologically founded justifications for focusing on extended family links.

First, evolutionary biology suggests preferences may be defined over the family dynasty. This aspect of behavior is commonly modelled within an overlapping generations framework. Moreover there are specific inter-generational investments – such as those into children's education, bequests, transfers to parents, and the choice of marriage partners – that have no counterpart in the relationships between friends.

Second, the extended family provides a well defined reference group from which to empirically identify social influences on behavior [Manski 1993]. Third, social ties between family members may differ from those with friends because the former are strong in the sense they are long term, embody mutual trust and reciprocity, and are not easily undone [Granovetter 1985]. This may for example make it easier to enforce implicit contracts within families than among friends, especially when formal legal institutions are weak. In addition, the process underlying the formation of family networks is different from that underlying the formation of friendship networks. For example, households have less scope to strategically place themselves within a family network, and family ties are necessarily bidirectional.

The paper is organized into five sections. Section 2 describes the *Progresa* program and evaluation data. Section 3 discusses how we exploit information on paternal and maternal surnames to identify family links outside of the household. Section 4 describes the data on family networks. Section 5 concludes with a discussion of how the constructed information on extended families can be utilized in future research.

2   The *Progresa* Program and Evaluation Data

Progresa is an ongoing publicly funded social assistance program in Mexico, targeted towards poor households. The program began in 1997 and was initially offered to 140,544 households in 3,369 villages. The program provides grants to households in the form of conditional cash transfers related to their children's school attendance, and attendance at local health clinics. By the end of 1999 the program had expanded to cover more than 2.6 million recipient households throughout

4

rural Mexico [Skoufias 2005].[1]

To determine a household's entitlement to *Progresa* transfers, in 1997 households were classified as either being eligible (poor) or non-eligible (non-poor) according to a household poverty index. This index is calculated as 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, rather than current, income.[2] In October 1997, 52% of households were classified to be eligible for *Progresa*.

To evaluate the effectiveness of *Progresa*, an experimental research design was implemented and household data collected on a panel of around 24,000 households every six months in 506 villages between March 1998 and November 1999.[3] Of the 506 villages, 320 were randomly assigned to the treatment group, namely locations where *Progresa* would later be implemented, and 186 villages were assigned to be control villages. Figure A1 summarizes the research design.

The first two waves of household panel data were collected pre-program, in October 1997 and March 1998, namely before any cash transfers from *Progresa* were distributed. The remaining waves, collected in October 1998, May 1999, and November 1999, all correspond to the post-program period. We restrict our analysis to the 22,553 households that are surveyed in the first and third waves of the data. As we aim to compare and contrast the extended family links of heads and their spouses, we focus attention on the 85% of households that are couple headed, where the head is always defined to be male.

3   Construction of Extended Family Links

3.1   Information on Surnames and the Matching Algorithm

To build data on the existence of extended family links between households in the same village, we exploit information on surnames provided in the third wave of the evaluation data. We combine this information with the patronymic naming convention in Mexico to build two types of extended family link for each household – (i) inter-generational family links, such as those from the head of

---

[1] Evidence that the program had a significant impact on the welfare and human capital investment of poor rural families in Mexico contributed to the decision in 2000 of the Vicente Fox Quesada administration to continue with *Progresa* and to expand its coverage to urban areas in 2003 under the new name of *Oportunidades*.

[2] The cutoff point in the household poverty index that determines eligibility varies across regions of Mexico.

[3] These villages are located in seven out of thirty one states in Mexico. These states are Guerrero, Hidalgo, Michoacan, Puebla, Queretaro, San Luis Potosi, and Veracruz, and they are mainly located in the southern half of Mexico. Villages were selected on the basis of a marginality index constructed from information on the share of illiterate adults in the village; the share of dwellings without water, drainage systems, electricity, and with floors of dirt; the average number of occupants per room in village households; the share of population working in the primary sector; the distance from other villages, and the health and school infrastructures present in the village.

a household to his parents, from the spouse of the head of the household to her parents, and from the head and spouse of the household to their adult sons and daughters; (ii) intra-generational family links, such as those from the head of the household to his brothers and sisters, and similarly those for his spouse.[4]

It is important to first note that Mexicans have *two* surnames – the first is inherited from the father's paternal lineage and the second from the mother's paternal lineage. For example, former Mexican president Vicente Fox Quesada would be identified by his given name (Vicente), his father's paternal name (Fox) and his mother's paternal name (Quesada).

In the evaluation data, respondents were asked, "*Tell me the complete/full name with all surnames for all the members of this household, starting with the head of household*". Respondents were then asked to provide the – (i) given name; (ii) paternal surname; and, (iii) maternal surname, for each member of the household.[5] Hence a household that is headed by a husband and wife, has four associated surnames – the paternal and maternal surname of the head, and the paternal and maternal surname of his wife.[6]

To define each inter and intra generational link, we use information on two of these four surnames. Figure 1 provides an illustration of the matching algorithm used to build each family link. Consider household **A** at the root of the family tree. The head of the household has paternal and maternal surnames $F1$ and $f1$ respectively. His wife has paternal and maternal surnames $F2$ and $f2$ respectively.[7]

---

[4]Two concerns arise from the fact that the surnames information used to construct family networks is measured in the first wave of post-program data collection (October 1998, wave 3). First, households may have incentives to endogenously respond to the program by changing the structure of households, in particular, by artificially forming new households in order to increase the number of eligible individuals in the family. This concern is partly ameliorated by the fact that a register of eligible households was drawn up in the baseline wave of data, and only households recorded to be eligible at that point were later entitled to receive transfers. In addition we note that although there is an increase in the number of households from the baseline survey to wave 3, this increase is proportionately the same in both treatment and control villages. A second concern is that the program may affect the migration of the household head or of his spouse. One piece of evidence against this is that only .35% (.50%) of households in wave 3 (wave 5) report having a migrant head or spouse. Moreover the share of households with such migrants does not differ across treatment and control villages.

[5]The precise wording of the question in Spanish is as follows, "*Dígame por favor el nombre completo con todo y apellidos de todas las personas que viven en este hogar, empezando por (jefe del hogar) – (i) nombre; (ii) apellido paterno; (iii) apellido materno*".

[6]We manually cleaned the names data according to the following criterion. We first removed non-alphabetical characters, replaced "Sin Apellido" (no surname) with missing values, and corrected some obvious typos based on intra-household surname checks. Second, we imputed a small number of missing female surnames from wave 2 (as wave 2 data only contains information on surnames for transfer recipients), and finally, we verified the surnames data using the same information from wave 5. No information on surnames is available in the first wave of data.

[7]We use the convention that the head's surnames are written in black, and those of his wife are written in red italics. Paternal surnames are indicated in upper case and maternal surnames are indicated in lower case. First names are not shown as they are not relevant for the construction of extended family links. Each household in the family tree is assumed to be couple headed purely to ease the exposition. In Anglo Saxon countries, $F1$ would correspond to the family name and $F2$ would correspond to the mother's maiden name.

The children of the couple in household **A** will adopt the paternal surnames of their father ($F1$) and mother ($F2$). Hence we define there to be a parent-son relationship between households **A** and **B** if – (i) the paternal surname of the head in household **B** is the same as the paternal surname of the head in household **A** ($F1$); and, (ii) the maternal surname of the head in household **B** is the same as the paternal surname of the spouse in household **A** ($F2$). Clearly, parent-daughter extended family relationships can be similarly defined. In addition, intra-generational family ties between siblings can be identified. For example, the heads of households **B** and **C** are identified to be brothers if they share the same paternal and maternal surnames. This relationship can be identified even if the parents of **B** and **C** themselves are not heading their own household, but are perhaps co-residing with one of their adult children.

While in Figure 1 we assume all households are couple headed, this need not be the case. In order to deal with this issue, we combine our matching algorithm with information on whether households are couple headed or not, and the gender of the head if they are single headed, to precisely define each extended family link.

There are some limitations on the extent to which information on paternal and maternal surnames can be used to accurately measure the presence and number of extended family links. We now make clear the limits of our matching algorithm, and some refinements we employ to address these concerns. In the next subsection we discuss in more detail the possible forms of measurement error in family links we identify, and provide evidence on the prevalence of each type of error in the data.

