DISCUSSION PAPER SERIES

IZA DP No. 4499

Village Economies and the Structure of Extended Family Networks

Manuela Angelucci
Giacomo De Giorgi
Marcos A. Rangel
Imran Rasul

October 2009

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Village Economies and the Structure of Extended Family Networks

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

Discussion Paper No. 4499
October 2009

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA Discussion Paper No. 4499
October 2009

ABSTRACT

Village Economies and the Structure of Extended Family Networks*

This paper documents how the structure of extended family networks in rural Mexico relates to the poverty and inequality of the village of residence. Using the Hispanic naming convention, we construct within-village extended family networks in 504 poor rural villages. Family networks are larger (both in the number of members and as a share of the village population) and out-migration is lower the poorer and the less unequal the village of residence. Our results are consistent with the extended family being a source of informal insurance to its members.

JEL Classification: J12, O12, O17

Keywords: extended family network, migration, village inequality, village marginality

Corresponding author:

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

* This research was supported by an IRB approval from the University of Chicago. The paper has been screened to ensure no confidential information is revealed. We thank the editor, Gary Solon, and two anonymous referees for suggestions that have helped improve the paper. We also thank Oriana Bandiera, Martin Browning, John Ermisch, and Alfonso Miranda for useful comments. All errors remain our own.

1 Introduction

Village economies in developing countries are characterized by missing or imperfect formal credit and insurance markets. Yet, there is a high need for insurance, caused by large income fluctuations and by limited income diversification. In this context, informal institutions emerge as resource-sharing devices. The extended family, that is, the set of related households living in the same village, might be one such institution.

There are good reasons to focus on extended family networks as an important informal institution. First, evolutionary biology suggests preferences are defined over the family dynasty. Second, there are specific inter-generational investments – such as those into children's education, bequests, and the choice of marriage partners – that have no counterpart in relationships between friends, and provide long run incentives for family members to reciprocate in implicit contractual arrangements. Third, the costs of transacting within the family network are likely low if its members have easy access to information about their relatives and possess mechanisms to punish relatives that renege on agreements [La Ferrara, 2003].

There is a large literature on risk sharing in developing countries. This literature shows that the village as a whole typically does not reach a Pareto efficient allocation of resources [Townsend 1994, Ligon 1998, Dercon and Krishnan 2000]. At the same time, there is stronger evidence of efficient risk-sharing within ethnic groups [Deaton 1992, Udry 1994], sub-castes [Mazzocco and Saini 2005, Munshi and Rosenzweig 2005], and family and friends [Rosenzweig 1988, Fafchamps and Lund 2003, La Ferrara 2003, Cox and Fafchamps 2007, Fafchamps 2008]. There is also evidence that interpersonal loans, transfers, and labor exchange arrangements occur primarily within long-standing interpersonal networks such as the extended family [Ellsworth 1985, Lucas and Stark 1985, Krishnan and Sciubba 2004]. Extended families affect also their members' inter-generational transfers [Cox and Jakubson 1995, Altonji *et al* 1997, La Ferrara 2003, Behrman and Rosenzweig 2006] and their children's education [Weiss and Willis 1985, Loury 2006]. The transfers between relatives may be crowded out by public transfers [Albarran and Attanasio 2003, Cox and Fafchamps 2007].

In this paper we explore the relationship between the structure of extended family networks and the characteristics of the local village economy. There are good reasons to believe such factors should be empirically correlated, and these are driven by the ability of households to more easily share resources with others they are related to, than with other households all else equal.

In particular, on one hand the need for resource-sharing is greater in poorer or more marginal villages. In more marginal villages, that might be so defined because they are geographically remote, we expect there to be larger networks, or more dense networks, because the costs of forming links outside the village is higher. This might be exacerbated because informal contracts are more difficult to fulfil when income is very low and hence the extended family acts as an effective contract enforcement institution [Coate and Ravallion 1993].[1]

[1] Gertler *et al* [2006] provide one important counter example to this – using data from Indonesia they find that

On the other hand, the establishment of resource sharing family networks may help to smooth consumption and increase investment, hence endogenously improving the wealth in the village economy. If this effect dominates overall, we expect to observe larger and denser family networks in less marginal villages.

There are also mechanisms relating village level inequality and family network structures. For example, the gains from informal risk sharing increase as the incomes of the sharing group become less correlated. The level of village inequality may be generated by different income profiles of the households, and could be correlated with differential sources of risk. On the one hand, this implies family networks should not be geographically concentrated, as they will then be more likely to be subject to common risks, i.e. weather or pests. It is therefore possible that risk is smoothed across villages through the strategic choice of marriage partners for children, and in particular daughters [Rosenzweig and Stark 1989], and out-migration [Winters *et al* 2001, Davis *et al* 2002, Munshi 2003, Angelucci 2006].[2]

On the other hand, as the geographic dispersion between network members increases, the transactions costs of maintaining interpersonal links, monitoring the behavior of others in the network, and providing non-monetary transfers such as labor exchange, also increase. In addition, as the likelihood of out-migration for a given individual increases, this may reduce the likelihood that other individuals will want to enter cooperative arrangements with them [Schiff 2002]. Lastly, credit constraints may limit out-migration from poorer villages.

Even if individuals were to never move across villages, marriage markets provide another important mechanism through which village inequality and family network structures in the same village relate to each other. If matching in the marriage market is assortative in terms of family resources, inequality leads to less integrated villages as households marry within wealth classes. Inequality within villages then becomes self perpetuating as family networks do not integrate with each other over time [Kremer 1997, Fernández *et al* 2005, Ermisch *et al* 2006].[3]

To summarize, there are many mechanisms through which the marginality of the village economy and inequality within it, relate to the family network structures that are observed within the village. The theoretical predictions on how these factors interplay with each other are neither well established empirically, nor clear cut theoretically. This paper is a first attempt to document these relationships and establish which effects dominate overall. We view this as the starting point of a

the presence of family ties, or other forms of social capital, has little effect on the ability of households to insure consumption against one-off health shocks.

[2]In developed countries, Bentolila and Ichino [2006] provide evidence from four OECD countries consistent with the view that family support and the Welfare State substitute each other in mitigating the consequences of unemployment shocks. Alesina and Giuliano [2007], using data from 70 countries in the World Value Survey, show that strong family ties imply more reliance on the family as a provider of goods and services, and less reliance on the market and on the government for social insurance.

[3]In addition, to the extent that village inequality generates distinct group identities, this can affect the accumulation of social capital and provision of local public goods [La Ferrara 2003]. Theoretically, the effect of inequality on public goods provision is again ambiguous and depends on the characteristics of the public good [Olson 1965, Bardhan *et al* 2002, Banerjee *et al* 2001]. Again this can affect the demand for extended family members to remain co-located in the same village in order to provide public goods to each other.

broader research agenda.