We only define extended family links that exist within the same village. There are three reasons for this. First, family that are in close geographic proximity are more likely to influence household behavior. Second, villages were partly selected into the evaluation data because they were geographically remote. Hence there is a relatively low probability of extended family residing in neighboring villages. Third, the number of households reporting a member to have permanently migrated away is less than 4%, so that within village family transfers are likely to be more important that those from other locations.[8]

Second, consider links from household i to household j where household j is single headed. As shown in Figure 1, the fact that household j is single headed does not affect the construction of links from the head and spouse of household i either to their children or to their siblings. However, links from the head (spouse) of household i to the household of his (her) parents can only be identified if *both* his (her) parents are alive and live together. This is because this particular

---

[8]In Angelucci *et al* [2006] we provide evidence from the *Mexican Family Life Survey* (MxFLS), that contains information on the number of family links that exist in *any* location, to verify that the within village links that we identify in the *Progresa* data are indeed no larger than those measured in the MxFLS for a comparable subsample of households.

family link definition relies on information from household j on the paternal surnames of both the head and spouse.[9]

Support for the accuracy of the matching algorithm can be found using 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. Reassuringly, this is true for 97% of households in the *Progresa* data using our matching algorithm. On the other hand, there is no upper bound on the number of links that household i may have to their children or siblings. To reduce potential errors, we therefore combine our matching algorithm with the following information when defining family links – (i) we limit inter-generational links (parent-child, child-parent) to exist when the relevant individuals have at least 15 years age difference, and no more than 60 years age difference between mother and child; (ii) we limit intra-generational links (siblings) to exist when the relevant individuals have at most 30 years age difference.

3.2   Measurement Error in Family Links

There are a number of potential sources of measurement error in the recorded surnames data that can be checked for. The first arises from the convention that women change their maternal surname to that of their husband's 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 6.9% of households do we observe the spouse's maternal surname to be the same as that of her husband. This figure therefore 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 might not recall his wife's maternal surname and simply replace it with her paternal surname. This is plausible given that his children would actually receive his wife's paternal surname as their second surname. Reassuringly, this occurs in only 5.9% of households.

Third, the accuracy of the surname data may be questioned in households where the paternal and maternal surnames of both the head and spouse are all reported to be identical. There are 1.7% of households in which this is the case, although the figure drops to .5% if we exclude households with the most common surname in the data.

With these caveats and concerns in mind, we note finally that there are potential forms of measurement error in family links that simply cannot be addressed. The first obviously arises

---

[9]Again, in Angelucci *et al* [2006] we use the MxFLS data to verify the extent to which the *Progresa* data underestimates the number of links from household i to the parents of the head or spouse because some of those parents will undoubtedly be widowed. However we note that female widows above 40 are 37% more likely to live as a dependent than head an independent household, relative to a similar married woman.

from any remaining typos in the surname data. Second, there might 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. As mentioned before, one indication that this problem may not be first order, is that 97% of households are matched to two or fewer parental households.

Third, consider a scenario in which a woman'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 in the same family network, we may overestimate the closeness of the family tie.

4   Descriptive Evidence

4.1   Surnames Data

Table 1 provides descriptive evidence on surnames, split according to 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 given the patronymic naming convention, paternal surnames have a greater survival rate across generations than maternal surnames.

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 is also partly due to spouses moving into the 506 villages in the data from villages outside of the evaluation sample.

The majority of surnames are mentioned at least twice. Figure 2 shows the frequency distribution of surnames for the most frequently mentioned names. For each surname type, there is one particular name that covers around 9% of households. For the head's paternal surname, the median household has the 55th most common surname.

The third row highlights that the probability (without replacement) of two randomly matched households in the sample 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, perhaps at the time of marriage.[10]

---

[10]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 $P_i = \left(\frac{n_i}{N}\right) \cdot \left(\frac{n_i - 1}{N - 1}\right)$, and the expected number of households in the population with name i is $E_i = n_i \cdot \left(\frac{N-1}{N}\right)$. The values

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 type is orders of magnitude larger than that in the population. 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. On the other hand, the expected number of same surname matches in the village is orders of magnitude higher than if households were randomly allocated by surname types into villages.[11]

Note that this expected number of matches in the village are based on the use of only one surname, and so provides an upper bound on the total number of specific extended family links we define, which are always based on two name matches.

Figures 3A and 3B show how the probability and expected number of same surname matches in the village vary by village size, as defined by the number of households in the village. We see that the probability of a same surname match declines with village size, but not at the same rate as the increase in village size. Hence the expected number of same surname matches increases with village size overall. This again suggests some degree of sorting by each surname into villages.

The final row sheds more light on the degree of sorting of households into villages by the surname type, as measured by the odds ratio. This is defined as the ratio of the probability that two randomly chosen households from within the same village have the same surname, divided by the probability that two randomly chosen households from the population of *Progresa* villages have the same surname. The odds ratio thus suggests that 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.[12]

4.2   The Number of Family Links

The upper panel of Table 2 provides information on the number of family links each household has to other households in the village, as generated by the matching algorithm described in Section

---

reported in Table 1 are the averages of P_i and E_i over all surnames i.

[11]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 denotes 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 $e_{iv} = n_{iv} \cdot \left(\frac{n_v-1}{n_v}\right)$. The values reported in Table 1 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 take account of the fact that the same name may be reported to different extents in different villages.

[12]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 $\frac{\widetilde{p}_i}{P_i}$.

3. The columns of the table split family links into – (i) inter-generational links, namely those to parents of the head or spouse, and those to the adult children of the head and spouse; and, (ii) intra-generational links, namely those to siblings of the head and spouse. Each row in the table splits each type of extended family link into those from the head of the household, and those from his spouse. The sample is limited to couple headed households that have at least one extended family link in the village. Around 81% of couple headed households have at least one extended family member heading their own household in the village.

The lower panel of Table 2 uses information from the household roster from wave 3 of the data to construct information on the corresponding family links that co-reside inside the household. Again, using information on the relationship of each person to the household head, we can decompose these family links inside of the household into those from the head, and those from his spouse.

Consider first links to parents. Table 2 shows that parents are more likely to reside outside of the household than inside the household of their adult children. The number of parents present is higher for the head than for his spouse, and this is true for parents inside and outside the household. This is consistent with either the spouse migrating to the village, or because she has moved in with her husband's family within the same village.

On links to adult children, by construction, the number of family links to children are identical for head and spouse. Given the age of respondents, there are many more young children inside than adult children outside the household. On links to siblings, we find that these are more likely to be outside than inside the household, which is again as expected given the age of respondents. The number of siblings is higher for the head than his spouse, and this is again true for family links both inside and outside of the household.

Overall, the figures emphasize that the extended family networks of male head of households are considerably greater than those of their spouse. This is true for family links that co-reside within the same household, and for family links that reside in other households in the same village. The source of this difference may arise from women moving households and potentially villages, at the time of marriage. In support of this we note that in wave 4 of the data, wives were asked about where they went to live after marriage – 49.3% stated that they went to live with their in-laws after marriage, and only 6.5% report living with their own parents after marriage. Moreover, 85% of spouses that have their parents present in the village report remaining in the same village at the time of marriage. The corresponding figure for spouses that have no parental links in the village is only 59%.

Along other margins, women with and without parents in the village are similar. For example, age at marriage is not significantly different between women with and without parents, and the proportion of women that report that their in-laws originally proposed the marriage is 56% for

both women that have their parents present and those that do not.

These results are in the same spirit as the findings of Rosenzweig and Stark [1989] who examine marital arrangements in rural India. They find women move to distant villages at the time of marriage and argue that in an economy characterized by information costs and spatially covariant risks, such marriages can be viewed as a consumption smoothing device. Indeed they find that marriages accompanied by migration of the daughter significantly reduce the variance of food consumption in the household of origin. We leave for future research a more complete analysis of the relationship between the marriage market and consumption smoothing in rural Mexico.

Finally, in Table 2 we also decompose the overall variation in each statistic into that arising from variation between villages, and that arising from variation across households in the same village. Reassuringly, in each case we find that the within village variation is around ten times as great as the between village variation. This highlights that variations in family structure occur within the same village, and it is not that there are systematically different types of family structures in each village.[13]

4.3   Correlates of Family Links

The presence of extended family members is unlikely to be exogenously determined. The earlier literature has explored a number of mechanisms that 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]. In this paper we present correlations between three types of factor and the presence of extended family ties.

First, there is a somewhat mechanical correlation between the age of the head and spouse and the likelihood that their parents and adult children are in close proximity.

Second, wealthier family dynasties may enjoy higher fertility and lower mortality, and so are more likely to have extended family members present, other things equal. Wealthier families may also be more likely to own land. As rural land markets are typically missing, the ability to inherit land, or to acquire land specific human capital, may lead adult children to be more likely to remain

---

[13]For the variance decomposition to sum to the total variance in an unbalanced panel, it is necessary to weight the between component by the number of households in the village, denoted H_v. If N denotes the number of observations in the sample and there are N_v villages, the decomposition of the total variance of the number of family links of household h in village v, L_{iv}, into the within and between components, as reported in Table 2, is,

$$\frac{1}{N}\sum_{v=1}^{N_v}\sum_{i=1}^{H_v}\left(L_{iv}-\overline{L}\right) = \frac{1}{N}\sum_{v=1}^{N_v}\sum_{i=1}^{H_v}\left(L_{iv}-\overline{L}_v\right) + \frac{1}{N}\sum_{v=1}^{N_v}H_v\left(\overline{L}_v-\overline{L}\right).$$

within the village than otherwise.

A third mechanism behind why households may partition and form larger extended family networks is through the need to insure against idiosyncratic income shocks. Such insurance can be gained through the strategic marriage of daughters into families with less correlated income shocks, or through the 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 following extended family links – (i) inter generational links such as the parents of the head (spouse), and links to adult children; (ii) intra generational links such as the brothers and sisters of the head and spouse; and, (iii) whether household h has any extended family link.

For each link type, L_{jh}, we control for the ages of the head and spouse, and dummy variables for whether they are working, 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 the household size as measured at baseline.[14]

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, if some villages are differentially subject to climatic shocks, that alters the need for households to insure each other against village level shocks and may cause alternative patterns of extended networks to form.

Standard errors are clustered by village and we report log odds ratios throughout. Hence tests of significance are relative to the log odds ratio being equal to 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 3, 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 of the household and resident in the village – the log odds ratio is significantly less than one, 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 women move village at the time of marriage.

Second, literate heads and spouses are more likely to have their parents present. If such corre-

---

[14]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%.

lations 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 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. Households in which at least one member has temporarily migrated in the last year – 18.5% of all households – are more likely to have adult children 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 status on the presence of any extended family ties. Whether the head and spouse speak an indigenous language also does not predict the presence of extended family ties. This is reassuring because the 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. There are at least three explanations of this. First, there may be persistent differences in fertility levels within the same family dynasty across generations. Second, the presence of extended family members reducing the costs of having and raising children because extended family members are able to supply of time, labor, and other resources to the household. Third, there may be pressure for households to partition once they become sufficiently large.

A comparison across the Columns of Table 3 is also informative. Consider first the presence of the parents of the head and spouse. In general, the control variables have a differential effect of the likelihood that the parents of the head are present, vis-à-vis the parents of the spouse being present. This is in line with the earlier evidence suggesting that 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.[15]

---

[15]These differential effects of a given control variable across two columns can be easily tested for using a t-test. This is based on first stacking any two columns of data, say for L_{jh} and $L_{j'h}$, defining the dependent variable to be equal to one if either L_{jh} or $L_{j'h}$ are equal to one and zero otherwise, and then introducing a complete set of interactions between each control and a dummy variable that is equal to one if the observation is from the first column, and is zero otherwise. The null hypothesis for the t-test is that the interaction term is zero.

4.4 Family Networks

Having described the construction of family links between any given pair of households in the same village, and the characteristics that correlate with the existence of such a link, we now describe the characteristics of the family network as a whole. To be precise, consider the following scenario in which the heads of households i and j are linked because they are brothers. Household j may itself be linked to household k, say, because the parents of the spouse of household j reside in household k. Households i and k then lie within the same family network, even though they do not have a *direct* family link between them. Following the terminology in Wasserman and Faust [1994], households i and j are said to be of distance 1 from each other and households i and k are of distance 2 from each other. Two households i and j are then defined to be within the same family network if the distance between them, d_{ij}, is finite.

We use tools from the analysis of social networks to describe family networks in *Progresa* villages. For this analysis we do not consider households that are single nodes – namely those unconnected to any other household in their village. We therefore consider family networks with at least two members. There are 2196 such family networks, covering 17,030 households.[16]

To begin with, Figure 4A shows the number of family networks in a village. There are around five to ten different family dynasties within the same village. Figure 4B shows the size distribution of family networks. On average, there are 7.8 households within the same family network. On average, 17% of households within a given village are part of the same family network.

The upper panel of Table 4 provides precise information on these statistics.[17] In addition we provide information on the diameter of the network. This is the largest distance d_{ij} between any two households in the family network. This is around 2.5, and one implication of this is that family networks in *Progresa* are unlikely to span across more than three generations. Note finally that there is considerable variation in each statistic both within family networks in the same village, and between family networks in different villages.

To emphasize how the structure of family networks can differ between villages, Figure 5 provides a graphical representation of family networks in two villages of median size (36 households). The upper panel shows a village with a relatively disperse set of family networks. There are five families

---

[16]Single nodes correspond to three types of household. First, those in which all family relations are living under the same roof. Second, there may households for whom there are no children or siblings of the head or spouse heading their own households in the village, and for whom the parents of the head and/or spouse are outside of the household but are widowed. Finally, there may be genuine cases of households for whom there are no children or siblings of the head or spouse heading their own households in the village, and for whom the parents of both head and spouse have either died or do not reside in the same village.

[17]There is one observation per family network so that each network has the same weight irrespective of the number of households within it. In the household level statistics in the lower panel of the table, there is one observation per household and so family networks that are larger will have more weight associated to them.

in which there are only two households present, one large family network in which there are ten households, and 16 households that have no family ties to any other household in the village. In contrast the lower panel shows another village – which recall is of the same size – but in which there exists one large family network containing 30 households, and 6 single node households.

The lower panel of Table 4 then provides statistics at the household level for households in family networks. The first statistic is the degree of the household which is simply the number of family links the household has to other households in the village. The average household in a family network is linked to 4.39 other households in the same village. The second column shows the mean distance between any two households in the same family network is 2.93.

The last statistic, on central closeness, is a function of the inverse of the mean distance between household i and other households in its family network. To understand the intuition behind this measure, consider the network shown in the lower panel of Figure 5. In this family, some households are in denser parts of the network in the sense that households linked to them are also more likely to be linked to each other, and so have lower central closeness. In contrast, other households are at the periphery of the network and so have higher central closeness. In order for the statistic to be comparable across networks of different size, the inverse of the mean distance between household i is multiplied by the size of the network minus one. The index ranges from zero to one. We see that for the average household this index is close to .5, although it varies considerably within households in the same family network.

Although the previous descriptive evidence opens up the possibility of there being differential impacts of *Progresa* depending on the presence of the extended family, this subsection has high-lighted another potential dimension of heterogeneous responses arising from the position of the household in the family network.[18]

4.5   Poverty Within Families

In this final stage of descriptive analysis we use the information on family networks data to shed light on the similarity in characteristics between households within the same family, and to show how this differs depending on the type of extended family link between households. In particular we focus on the poverty index and eligibility status of households within the same family network.

The first row of Table 5 reports the correlation in the poverty index between various pairs of households. The first column shows that if two households are chosen at random from within the same village, the correlation in their poverty indices is .307. This correlation is higher at

---

[18]There is a growing theoretical literature on the strategic behavior of agents within network structures. For example, Galeotti *et al* [2006] develop a framework to analyze strategic interactions when each agent's underlying network affects their payoffs. Bloch *et al* [2005] develop a model of insurance and information exchange within networks.

.413 for any two families within the same family network. The remaining columns show how this correlation varies across and within generations of the same family.

Focusing first on inter-generational family ties, the correlation in poverty tends to be higher among parent-son links than among parent-daughter links. This may be because marriage is used as a mechanism by which daughters marry into households that have less correlated shocks to those of her family, and hence allows extended families to diversify their exposure to risk from income and other shocks [Rosenzweig and Stark 1989]. This also raises the possibility that net transfers from parents to their sons outside of their household may differ from those to their daughters outside of their household, everything else equal.

This intuition is somewhat confirmed when we examine the correlation in households related through same gender intra-generational family links. The correlation is higher among brothers than sisters. This raises the possibility that any transfers among siblings may in part relate to the gender of the giving and receiving sibling, all else equal.

The remaining rows provide information on the similarity of eligibility status among families. For two randomly chosen households in the same village, the probability they are both eligible is .581, and this is only slightly higher at .622 among two households within the same family network. However, breaking down the family ties into inter and intra-generational links, two important results emerge. First, this probability is lower among households linked through inter-generational family links than would be the case for two randomly chosen households. In contrast, this probability is higher among households tied through intra-generational family links than would be the case for two randomly chosen households. While this may in part be driven by the definition of eligibility – which is partially based on housing assets which older individuals have had more time to accumulate – these findings open up the possibility that the extent of transfers within families will differ significantly between inter and intra-generational links.