We provide evidence from rural Mexico on how the village level of development and inequality relate to the structure of extended family networks in terms of size and characteristics of the links. Our two key variables are: (i) an index of the development or marginality of the village designed by the Mexican federal government; (ii) an index of household's permanent income, from which we construct the Gini coefficient for each village. This coefficient proxies the extent of within-village inequality in the distribution of income, assets, and land. These two indices capture long run features of the economic environment in each village.

We use household data from the evaluation of *Progresa*, a social assistance program in rural Mexico. These data have two key features that make them ideal for the purposes of this project. First, the data are a census of 504 villages and therefore provide information on all their residents. Second, they contain people's last names. We combine information on the paternal and maternal surnames of heads of households and their spouses with the dual-surname convention in Mexico to obtain a complete mapping of the within-village extended family networks, covering around 22,000 households and over 130,000 individuals.[4]

We consider two types of relatives: related households from different generations, which we call inter-generational links, such as those from the head and spouse to their parents, and to their adult sons and daughters; related households from the same generation, labelled intra-generational links, such as those from the head and spouse to their brothers and sisters.

We present three main findings. First, as conjectured, poorer villages have larger family networks both in terms of number of households embedded and as a share of total village population. These findings are consistent with insurance motives driving the formation of family networks. All else equal, a larger network is better able to smooth away idiosyncratic shocks, and therefore produce a higher level (possibly full) of risk sharing.

Second, we document that both intra *and* inter-generational ties from households are more likely to remain co-resident within the same village, the poorer and the more equal the village. This applies equally to the blood relatives of the head of household *and* their spouse. Hence all households within the family network – irrespective of how they are related – appear to be part of informal insurance arrangements.

Finally, we provide evidence that decisions over household's *permanent* location choices are correlated with the underlying features of the village economy in which their extended family reside. In particular, permanent out-migration is more likely to occur from more unequal villages. One possible explanation for this finding is that family networks are less important when there is enough idiosyncratic income variation within a village. In this case migrants might provide insurance against village level shocks. This is further supported by evidence that these migrants do indeed send remittances back to their extended family. On the other hand, temporary migration does not appear to be correlated with village inequality.

Overall, the results support the notion that in poorer and more equal villages, the benefits of

[4] In a similar spirit, Guell *et al* [2007] use surnames in Spain to document patterns of intergenerational mobility.

geographically dispersed family networks – such as being subject to different weather related risks – do not offset the costs of the network being dispersed – such as increased transactions costs of enforcing cooperative agreements within the network.

These findings have important implications for the design and evaluation of policy interventions. For example, targeting the poorest villages implies targeting mainly households that are part of an extended family. The few households with no extended family in the village may be less likely to benefit from such interventions.

When evaluating such targeted interventions and thinking through the likely consequences of scaling up, our results suggest two factors need to be borne in mind. On the one hand, if there are complementarities between the way in which policy interventions are designed and the presence of the extended family, then any measured behavioral responses to the policy needs to be decomposed into that attributable to the policy itself and that due to the presence of the extended family. Without such an analysis, the expected program effects as the policy expands to less marginal villages are likely to be overestimated because in those locations family networks are less dense. On the other hand, if transfers take place within extended families, then policy responses may be muted in the most marginal villages where any transfers are shared within the extended family. Ignoring such effects would lead to underestimating the true policy effects as the program expands to less marginal locations.

A natural question that arises from this analysis is whether, ultimately, the presence of extended family members influences household behavior and outcomes. Indeed we see this paper as part of a broader research agenda building on previous work on consumption [Altonji *et al* 1992], intergenerational transfers [Cox and Jakubson 1995, Altonji *et al* 1997, La Ferrara 2003, Behrman and Rosenzweig 2006], and children's education choices [Loury 2006]. For example, Angelucci *et al* [2007] show that, when households are offered a 66% subsidy to send their children to secondary school, only households that are part of a connected family, hence able to pool resources, increase school attendance. Households eligible for this partial subsidy, but without relatives in the village, do not increase secondary school enrollment.

2 Data

2.1 The Progresa Evaluation Data

We use household data collected to evaluate the *Progresa* social assistance program in rural Mexico. The data are a census of 504 villages, whose 24,000 households are interviewed every six months between October 1997 and November 1999.[5] The data we use are collected in November 1998, about 6 months after the beginning of the program, or before its start. In this way, none of our key variables are likely influenced by the presence of the program, which was initially offered only

[5] The villages are located in seven states – Guerrero, Hidalgo, Michoacan, Puebla, Queretaro, San Luis Potosi, and Veracruz – and are mainly located in central Mexico.

in a random subset of villages.

Villages were selected on the basis of a village marginality index constructed in 1995 from information on – (i) the share of illiterate adults in the village; (ii) the share of dwellings without water, drainage systems, electricity, and with floors of dirt; (iii) the average number of occupants per room in households; (iv) the share of the population working in the primary/agricultural sector; (v) health and school infrastructures present in the village; (vi) distances from other villages. The index is defined so that it increases as the village becomes more marginal or worse off. In order to more easily interpret the effect of village marginality on family networks, we standardize the marginality index across all villages so that it has mean zero and standard deviation one.[6]

The village marginality index provides a summary statistic characterizing average levels of poverty within the village, the vulnerability of the population to income shocks, and the integration of the village economy into local markets and the wider economy. All these channels will influence the location choices of households and therefore the structure of extended family networks that reside in any given village, as has long been argued by demographers [Taylor *et al* 1996]. We use this index as a measure of village long-term poverty. From now on we will use village poverty and marginality interchangeably.

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 welfare index. This index is calculated as a weighted average of – (i) household income (excluding children); (ii) household size; (ii) ownership of durables, land, and livestock; (iv) educational attainment; (v) physical characteristics of the dwelling. The index gives relatively greater weight to the correlates of permanent, rather than current, income. The household welfare index, on which the village inequality index is based, is defined for each of the seven regions in the *Progresa* data. We therefore control for permanent differences across regions throughout the empirical analysis.

We use this household welfare index to construct a measure of inequality at the village level – defined to be the Gini coefficient of the welfare index across all households in the village. This summary statistics captures inequality in income, acquired human capital, assets, and land holdings, across a census of households in the village. Again, this captures a number of channels through which the location choices of households are influenced, such as the need for and ability of family members to insure each other.[7]

Table 1 shows the mean and standard deviation of these two indices. By construction, the village marginality index has mean zero and standard deviation one. The Gini coefficient is, on average, 8.34. This is considerably lower than measures of the Gini coefficient for most developing countries as a whole, which are typically based on household's current income. This is to be expected given the coefficient here is – (i) based both on a measure of household's permanent

[6] The marginality index is normalized by region (because poverty lines differ by region) and because the number of villages differs by region, the mean of the index is not identically equal to zero.