To shed more light on the eligibility status within poor and non-poor families, the remaining rows provide information on the probability household i is eligible conditional on the eligibility status of household j. Consider first a pair of households in the family linked through a head-parent relationship. We see that among non-poor families, the probability the household of the head is non-poor conditional on his parents being non-poor is $1 - .541 = .459$. In contrast, the probability the household of the head is poor conditional on his parents being poor is .767. The probability of a spouse being poor conditional on her parents being non-poor is .381, and the probability she is poor conditional on her parents being poor is .759. For each probability, the standard error is always marginally higher among family links involving females than the corresponding male defined links. This suggests the economic status of women's households are more diversified than those of males within the same family network, other things equal. We leave for future work a

complete study of the interplay between marriage markets and the inter-generational persistence of poverty.

5   Conclusions

This paper has described the presence and characteristics of extended families in rural Mexico. Using data from the *Progresa* social assistance program we exploit information on the paternal and maternal surnames of household heads and their spouses and the patronymic naming convention to identify the inter and intra generational family links of each household to others in the village. We then use the constructed data on family networks to describe – (i) whether husbands and wives differ in the extent to which members of their extended family are located in geographic proximity; (ii) the characteristics that predict the existence of extended family links; (iii) the similarity of households within the same family network in terms of their poverty, and how this differs within and between generations of the extended family.

We view this as a first step of a broader research agenda in which the challenge lies in understanding whether and how such networks shape household behavior. Before discussing particular issues that the constructed data on extended family networks can be used to explore, it is useful to reiterate two key features of the *Progresa* research design that should be exploited in any analysis of the effects of extended family networks on household behavior. First, villages in the evaluation data are randomly assigned into treatment and control groups. Hence, identification of the causal effect of *Progresa* on eligible households arises from a comparison of eligible households in treatment villages, to eligible households in control villages.

Second, all eligible and non-eligible households in each village are surveyed. Hence the causal effect of *Progresa* on non-eligible households can also be identified. Such within village spillover effects of the program on household behavior are identified from a comparison of non-eligible households in treatment villages, to non-eligible households in control villages.[19]

These features of the data are exploited in Angelucci *et al* [2006] where we estimate whether a household's response to *Progresa* in terms of its secondary school enrolment choices, are heterogeneous depending on the existence and characteristics of extended family members in the same village.[20] To reiterate, the *Progresa* research design allows us to identify these responses both for

---

[19]Earlier papers using *Progresa* have found evidence of the program on non-eligible households [Bobonis and Finan 2005]. *Progresa*'s spillover effects are not however limited to schooling outcomes. Angelucci and De Giorgi [2006] show that the program also increases consumption among non-eligible households. The mechanism behind this is that the liquidity injection of the program benefits all village residents by improving insurance against risk, thus enabling households to borrow more and reduce their stock of precautionary savings.

[20]We thus build on the earlier work of Schultz [2004], Attanasio *et al* [2005], and Behrman *et al* [2005] who estimate the effects of *Progresa* on schooling enrolment, but do not explore whether these effects are heterogeneous

eligible and non-eligible households. This is important because if resources are transferred within family networks, this opens up the possibility of non-eligible households benefiting from *Progresa* if they have extended family links that are eligible for the program.

In that paper we estimate whether a household's response to *Progresa* differs according to the whether any extended family ties are present or not, and conditional on the presence of extended family, whether the precise type of extended family link present influences the response to *Progresa*. We distinguish between inter and intra-generational family links in the village, and between the extended family links of the head of the household vis-à-vis those of his spouse. Finally, we consider if the eligibility status and other characteristics of extended family links themselves matter, and whether there are similar effects of extended family members that co-reside within the household.[21]

The constructed data on extended family networks can also be used to shed light on the provision of insurance in village economies. In particular by combining information on family networks with the detailed information on consumption that is collected in the *Progresa* surveys, we aim to shed light on whether households within extended families are able to insure each other against idiosyncratic shocks [Townsend 1994, Dercon and Krishnan 2000]. One mechanism through which such insurance may occur is the marriage market. In particular, following Rosenzweig and Stark [1989], the data on extended families can be used to understand how families use the marriage market to insure themselves against risks by marrying daughters into other family networks that face shocks less correlated to their own.

We view this research agenda as helping to not only understand the behavior of households in a wider context of them being embedded in a family network, but also to emphasize the need to collect information on the familial ties between households in future survey data.

References

[1] ALBARRAN.P AND O.ATTANSIO (2004) Do Public Transfers Crowd Out Private Transfers? Evidence from a Randomized Experiment in Mexico, in *Insurance Against Poverty* (editor S.Dercon), Oxford: Oxford University Press.

[2] ALTONJI.J.G, F.HAYASHI, AND L.J.KOTLIKOFF (1992) "Is the Extended Family Altruistically Linked? Direct Tests Using Micro Data", *American Economic Review* 85: 1177-98.

according to the presence and characteristics of extended family members.

[21]There is evidence from the US that schooling choices within extended families are correlated. Loury [2006] uses NLSY data to show the extent to which the characteristics of extended family networks are correlated to child schooling outcomes. Aunts, uncles, and grandparents's education are significantly correlated to college attendance probabilities and test score results of their younger relatives. In some cases, the sizes of the estimated effects are large enough to substantially narrow the achievement gap between disadvantaged and other youth.

[3] ALTONJI.J.G, F.HAYASHI, AND L.J.KOTLIKOFF (1997) "Parental Altruism and Inter Vivos Transfers: Theory and Evidence", *Journal of Political Economy* 105: 1121-66.

[4] ANGELUCCI.M AND G.DE GIORGI (2006) Indirect Effects of an Aid Program: How Do Liquidity Injections Affect Non-Eligibles' Consumption?, mimeo Stanford University.

[5] ANGELUCCI.M, G.DE GIORGI, M.A.RANGEL, AND I.RASUL (2006) Family Networks and Schooling Outcomes: Evidence from a Randomized Social Experiment, mimeo University College London.

[6] ATTANASIO.O AND V.LECHENE (2002) "Tests of Income Pooling in Household Decisions," *Review of Economic Dynamics* 5: 720-48.

[7] ATTANASIO.O, C.MEGHIR, AND A.SANTIAGO (2005), Education Choices in Mexico: Using a Structural Model and a Randomised Experiment to Evaluate *Progresa*, IFS Working Paper EWP05/01.

[8] BANDIERA.O AND I.RASUL (2006) "Social Networks and Technology Adoption in Northern Mozambique", *Economic Journal* forthcoming.

[9] BECKER.G.S (1981) *A Treatise on the Family*, Cambridge: Harvard University Press.

[10] BEHRMAN.J.R, P.SENGUPTA, AND P.TODD (2005), "Progressing Through *Progresa*: An Impact Assessment of a School Subsidy Experiment in Rural Mexico", *Economic Development and Cultural Change* 54: 237-76.

[11] BEHRMAN.J.R AND M.R.ROSENZWEIG (2006) "Parental Wealth and Adult Children's Welfare in Marriage", *Review of Economics and Statistics* 88: 496-509.

[12] BERNHEIM.D, A.SHLEIFER, AND L.H.SUMMERS (1985) "The Strategic Bequest Motive", *Journal of Political Economy* 93: 1045-76.

[13] BLOCH.F, G.GENICOT, AND D.RAY (2005) Informal Insurance in Social Networks, mimeo New York University.

[14] BOBONIS.G AND F.FINAN (2005) Endogenous Peer Effects in School Participation, mimeo, University of California Berkeley.

[15] COX.D AND G.JAKUBSON (1995) "The Connection Between Public Transfers and Private Interfamily Transfers: Theory and Evidence", *Journal of Public Economics* 57: 129-67.

[16] CHEN.Z AND F.WOOLLEY (2001) "A Cournot-Nash Model of Family Decision Making", *Economic Journal* 111: 722-48.

[17] CHIAPPORI.P.A (1988) "Rational Household Labor Supply", *Econometrica* 56: 63-90.

[18] CONLEY.T AND C.UDRY (2005) Learning About a New Technology: Pineapple in Ghana, mimeo, Yale University.

[19] DERCON.S AND P.KRISHNAN (2000) "In Sickness and in Health: Risk Sharing Within Households in Rural Ethiopia", *Journal of Political Economy* 108: 688-727.

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

[21] FOSTER.A AND M.ROSENZWEIG (1995), "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture", *Journal of Political Economy* 103: 1176-209.

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

[23] GALEOTTI.A, S.GOYAL, M.O.JACKSON, F.VEGA-REDONDO, AND L.YARIV (2006) Network Games, mimeo Caltech.