[7] The Gini coefficient satisfies four desirable properties of any measure of inequality – the anonymity principle, the population principle, the relative income principle, and the Dalton principle. Ray [1998] provides a detailed discussion of these properties, and a comparison of the Gini coefficient with alternative measures of inequality.

income and so is less likely to reflect transitory shocks to income or wealth; (ii) reflects only within village, and not across village, inequality.

Table 1 also presents information on the average village size, as defined by the number of households in the village. The sampled villages are relatively small so that family networks may be especially important determinants of household behavior and outcomes in this setting.

Figure 1A shows a scatter plot of the village marginality and inequality indices. The two are not much correlated – the correlation coefficient between the two is -.043. This helps us to separately identify the relationship between each index and the structure and composition of extended family networks. Figure 1B shows the two measures of the village economy are not much related to village size. The correlation in the village marginality (inequality) index and village size is -.151 (.009). In some cases, this finding eases concerns that we spuriously attribute any mechanical correlation between village size and networks to these features of the economic environment.[8]

2.2 Constructing Extended Family Links

We map the relations between households in the same village matching surnames, which are provided in the third wave of data. Mexicans use *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 to provide the – (i) given name; (ii) paternal surname; and, (iii) maternal surname, for each household member. Hence couple-headed households have four associated surnames – the paternal and maternal surnames of the head, and the paternal and maternal surnames of his wife.[9]

Figure 2 provides an illustration of the matching algorithm. To define each family link, we use information on *two* of the four surnames. 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.[10]

[8] If however the indices have a causal effect on networks and village size also affects networks and these two effects offset each other, then the low correlation between the indices and village size is not as suggestive.

[9] The precise wording of the question in Spanish is, "*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*". We cleaned the surnames data as follows – (i) we removed non-alphabetical characters, replaced "Sin Apellido" (no surname) with missing values, and corrected some obvious typos based on intra-household surname checks; (ii) we imputed a small number of missing female surnames from wave 2; (iii) we verified surnames using the same information from wave 5, and verified the relationship to the household head using wave 1 data. No information on surnames is available in the first wave of data. The head of household is originally defined to be the main income earner. In a very small number of cases the head of a couple headed household is reported to be a women. To keep the exposition clear, we redefine the head to be male in such cases.

[10] 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, $F1$ corresponds to the family name and $F2$ corresponds to the spouse'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** ($F2$). Parent-daughter relationships can be similarly defined. Moreover, intra-generational family ties between siblings can also 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.

In Figure 2 we assume households are couple headed solely to ease the exposition. To deal with the 15% of households that are single headed we use information on the gender of the head to accurately define each family link. Finally, we impose the following restrictions when defining family links – (i) inter-generational links 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) intra-generational links exist when the individuals have at most 30 years age difference.[11]

In presenting descriptive evidence on extended family networks as a whole, we use information from both single- and couple-headed households that are surveyed in the first and third waves of the data. For the later econometric analysis we restrict attention to the 18,979 couple-headed households, as this allows us to compare the extended family links of heads and their spouses. In the Appendix we discuss the forms of potential measurement error that can occur using our matching algorithm and provide evidence on the likely prevalence of these errors. With these caveats in mind, it is important to reiterate that all links are defined across households on the basis of *two* surname matches and the fact that these two surnames map precisely to two of the four surname types. For the empirical analysis, we only exploit information on whether such family links exist, and not the number of links.

One concern is that extended family networks might be formed through assortative matching and inheritance, and this might influence the measure of village inequality. In the Appendix we provide more detailed descriptive evidence on the distribution of surnames in villages. We later present evidence on the distribution of wealth within family networks.[12]

[11] There are of course limits to which information on surnames can be used to construct family ties. Consider links from household i to a single headed household j. As Figure 1 shows, 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 resident together. This is because this particular family link is identified using information from household j on the paternal surnames of both the head and spouse. However this is unlikely to be a major issue in practice. For example we note that female widows aged above 40 are 37% more likely to live as a dependent within a household, rather than head their own household, relative to a similar married woman. These single parents are then recorded in the household roster. The fact that most unmarried elderly women in Mexico reside with kin is documented in the demography literature [De Vos 2002].

[12] See Guell *et al* [2007] for more evidence and tests of assortative matching by surname using data from Spain.

3 Descriptive Evidence

3.1 Village Characteristics

Table 2 reports the share of households within a village that have each type of family link. 22% of couple-headed households are isolated, in that they have no extended family members resident in the village. The incidence of not having extended family members in the same village is therefore at least as high as the incidence of single headedness – the causes and consequences of which have been much documented in the literature – which affects 15% of households.

The remaining columns show that 16.4% of households have ties between the head of household and that of his parents – this is around double the number of parental links from the spouse to her parents in the village. The same inter-generational pattern occurs when we consider links from the head and spouse to their adult children – households are around twice as likely to have at least one son resident in the village than at least one daughter.

Intra-generational ties are more prevalent than inter-generational ties. The most likely type of extended family tie to exist is that between brothers – 45.2% of all household heads have at least one brother in the village. In contrast, only 23% of households have ties from the spouse to her sister.[13]

3.2 The Number of Extended Family Links

Table 3 documents the number of specific family links each household has to others in the village. These are reported for connected households, namely those households that are embedded within an extended family network in the village and so have at least one family member resident in the village. As we focus on couple headed households, we are able to report each type of link from the head and spouse separately.

The first column shows that the number of parents present outside the household and within the village is higher for the head than for his spouse. The number of links to adult children are, by construction, identical for head and spouse. On average, respondents have .443 sons that head their own household in the village, and .209 daughters. In terms of sibling links, heads again have more links to their siblings than do their spouses. In total, heads of household have blood ties to 3.34 others in the same village, and their spouse is connected to 2.54 other households. Given that by construction links to adult sons and daughters are the same for heads and their spouses, this implies that on average, a household is directly connected to 5.23 other households in the village.

Three further points are of note. First, Table 3 decomposes the variation in each statistic into that between and within villages. While in most cases there is more variation within than between villages, the proportion of the variation that is attributable to cross village differences tends to be

[13] A greater share of households have ties from the head of household to a spouse of another household (.285) than from the spouse to the head (.262). This reflects the fact that we restrict attention to ties from couple-headed households to both couple- and single-headed households. These figures thus highlight that single-headed households are more likely to be female headed, as expected.

higher for female defined links than for male links. This might reflect the fact that there is more sorting over locations by women in response to economic conditions, as we document throughout our analysis.

Second, we can also use the household roster to measure family ties that are coresident. When doing so we note that parents are significantly more likely to reside outside of the household than within it. On average, the head (spouse) has .06 (.02) parents resident with him (her). In other words, there are at least ten times as many parents outside the household and within the village, than are coresident with respondents. Given respondents' ages, there are many more young children coresident inside the household than adult children outside the household. Sibling links are also many times more likely to be outside the household, which is as expected given the age of respondents. Heads have more siblings links than their spouses, and this is again true for links both inside and outside the household.[14]

Third, the evidence in Table 3 shows that the majority of extended family links are those of male heads of household. A key source of this difference arises from women moving households and villages at the time of marriage. To shed light on this, we exploit data on spouse's marital history. 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. The key difference between spouses with and without parents resident in the village is that 85% of spouses that have their parents present in the village report remaining in the same village at the time of marriage. The figure for spouses that have no parental links in the village is only 61%. In line with the evidence from rural Indian in Rosenzweig and Stark [1989], these descriptives hint that in rural Mexico, parents may choose to marry their daughters into distant villages as a consumption smoothing device.

To provide external validity to the links, the Appendix presents similar information from the *Mexican Family Life Survey,* that was collected in rural areas over a comparable time period.

3.3 Extended Family Networks

We now use tools from social networks analysis to describe the extended family networks as a whole. There are 2196 family networks, covering 17,030 households that may be either single- or couple-headed.[15]

Figure 3A shows there are around five to ten different family dynasties within the same village. Figure 3B shows the size distribution of family networks. As reported in Table 4, there are on average 7.8 households within the same family network. Since a household is has 5.23 immediate

[14] In October 1998 the average age of heads (spouses) among couple headed households is 45.0 (40.5).

[15] To understand when two households are defined to be in the same family network, consider a 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. Households i and j are said to be of distance one from each other and households i and k are of distance two 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.

relatives (or distance-one relatives), as shown in Table 2, each household has two to three ties of distance two or more in the village. About 17% of households within a given village are part of the same family network, and Figure 3C shows the overall variation in the share of village households that are part of the same dynasty.

Taken together, this descriptive evidence highlights that most villages have multiple family networks. That is, it is not the case that all households become connected to all other villagers. This is the case despite the sampled villages being of relatively small size. This finding is consistent with papers studying endogenous networks formation. These paper consistently show that, even within a small local economy, multiple networks arise in equilibrium [Genicot and Ray 2005, Bramoullé and Kranton 2008, Karlan *et al* 2009].

Table 4 provides further information on statistics that describe the family network. The average distance between any two households in the same network is 2.12. The degree of the average network, defined as the number of households each is directly connected to, is 2.07. This is smaller than the average number of links each household in the sample has, as reported in Table 3, because in Table 4 there is one observation per family network and all are equally weighted irrespective of the number of households that are embedded within the network. This highlights that there are many family networks that comprise only two households, as shown in Figures 3B and 4. The diameter of the network, defined as the largest distance d_{ij} between any two households i and j in the network, is around 2.5 on average. Hence family networks are unlikely to span across more than three generations, as expected.

The table also decomposes the variation in each statistic into that between villages and that within the same village. We note there is considerable variation arising from both dimensions of the data. Hence there are considerable differences in the structure of family networks in different villages, and in the network structure of family dynasties that reside in the same village.

The final column presents evidence on the welfare index of households within the same network. As this is defined by region, we standardize the measure within region to ease comparisons across regions. As all households in the family network are defined to be in the same village and hence region, by construction this has mean zero. There is considerable variation in household welfare within the same extended family network – this variation is almost as great as that across villages. This variation rules out a strong positive assortative matching in marriage markets by family wealth and may be potentially related to the need for establishing informal insurance arrangement within the extended family.[16,17]

To emphasize how the structure of family networks can differ between villages, Figure 4 provides a graphical representation of family networks in two villages of median size, corresponding to 36

[16] We can also test for assortative matching in wealth by computing the standard deviation of the poverty index used to determine program *Progresa* eligibility for each family network and each village. If there is positive (negative) assortative matching then the ratio of network over village standard deviations in poverty, $\frac{sd^n}{sd^v}$, should be less (more) than one. The computed ratios are centered around one, rejecting the hypothesis of assortative matching.

[17] This finding is slightly different from the results found in Guell *et al* [2007] using surnames in Spain where they document that richer Catalans have less frequent surnames.

households. The upper panel shows a village with a relatively dispersed set of family networks. There are five families in which there are only two households present, one large, ten-household family network, and 16 households without family ties to any other household in the village. In contrast the lower panel shows a very inter-connected village, with one large family network containing 30 households, and 6 households with no relatives in the village. In the remainder of this paper we shed light on whether and how such villages differ in their family network structures because of the underlying features of the village economy as captured by the village marginality and inequality indices.

4 The Relationship Between Economic Characteristics and Extended Family Networks

4.1 Village Level Analysis: Extended Family Networks

We now estimate the relationship between characteristics of the village economy on the number and structure of family networks. To begin with, we consider how the village marginality index in village v in region s (M_{vs}) and the village inequality index (Q_{vs}) correlate with the number of family networks in the village (N_{vs}). We estimate the following OLS regression at the village level,

$$N_{vs} = \alpha_s + \beta M_{vs} + \gamma Q_{vs} + \delta S_{vs} + u_{vs}, \tag{4.1}$$

where α_s denotes region fixed effects that capture differences across family networks due to the geographic location of the village. We also condition on the size of the village (S_{vs}) as there is expected to be a positive correlation between village size and the number of family dynasties present within it. We compute robust standard errors throughout. We report the estimates in Table 5.

Column 1 shows there are a greater number of family networks in villages that – (i) are less marginal and so better off; (ii) have greater levels of inequality, as measured by the Gini coefficient for permanent income; (iii) are larger, as measured by the number of households within them, as expected.