[24] GRANOVETTER.M.S (1985) "Economic Action and Social Structure: The Problem of Embeddedness", *American Journal of Sociology* 91: 481-510.

[25] JENSEN.R.T (2003) "Do Private Transfers Displace the Benefits of Public Transfers? Evidence from South Africa", *Journal of Public Economics* 88: 89-112.

[26] LOURY.L.D (2006) "All in the Extended Family: Effects of Grandparents, Aunts, and Uncles on Educational Attainment", *American Economic Review Papers and Proceedings* 96: 275-8.

[27] LUNDBERG.S, R.POLLAK, AND T.WALES (1997) "Do Husbands and Wives Pool Resources?: Evidence from the UK Child Benefit", *Journal of Human Resources* 32:463-80.

[28] MANSER.M AND M.BROWN (1980) "Marriage and Household Decision Theory - A Bargaining Analysis", *International Economic Review* 21: 21-34.

[29] MANSKI.C.F (1993) "Identification of Endogenous Social Effects: The Reflection Problem", *Review of Economic Studies* 60: 531-42.

[30] MCELROY.M AND M.HORNEY (1981) "Nash-bargained Decisions: Towards a Generalization of the Theory of Demand", *International Economic Review* 22: 333-49.

[31] MUNSHI.K AND ROSENZWEIG.M (2005) Why is Mobility in India so Low? Social Insurance, Inequality, and Growth, mimeo Brown University.

[32] QIAN.N (2006) Missing Women and the Price of Tea in China: The Effect of Sex-Specific Earnings on Sex Imbalance, mimeo Brown University.

[33] RANGEL.M.A (2005) Alimony Rights and Intra-household Allocation of Resources: Evidence from Brazil, University of Chicago, Harris School Working Paper Series 05.5.

[34] RANGEL.M.A (2006) Allocation of Resources Within Extended Family Households, mimeo University of Chicago.

[35] RAUT.L.K AND L.H.TRAN (2005) "Parental human Capital Investment and Old-Age Transfers from Children: Is it a Loan Contract or Reciprocity for Indonesian Families", *Journal of Development Economics* 77: 389-414.

[36] ROSENZWEIG.M AND O.STARK (1989) "Consumption Smoothing, Migration, and Marriage: Evidence From Rural India", *Journal of Political Economy* 97: 905-26.

[37] RUBALCAVA.L, G.TERUEL, AND D.THOMAS (2004) Spending, Savings and Public Transfers Paid to Women, CCPR Working Paper.

[38] SCHULTZ.T.P (1990) "Testing the Neoclassical Model of Family Labor Supply and Fertility", *Journal of Human Resources* 25: 599-634.

[39] SCHULTZ.T.P (2004) "School Subsidies for the Poor: Evaluating the Mexican *Progresa* Poverty Program", *Journal of Development Economics* 74: 199-250.

[40] SKOUFIAS.E (2005) Progresa and Its Impacts on the Human Capital and Welfare of Households in Rural Mexico: A Synthesis of the Results of an Evaluation by IFPRI, Research Report 139, Washington D.C.: International Food Policy Research Institute.

[41] THOMAS.D (1990) "Intra-household Resource Allocation: An Inferential Approach", *Journal of Human Resources* 25: 635-64.

[42] TOWNSEND.R (1994) "Risk and Insurance in Village India", *Econometrica* 62: 539-91.

[43] WASSERMAN.S AND K.FAUST (1994) *Social Network Analysis: Methods and Applications*, Cambridge: Cambridge University Press.

Table 1: Descriptive Statistics on Surnames, by Surname Type

Mean, standard errors in parentheses, percentages in brackets

	Head's Paternal Surname	Head's Maternal Surname	Spouse's Paternal Surname	Spouse's Maternal Surname
	(F1)	(f1)	(F2)	(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	9.50×10^{-6}	7.54×10^{-6}	8.60×10^{-6}	8.33×10^{-6}
	(5.48×10^{-6})	(4.16×10^{-6})	(4.95×10^{-6})	(4.95×10^{-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 2: The Number of Family Links, by Family Tie

Couple Headed and Connected Households

Mean, standard deviation between village in parentheses, standard deviation within village in brackets

	Parent	Children Aged 0-16	Adult Children	Siblings	All
		Outside of the Household and in Village			
From head of household to:	.461	-	.652	2.23	3.34
	(.094)		(.487)	(.601)	(1.01)
	[.837]		[2.09]	[3.00]	[3.93]
From spouse of household to:	.250	-	.652	1.63	2.54
	(.083)		(.487)	(.611)	(1.01)
	[.656]		[2.09]	[2.90]	[3.90]
		Inside of the Household			
From head of household to:	.062	3.23	.821	.033	4.15
	(.033)	(.223)	(1.83)	(.037)	(.310)
	[.271]	[2.20]	[3.00]	[.256]	[2.92]
From spouse of household to:	.018	3.23	.821	.013	4.09
	(.049)	(.223)	(1.83)	(.019)	(.299)
	[.139]	[2.20]	[3.00]	[.131]	[2.90]

Notes: The sample is restricted to couple headed households that can be tracked in the first and third waves of *Progresa*, that have at least one extended family tie present in the village. We define the head of the household to be the male among the couple. The standard deviations between and within villages are calculated taking account of the fact that there are unequal number of households across villages. By construction, the number of family links to parental households is always two conditional on such a family link existing. By construction, the number of children of the couple inside and outside of the household are identical for the head and the spouse. Adult children are defined to be at least 17 years of age.

Table 3: Determinants of Extended Family Links

Couple Headed Households

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

Type of Family Link:	Inter-generational Family Links			Intra-generational Family Links				
	Parents of Head	Parents of Spouse	Adult Child	Brothers of Head	Sisters of Head	Brothers of Spouse	Sisters of Spouse	Any Link
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	.811
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 4: Family Network Descriptives

Means, standard deviation between villages in parentheses, standard deviation within villages in brackets

Network Level	Size of family network	Network size/number of households in village	Diameter
Mean	7.76	.167	2.45
Standard deviation between villages	(9.65)	(.149)	(1.18)
Standard deviation within villages	[11.3]	[.153]	[2.11]

Household Level	Degree	Mean Distance	Central closeness
Mean	4.39	2.93	.507
Standard deviation between networks	(2.05)	(.686)	(.205)
Standard deviation within networks	[4.02]	[1.11]	[.189]