Columns 2 to 4 show how M and Q relate to the size of the largest family network, looking at – (i) the absolute number of network members, (ii) the share of households in the village that are part of the largest family network, and, (iii) the share of households that are part of any family network. Conditional on region, – the size and the village share of the largest network are significantly higher the higher the village marginality (or poverty) and the higher the village equality. Moreover, there is a significantly greater share of households with relatives in villages that are more equal, although this variable is has no significant correlation to village poverty.[18]

[18]These results are qualitatively robust to using an alternative measure of inequality within the village such as the coefficient of variation or Theil index – with a wide range of sensitivity parameters – based on the household

The results are consistent with the idea that the local economy interplays with the structure of the extended family. More marginal villages are both poorer and more isolated from the wider economy. Therefore its residents have a high need for insurance but a limited scope for sharing risk with geographically distant households. In this context, it is not surprising that the local family network is bigger. That is, in marginalized villages the benefit of having a dispersed family, which ensures the income shocks are not strongly correlated, is probably more than offset by the costs of the network being dispersed – such as increased transactions costs of enforcing cooperative agreements within the network. In a related way, in more unequal villages – where the ability of households to insure each other for example is greater – family networks are smaller.

The most intuitive way in which to benchmark the magnitude of the effects is in terms of the number of households in the largest network. Taking the estimates from our preferred specification in Column 2 and the descriptive evidence from Table 1 reveals that – (i) in villages with a marginality index larger by one standard deviation, the largest family network comprises 1.45 more households; (ii) a one standard deviation decrease in the village inequality index is associated with 1.02 more households in the largest family network; (iii) the magnitude of the effect of the marginality index (inequality index) is equivalent to 2.01 (1.41) more households in the village *per se*, relative to a mean village size of 44.7 households. This emphasizes that these features of the village economy are both qualitatively and quantitatively important correlates of the residents' family network structures.[19]

Since we are comparing networks in a cross section of villages, it is impossible to identify with complete certainty whether network structures endogenously respond to these characteristics of the village economy, or whether the features of the local economy are themselves shaped by the nature of extended family networks within them.[20] The next subsection shows how the underlying features of the village economy relate to the location choices of specific extended family members, conditional on household characteristics. This helps ease concerns over the results merely reflecting that households sort into villages on the basis of the characteristics of the village economy.

---

welfare index.

[19] All the results in Table 5 are qualitatively unchanged if we allowed village size to have a quadratic effect on each outcome. We have also explored whether there are non-linear effects of village size using a quadratic specification. The evidence suggests such an effect exists for all the outcomes except for the size of the largest family network. Exploring how this varies across villages of different size, we find that any non-linear effect is driven predominantly by villages above the median village size (36 households). More importantly, the magnitude of the effect is quantitatively small and so is not economically significant.

[20] Reltedly, one concern is that, as the extended family network data is based on information collected six months after *Progresa* was implemented, the program might have caused endogenous changes in networks in those villages that were randomly selected to receive *Progresa*. There is however little evidence of *Progresa* causing differential migration between treatment and control locations in the first few months of its operation, as measured by changes in the number of households, or permanent migrants. As a check, we estimated (4.1) using data only from the 184 control villages. As expected the estimates become more precise but of the 12 coefficients shown in Table 5 - (i) none change sign in this sub-sample; (ii) of the 11 significant coefficients, 9 remain significant at conventional levels; (iii) the two coefficients that become insignificant are on the village marginality index in Columns 2 and 3; these are insignificant because the standard error rises in this smaller sub-sample, the point estimates are almost unchanged (1.28 and .028 respectively); (iv) the one insignificant coefficient in Table 5 remains insignificant in this sub-sample as expected.

4.2 Household Level Analysis: Extended Family Links

In order to present the evidence on the relation between the underlying village economy and the structure of the family network we need to control for possible confounders at the household level.

This analysis sheds light on whether, conditional on household characteristics, features of the village economy continue to have robust correlations with the structure of extended family networks. In particular we establish the precise inter and intra-generational family links that bind members of each network across village economies with differing underlying characteristics.

To do this we define a dummy variable, L_{jhvs}, set equal to one if household h in village v in region s has at least one type-j link to another household in the same village, and zero otherwise. We can define define the following eight types of family link j – intra-generational family links, i.e. from the head (spouse) of the household to his (her) brothers and sisters; inter-generational family links, i.e. from the head (spouse) of a household to his (her) parents, and from the head and spouse of the household to their adult sons and daughters. We therefore estimate the following specification for whether household h in village v in region s has at least one type-j link to another household in the same village,

$$L_{jhvs} = \alpha_s + \beta M_{vs} + \gamma Q_{vs} + \delta S_{vs} + \lambda X_{hvs} + u_{jhvs}, \tag{4.2}$$

where X_{hvs} are household characteristics, and all other controls are as previously defined. An econometric concern with this specification is that the presence of a given link type-j is likely to be influenced by the presence of other link types, j', so that in effect we need to estimate a linear equation for each j. We therefore use seemingly unrelated regression (SUR) methods to estimate these j-equations allowing for cross-equation correlation in the error terms [Zellner 1962]. The results are reported in Table 6.[21]

As discussed in the introduction, a large literature has explored the mechanisms driving the presence of extended family members. We use this to help guide our choice of which characteristics to condition on in X_{hvs}. This allows us to then explore whether features of the village economy are related to the presence of extended family members over and above such household observables.

We condition on the following controls in X_{hvs}, as defined in the first wave of household data collected in October 1997. First, there is a mechanical correlation between the age of the head and spouse and the likelihood that their parents and adult children are in close proximity so we control for the ages of the head and spouse. To capture any ethnic differences across families, we control for whether the head and spouse speak an indigenous language. We also condition on the

[21] A concern that arises with this specification is that the dependent variable is a share and so lies between zero and one, which is not accounted for by the SUR estimates. In relation to this we note that the mean of the dependent variable lies well within the zero-one interval, as shown in Table 2. Moreover, we also experimented with estimating this system of eight equations for each inter and intra-generational link using a multivariate probit regression framework. This uses the method of simulated maximum likelihood based on the Geweke-Hajivassiliou-Keane (GHK) algorithm. The structure is therefore similar to the SUR framework except that explicit account is taken of the binary nature of each dependent variable. We found the results to be qualitatively similar using the two methods although the SUR estimates can be computed in considerably less time.

number of members resident in the household.

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 are more likely to own land and if land markets are missing, the ability to inherit land, or to acquire land specific human capital, may lead adult children to be more likely to remain within the village than otherwise [Foster and Rosenzweig 2002]. We therefore control for the years of completed schooling of the head and spouse, whether the head and spouse are working, whether the household owns its own home, any land, any livestock, whether the house has floors of dirt, and the household welfare index. Once this set of characteristics is controlled for, the village marginality index captures variation due to largely exogenous factors such as the remoteness of the village from urban centres and its level of integration into local markets.[22]

Table 6 reports the results. We first focus on the results in Columns 1 to 4 on whether households have inter-generational extended family links co-resident in the same village. We see that in nearly all cases, such inter-generational links are more likely to be co-resident in more marginal villages, and villages with less inequality.