Notes: The upper panel reports descriptive statistics on the family network as a whole. There is one observation per family network so that each network has the same weight irrespective of the number of households within it. The size of the network is the number of households in the network. The diameter of the networks is the longest distance between two households that exists in a network. We define two households that are directly connected to be of distance one to each other. The lower panel reports descriptive statistics on the family network at the household level. There is one observation per household and so family networks that are larger will have more weight associated to them. The degree of a household is the number of family links a household has. The mean degree is the mean of degrees across all network members. The central closeness is the inverse of the sum of the distances from household i to all other households in the locality, multiplied by the number of members of the network minus one. This index ranges from zero to one.
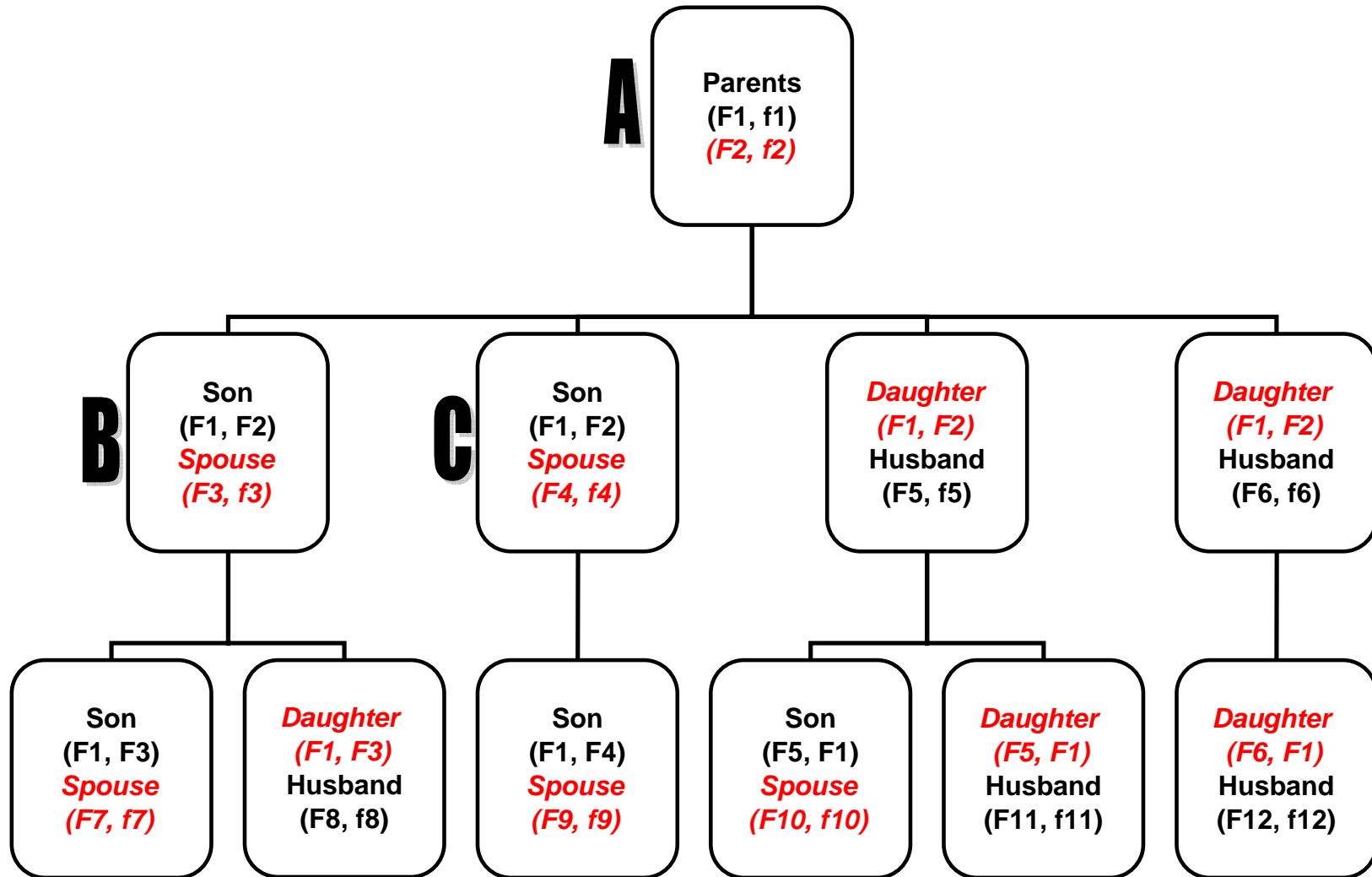
Table 5: Similarity of Characteristics Within Families, by Family Link

Mean, standard errors in parentheses clustered by family network

|	Random matching	Within family	Adult Child to Parent Inter-generational Links		Same Gender Intra-generational Links	
			Son to Parent	Daughter to Parent	Brothers	Sisters
Poverty index [correlation]	.307	.413	.467	.414	.456	.414
Prob[eligible i = eligible j]	.581	.622	.565	.529	.642	.648
	(.002)	(.006)	(.012)	(.017)	(.007)	(.011)
Prob[eligible i=1 | eligible j=0]	.441	.433	.541	.619	.406	.467
	(.006)	(.009)	(.015)	(.020)	(.011)	(.016)
Prob[eligible i=1 | eligible j=1]	.601	.661	.767	.812	.679	.716
	(.006)	(.011)	(.017)	(.021)	(.011)	(.015)

Notes: Standard errors are clustered at the family network level throughout. These numbers are calculated using one observation per pair of households in each village.

Figure 1: Family Tree

A **Parents**
(F1, f1)
(F2, f2)

B **Son**
(F1, F2)
Spouse
(F3, f3)

C **Son**
(F1, F2)
Spouse
(F4, f4)

Daughter
(F1, F2)
Husband
(F5, f5)

Daughter
(F1, F2)
Husband
(F6, f6)

Son
(F1, F3)
Spouse
(F7, f7)

Daughter
(F1, F3)
Husband
(F8, f8)

Son
(F1, F4)
Spouse
(F9, f9)

Son
(F5, F1)
Spouse
(F10, f10)

Daughter
(F5, F1)
Husband
(F11, f11)

Daughter
(F6, F1)
Husband
(F12, f12)

Notes: We use the convention that the head's surnames are written in standard (black) font, and those of his wife are written in (red) italics. Paternal surnames are indicated in upper case (F1, *F2*) and maternal surnames are indicated in lower case (f1, *f2*). First names are not shown as they are not relevant for the construction of extended family ties. Each household in the family tree is assumed to be couple headed purely to ease the exposition.

Figure 2: Distribution of Surnames, by Surname Type, For the Most Common Surnames

Head's Paternal Surname

Head's Maternal Surname

Spouse's Paternal Surname


Spouse's Maternal Surname
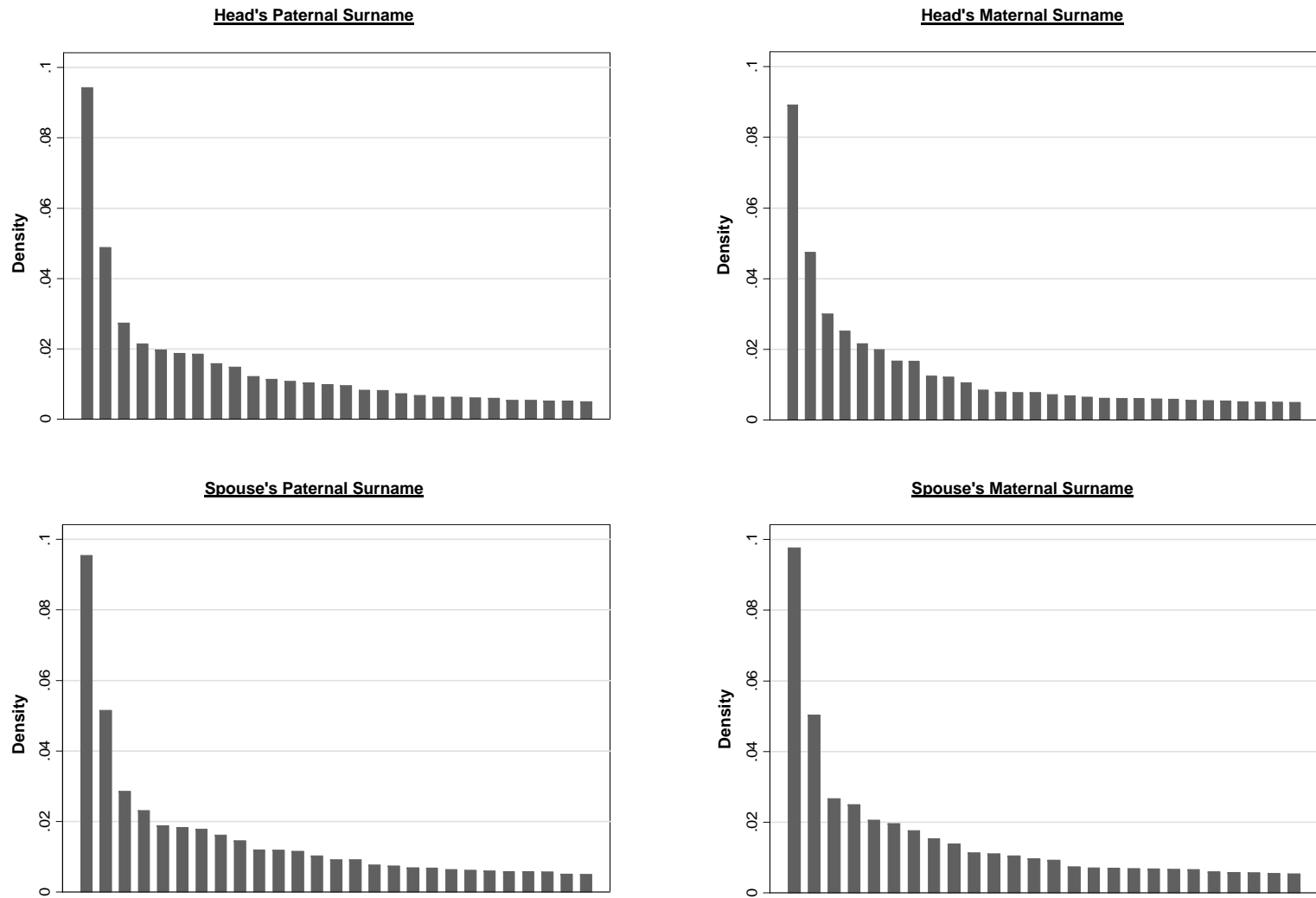
Notes: Each figure is plotted for those surnames that cover at least .5% of the population. The surnames themselves are not shown to ensure confidentiality.