The second set of results in Columns 5 to 8 focus on the presence of intra-generational links, namely siblings, within the same village. Again in most cases, siblings are more likely to reside in another household within the same village if the village is more marginal and more equal. As expected, the likelihood a household has any given type of link-j is always significantly increasing in the size of village.

Taken together with the previous results at the village level on the network characteristics as a whole (Table 5), the evidence points to the reason why family networks are larger in more marginal and more equal villages, being that extended family members, that are intra *and* inter-generationally linked to any given household, are more likely to reside within the same village, all else equal. Moreover, this applies equally to the blood ties of the head as well as his spouse. Finally, the results highlight that the structure of family networks relates to features of the village economy as a whole, over and above the characteristics of individual households within them.[23]

In the final Column, we use OLS to estimate how the local structure of the family network – in particular the density of the network – varies with characteristics of the village economy. The standard measure used in social networks analysis to measure local density is the local clustering

[22] We do not control for literacy because it is highly correlated with years of schooling – 89% (90%) of heads (spouses) have no formal schooling if they are illiterate.

[23] Two further points are of note. First, the set of households in Table 6 includes both connected and isolated households. The results are robust to restricting the sample to only those 78% of households that are embedded within a family network. Second, although not the focus of this paper, we note that the correlations between household characteristics and the presence of family ties have intuitive signs. For example, the mechanical correlations with age are as expected with older heads and spouses being less likely to have their parents outside of the household and resident in the village, and significantly more likely to have their adult children in other households in the village. Heads and spouses with more years of schooling are more likely to have their parents present. Finally, whether the head and spouse speak an indigenous language 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.

coefficient [Wasserman and Faust 1994]. This measures how many links from a household have links to each other, divided by the maximum number of links they could potentially have had. This statistic therefore lies between zero and one. For example, a local clustering coefficient of one implies all the relations of household i are directly related to each other. A local clustering coefficient of zero implies none of the family links of household i are directly related to each other. We have constructed the local clustering coefficient for all households embedded in extended family networks and then estimated how this correlates to characteristics of the village economy. The result in Column 9 of Table 6 shows that in poorer and more unequal villages, extended family networks are denser, all else equal.

In conjunction with the earlier SUR regression results, this suggests that in poorer and more unequal villages, specific familial ties are less likely to be present *per se*, but those that remain are more closely related.

It is natural to interpret these results in terms of the features of the village economy that drive extended family members to migrate away from the village. There are two key channels we identify. First, the evidence shows that as the village becomes better off and less marginal, extended family members are more likely to migrate to other locations. Given our measure of marginality and the fact that we condition on household characteristics including the household welfare index, the results imply the costs of out-migration from marginal villages is high. This is as expected given that more marginal villages are geographically remote and less integrated into local markets. Hence lower mobility in more marginal villages is likely to be caused by more binding credit constraints to finance migration.[24]

In addition, the demand for informal insurance is likely to be greater in more marginal villages where even small fluctuations in household income can cause large falls in welfare. The demand for extended family members to co-locate in the same village in such environments is likely to be great given the transactions costs of implementing insurance agreements among kinship groups are likely to be lower than with other households, all else equal.

The second channel we identify is that in more unequal villages, extended family members are more likely to migrate away from the village, all else equal. In such economic environments, the gains from informal risk sharing contracts increase as the income profiles of households differ and are subject to different sources of risk. On the one hand, this implies family networks should not be geographically concentrated, as they will then be more likely to be subject to common weather related risks, and because there may exist alternatives to informal insurance agreements with family members. On the other hand, as the geographic dispersion between network members increases, the transactions costs of maintaining interpersonal links, monitoring the behavior of others in the network, and providing non-monetary transfers such as labor exchange, also increase.

[24] That out-migration is negatively related with marginality is also consistent with the existing empirical evidence on migration which emphasizes the household level determinants of migration. Indeed, a pattern of migration consistent with the existence of credit constraints at the household level is documented by Angelucci [2006] using the *Progresa* data. More broadly, the empirical literature has established an inverse-U shaped correlation between the propensity to migrate and household wealth [Massey 1988, Faini and Venturini 1993, Hatton and Williamson 1998].

Our results suggest the former effect dominates overall – extended family networks become more geographically dispersed as village inequality rises.[25]

The results have implications for thinking through the location choices specifically of adult daughters and sons, which have received a lot of attention in the literature. First, they raise the possibility that marriages of daughters are used as strategic devices to geographically diversify risk where there is a high degree of village inequality. This is also consistent with the evidence from rural India presented by Rosenzweig and Stark [1989]. The results also support the notion that in more equal villages, there are fewer incentives for households to marry their daughters outside of the village, because they are able to assortatively match in the marriage market that exists within their own village.

The location choices of males are driven in a similar way – they are more likely to out migrate in less marginal villages again presumably because the costs of doing so are lower. At the same time, economic forces that tend to tie adult sons to the same location as their parents – such as the non-tradeability of labor exchange across villages, or the nature of household production that is tied to specific plots of land within the village – appear to be less binding when the village is better off and more equal [Rosenzweig 1988].[26]

4.3 Network Level Analysis: Migration

To complement the previous results, we now directly exploit information from the October 1997 data on whether any members of the extended family network have migrated away, and whether they have sent remittances conditional on migrating. We distinguish between permanent and temporary migration, where the former is asked about all household members who left in the previous 5 years the latter asks about the number of household members who left on a seasonal basis in the year before the survey.

Migration rates are higher among household that are part of family networks: while on average 3% (19%) of households report at least one permanent (seasonal) migrant, 16% (44%) of extended

[25] We also checked whether extended family networks are shaped by the incidence of natural shocks to the village. We did so using information from community surveys on the prevalence of natural shocks that afflicted the village such as droughts, floods, frosts, fires, pests, earthquakes, and hurricanes. We found no evidence that the incidence of natural shocks to the village relates to the extended family structures that reside within them. This is not surprising if individuals can anticipate the average level of such shocks in any location. Hence while the vulnerability of a location to naturally occurring shocks may well drive the long run location choices of extended family members, realizations of such naturally occurring shocks in the short run have little impact on the long term structure of family networks.

[26] We also explored whether the features of the village economy have heterogeneous effects across households on the formation of family networks. To do so we use the classification of households into poor and non-poor designed to determine eligibility for *Progresa* transfers. This classification is made in 1997 on the basis of the household poverty index. We then estimate (4.2) allowing the effects of M_{vs} and Q_{vs} to differ between those 52% of households classified as poor, and the remaining 48% classified as non-poor. We find no robust evidence that, conditional on household characteristics, the features of the aggregate village economy have heterogeneous effects across relatively poor and relatively non-poor households. Hence it is not the case that the location choices of only the most or least wealthy households within a family dynasty are affected by the economic environment. Indeed, as Table 4 shows, there remains considerable variation in the welfare index across households within the same family network in the same village.

family networks report at least one migrant. This is consistent with the existing literature on the importance of networks in migration decisions [Munshi 2003].

We estimate the following function of the probability of having at least one migrant in network n, village v and region s,

$$\text{Prob}(Y_{nvs} = 1) = \Phi(M_{vs}, Q_{vs}, S_{vs},\ N_{nvs}, \alpha_s,) \tag{4.3}$$

The dummy Y_{nvs} is one if at least one member of network n has permanently (temporarily) migrated away. M_{vs}, Q_{vs}, and S_{vs} denote the village marginality index, village inequality index, and village size. N_{nvs} denotes characteristics that describe the network as a whole as discussed in Section 3.2 – network size, average distance between any two households in the network, network degree and diameter, and the average welfare index across households within the network, which captures the migration incentives and constraints for those within the network. The region dummies, α_s, capture geographic differences in the propensity to migrate. We cluster the standard errors at the village level.

Table 7 presents the marginal effects estimated by probit. Column 1 shows that family networks in more unequal villages have more permanent migrants, consistent with the earlier results. On the other hand, there is no significant correlation between the village marginality and permanent migration likelihood. Column 2 explores whether at least one permanent migrant is reported to send remittances to any household in the family network – as is intuitive, remittances are more likely to be sent back to family networks that reside in more marginal and unequal villages, all else equal.[27] The remaining Columns repeat the analysis for seasonal migration. The seasonal migration likelihood and the likelihood of sending remittances are unrelated to the chosen village characteristics.

One possible interpretation of our results is that residents of unequal villages who can pool resources with other villagers use permanent migration as part of the response to the need for insurance against (village level) income shocks, further consistent with the evidence that these migrants send remittances back.

As a final point, we recognize that family networks can also change over time through fertility choices. to explore this, the final Column of Table 7 presents OLS estimates of fertility as a function of characteristics of the village economy and household characteristics. We see that in less marginal and more unequal villages, fertility rates are higher suggesting larger extended family networks in the future.[28]

[27] This contrasts with the findings of Munshi and Rosenzweig [2005] who provide evidence from India that those who migrate away from their sub-caste lose the services of that network, including mutual insurance arrangements.

[28] In this specification we do not control for the total household size, unlike for the other specification in Table 7. The results are robust to using a Poisson regression model that takes account of the dependent variable being a count.

5 Conclusions and Policy Implications

This paper is a descriptive exploration of the relationship between the local economy - poor rural Mexican villages - and the existence and shape of extended family networks. Using household and community data, we find that more marginalized and less unequal villages have larger family networks. In these villages the permanent migration likelihood is lower.

This evidence suggests that in economic environments characterized by – (i) imperfect markets, correlated income shocks, an absence of formal institutions of contract enforcement; and, (ii) where all households are equally poor, households choose to or have no option but to live close to their extended family members. The benefits of having a geographically dispersed network – such as being subject to less correlated income shocks due to climatic conditions – seem to be outweighed, on average, by the costs of such geographic dispersion – such as increased transactions costs of monitoring and making non-monetary transfers to others in the network, credit constraints that prevent people from migrating, or the low marriage marketability of individuals belonging to poor households.

This paper provides the first large-scale description of the structure of extended family networks in poor rural areas. We advocate incorporating questions on the extended family and other informal institutions into household surveys. There are already some examples in small scale village surveys, providing important insights on the role of social networks in learning about agriculture in Ghana [Conley and Udry 2005], child fostering in Burkina Faso [Akresh 2005], implicit insurance arrangements in Ethiopia [Dercon *et al* 2005], and informal borrowing in Peru [Karlan *et al* 2009].

Incorporating these data into surveys will help us understand the effects of the presence and characteristics of kinship networks on household behavior and welfare, integrating insights from anthropology and sociology [Gillin 1949, Carlos and Sellers 1972, Rothstein 1999] into mainstream economics.

This paper is the first step of a broader research agenda. As such, it has highlighted the importance of understanding both how the extended family is formed and how it affects the provision of insurance and investment capital for its members. In ongoing work in Angelucci *et al* [2009], we use this data at the household level to develop and test a model of insurance and investment within family networks. The basic assumption of the model is that resource transfers are less costly with the family network, than between unconnected households. The model and evidence shows that connected households experience lower variability in their consumption, and are more able to undertake investments, especially if investments are lumpy.

The results from the analysis in this paper have important implications for the design and evaluation of policy interventions. For example, targeting the poorest villages implies targeting mainly households that are part of an extended family. The few households with no extended family in the village may be less likely to benefit from such interventions. In scaling up a policy, our results suggest that its effects could be due to the interaction between family networks and the policy itself. The expected program effects as the policy expands to less marginal villages are likely to be overestimated because in those locations family networks are less dense. On the other hand, if transfers take place within extended families, then policy responses may be muted in the most marginal villages where any transfers are shared within the extended family. Ignoring the latter would lead to underestimating the true policy effects as the program expands to less marginal locations.

6 Appendix

6.1 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 suspicious 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.

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.

6.2 Sorting into Villages by Surname

Table A1 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. 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 partly reflects the patronymic naming convention implies spouse's paternal surnames have lower survival rates across generations than those of male heads of household. The majority of surnames are mentioned at least twice. In the entire sample, the most frequent paternal surname of the head covers 9% of households.

Within a village, the table shows that on average, there are 15.6 unique paternal surnames of the head – with a standard deviation of 9.17. There is a similarly wide number of unique surnames for the other surname types. To assess the degree of sorting by surname into village, we first note that the expected number of households in the *sample* with the same head's paternal surname is 13.3.[29] The next row reports the same information but at the village level. The fact that the expected number of households in the *village* with the same surname is significantly smaller than in the population implies households do not perfectly sort into villages by surname.[30]

6.3 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.[31]

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

[29] 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 expected number of households in the population with name i is $E_i = n_i.\left(\frac{N-1}{N}\right)$. The value reported in Table A1 is the averages of E_i over all surnames i.

[30] 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 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 1 are the weighted average of 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.

[31] As discussed earlier, 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.

spouse are alive is likely to be driven by spouses being younger than their husbands. Moreover, the 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.[32] 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.

References

[1] ALBARRAN.P AND O.P.ATTANASIO (2003) "Limited Commitment and Crowding Out of Private Transfers: Evidence from a Randomised Experiment", *Economic Journal* 113: C77-85.