Figure 3: Surnames and Village Size

A. Prob (same surname in locality) by surname type and number of households in the village
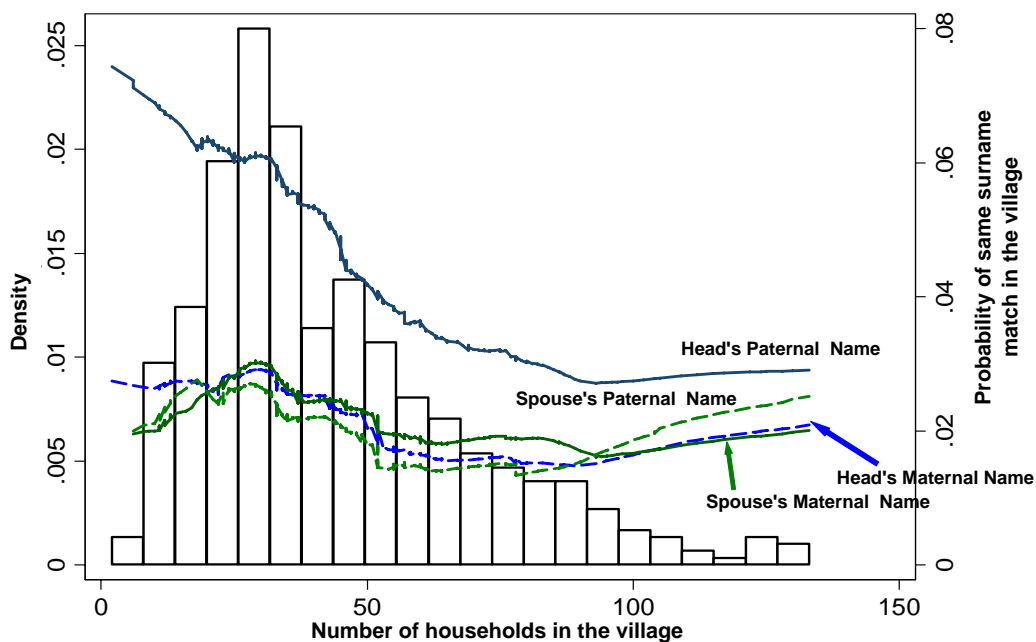
## B. Expected number of same surname matches by surname type and number of households in the village
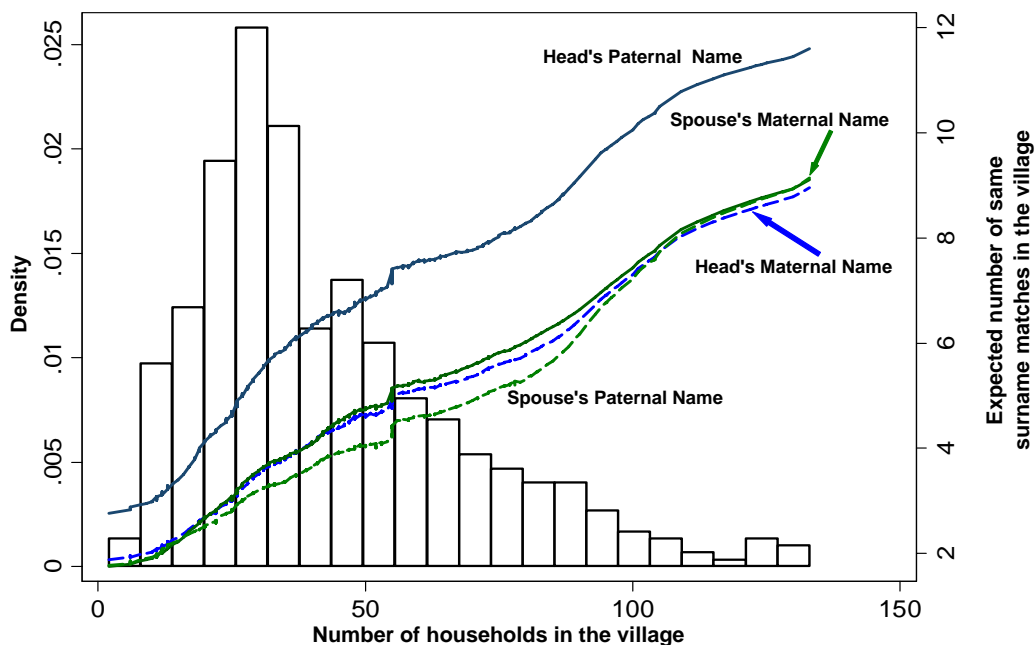
Notes: Figure A shows, for each surname type, a locally weighted regression of the probability of the same surname match in the village on the village size, which is defined as the number of households in the village. Figure B shows, for each surname type, a locally weighted regression of the expected number of same surname matches on the village size. The background of each figure shows a histogram of the village size. Both locally weighted regressions are estimated using a bandwidth of .8. To ease exposition, the largest .5% of villages are dropped from both figures.
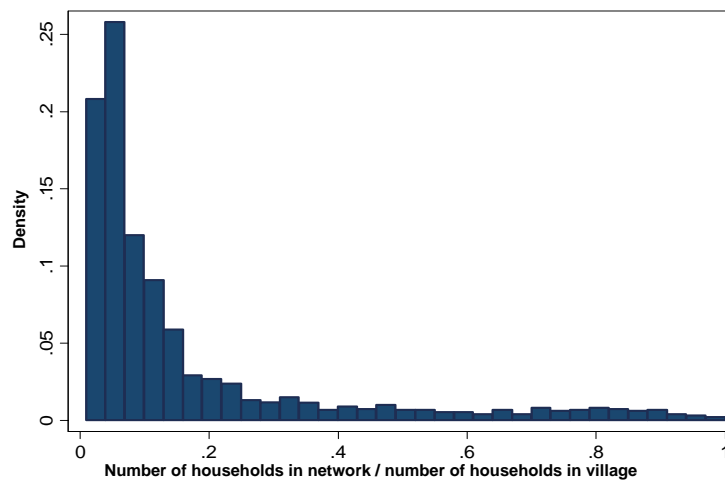
Figure 4: Family Network Descriptives

A. Number of Family Networks in the Village
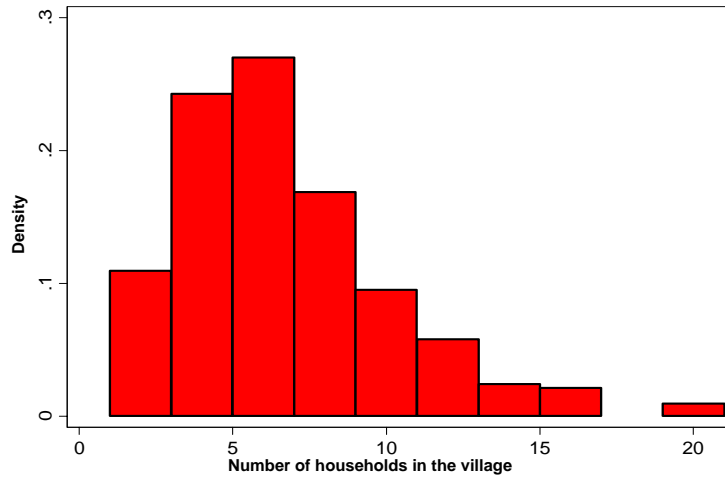
### B. Size Distribution of Family Networks
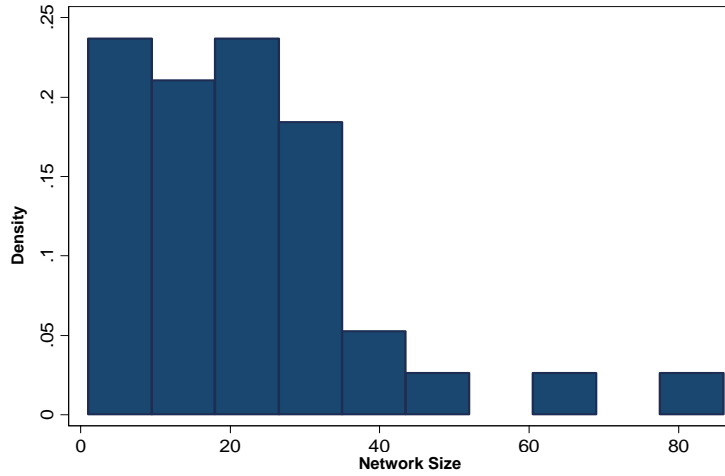
### C. Family Network Size as a Ratio of Number of Households in the Village

Notes: Each figure is constructed from those family networks with at least two households in them. Of the baseline sample of 22553 households that can be tracked over the first three waves of *Progresa*, 17030 (75.5%) of them are within family networks with at least two households.

Figure 5: Family Network Graphs, at Median Village Size
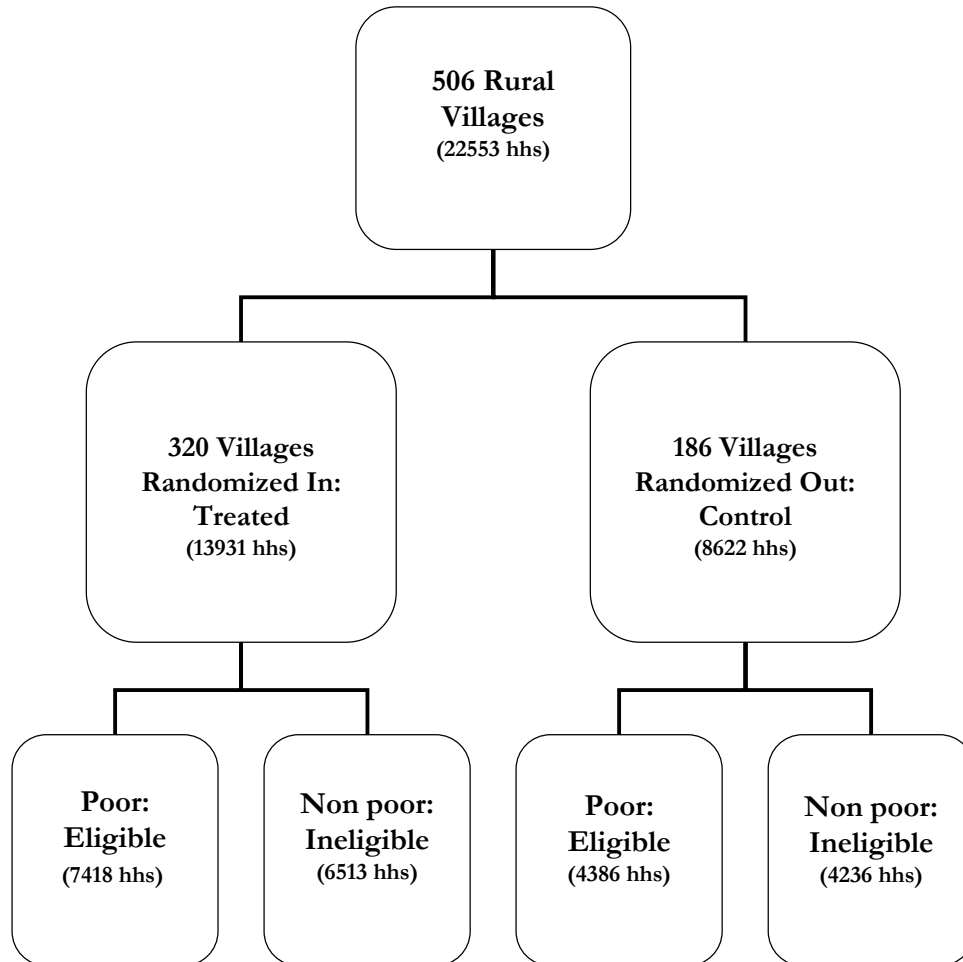
A. Disperse Village
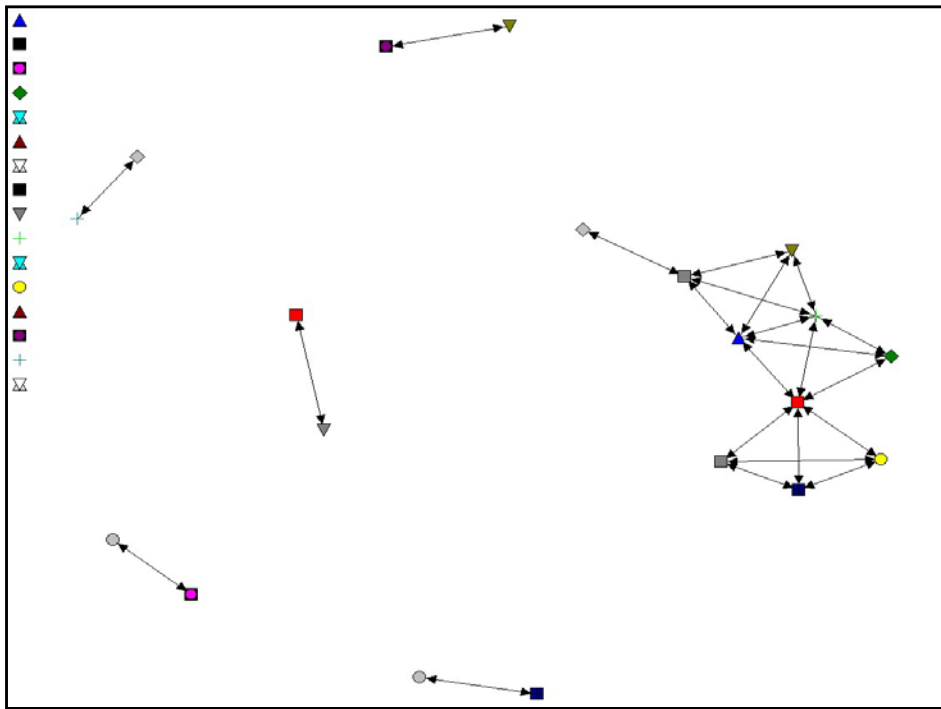
B. Interconnected Village
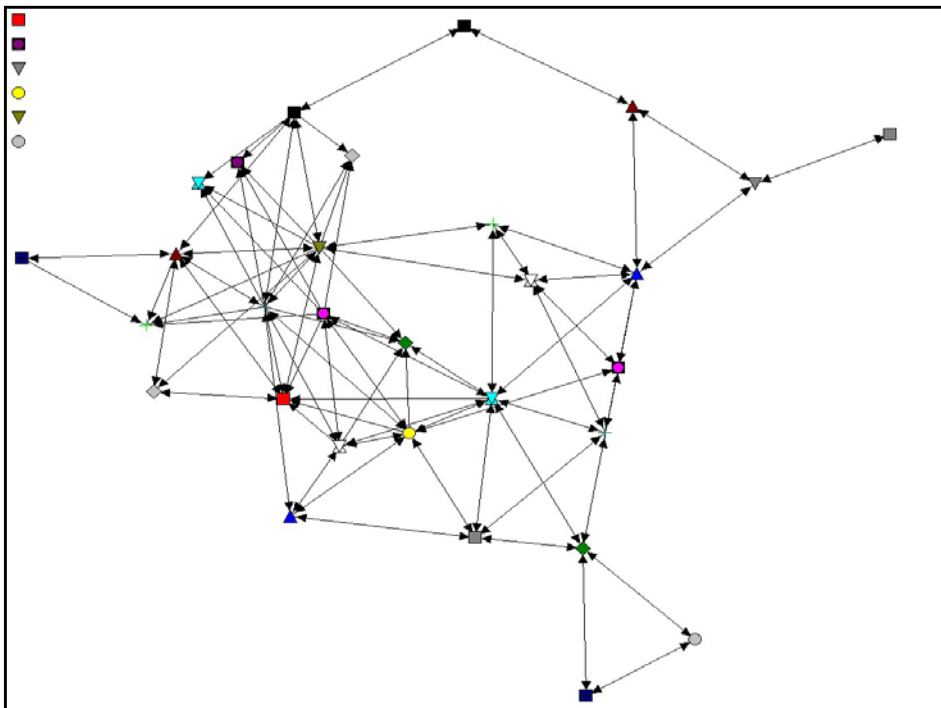
Notes: The two villages shown in Figures A and B have the same number of households in them. The number of households in each is 36, which is the median village size in the *Progresa* data. Each node represents a household. Each link between households correspond either to a parent/child link, a child/parent link, or a sibling link. Single node households that are not linked to any other households are shown in the top left hand corner of each graph. The figures are generated using UCINET.

Figure A1: Progresa Research Design

506 Rural Villages
(22553 hhs)

320 Villages Randomized In: Treated
(13931 hhs)

186 Villages Randomized Out: Control
(8622 hhs)

Poor: Eligible
(7418 hhs)

Non poor: Ineligible
(6513 hhs)

Poor: Eligible
(4386 hhs)

Non poor: Ineligible
(4236 hhs)

Notes: The sample is restricted to those households that can be tracked for at least the first three waves of the *Progresa* data, 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.