[2] ALESINA.A AND P.GIULIANO (2007) The Power of the Family, NBER Working Paper 13051.

[3] 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.

[4] 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.

[5] ANGELUCCI.M (2006) Aid Programs' Unintended Effects: The Case of Progresa and Migration, mimeo University of Arizona.

[6] ANGELUCCI.M, G.DE GIORGI, M.A.RANGEL, AND I.RASUL (2007) Family Networks and School Enrolment: Evidence from a Randomized Social Experiment, mimeo University College London.

[7] ANGELUCCI.M, G.DE GIORGI, M.A.RANGEL, AND I.RASUL (2009) Insurance and Investment in Family Networks, mimeo, University of Arizona.

[8] ATTANASIO.O.P AND J.V.RIOS RULL (2003) "Consumption Smoothing and Extended Families", in M. Dewatripont, L. Hansen and S. Turnovsky, (eds.) *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, Econometric Society Monographs, Cambridge, England: Cambridge University Press.

[32] 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.

[9] BANERJEE.A, D.MOOKHERJEE, K.MUNSHI, AND D.RAY (2001) "Inequality, Control Rights and Efficiency: A Study of Sugar Cooperatives in Western Maharashtra", *Journal of Political Economy* 109: 138-90.

[10] BARDHAN.P (1984) *Land, Labor, and Rural Poverty*, New York: Columbia University Press.

[11] BARDHAN.P, M.GHATAK AND A.KARAIVANOV (2002) "Inequality, Market Imperfections, and the Voluntary Provision of Collective Goods", in Baland.J, P.Bardhan and S.Bowles (eds.), *Inequality, Collective Action and Environmental Sustainability*, Russell Sage Foundation and Princeton University Press.

[12] BASU.K (1986) "One Kind of Power", *Oxford Economic Papers* 38: 259-82.

[13] 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.

[14] BENTOLILA.S AND A.ICHINO (2006) "Unemployment and Consumption Near and Far Away From the Mediterranean?", forthcoming *Journal of Population Economics.*

[15] CARLOS.M.L AND L.SELLERS (1972) "Family, Kinship Structure, and Modernization in Latin America", *Latin American Research Review* 7: 95-124.

[16] COATE.S AND M.RAVALLION (1993), "Reciprocity Without Commitment Characterization and Performance of Informal Insurance Arrangements", *Journal of Development Economics*, 40: 1-24.

[17] COX.D AND M.FAFCHAMPS (2007) "Extended Family and Kinship Networks: Economic Insights and Evolutionary Directions", forthcoming *Handbook of Development Economics* Volume 4.

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

[19] DAVIS.B, G.STECKLOW, AND P.WINTERS (2002) "Domestic and International Migration from Rural Mexico: Disaggregating the Effects of Network Structure and Composition", *Population Studies* 56: 291-309.

[20] DEATON.A (1992) "Saving and Income Smoothing in Côte D'Ivoire", *Journal of African Economies* 1: 1-24.

[21] 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.

[22] DE VOS.S (2002) "Kinship Ties and Solitary Living Among Unmarried Elderly Women: Evidence From Chile and Mexico", *Research on Aging* 22: 262-89.

[23] DUFLO.E AND A.BANERJEE (2007) "The Economic Lives of the Poor", *Journal of Economic Perspectives* 21: 141-67.

[24] ELLSWORTH.L (1989) Mutual Insurance and Non-Market Transactions Among Farmers in Burkina Faso", University of Wisconsin, Ph.D. Dissertation.

[25] ERMISCH.J.F, M.FRANCESCONI, AND T.SIEDLER (2006) "Intergenerational Mobility and Marital Sorting", *Economic Journal* 116: 659-79.

[26] FAFCHAMPS.M (2008) "Risk Sharing Between Households", forthcoming in the *Handbook of Social Economics.*

[27] FAFCHAMPS.M AND S.LUND (2003) "Risk-sharing Networks in Rural Philippines", *Journal of Development Economics* 71: 261-87.

[28] FAINI.R AND A.VENTURINI (1993) "Trade, Aid and Migrations. Some Basic Policy Issues", *European Economic Review* 37: 435-42.

[29] FERNANDEZ.R, N.GUNER, AND J.KOWLES (2005) "Love and Money: A Theoretical and Empirical Analysis of Household Sorting and Inequality", *Quarterly Journal of Economics* 120: 273-344.

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

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

[32] GERTLER.P, D.I.LEVINE, AND E.MORETTI (2006) "Is Social Capital the Capital of the Poor? The Role of Family and Community in Helping Insure Living Standards Against Health Shocks", *CESifo Economic Studies* 52: 455-99.

[33] GILLIN.J (1949) "Mestizo America", in R.Linton, (ed.) *Most of the World*, New York: Columbia University Press.

[34] GUELL.M, J.V.RODRIGUEZ MORA, AND C.TELLMER (2007) Intergenerational Mobility and the Informative Content of Surnames, mimeo, University of Edinburgh.

[35] HATTON.T.J AND J.G.WILLIAMSON (1998) *The Age of Mass Migration: Causes and Economic Impact*, New York: Oxford University Press.

[36] KARLAN.D, M.MOBIUS, T.ROSENBLATT, AND A.SZEIDL (2009) "Trust and Social Collateral", *Quarterly Journal of Economics*, forthcoming.

[37] KREMER.M (1997) "How Much Does Sorting Increase Inequality?", *Quarterly Journal of Economics* 112: 115-39.

[38] KRISHNAN.P AND E.SCIUBBA (2004) Endogenous Network Formation and Informal Institutions in Village Economies, Cambridge Working Papers in Economics, 462.

[39] LA FERRARA.E (2003) "Kin Groups and Reciprocity: A Model of Credit Transactions in Ghana", *American Economic Review* 93: 1730-51.

[40] LIGON.E (1998) "Risk-Sharing and Information in Village Economies", *Review of Economic Studies* 65: 847-64.

[41] 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.

[42] LUCAS.R AND O.STARK (1985) "Motivations to Remit: Evidence from Botswana", *Journal of Political Economy* 93: 901-18.

[43] MASSEY.D.S (1988) "Economic Development and International Migration in Comparative Perspective", *Population and Development Review* 14: 383-413.

[44] MAZZOCCO.M AND S.SAINI (2005) Testing Efficient Risk Sharing with Heterogeneous Risk Preferences, mimeo University of Wisconsin.

[45] MUNSHI.K (2003) "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market", *Quarterly Journal of Economics* 118: 549-97.

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

[47] OLSON.M (1965) *The Logic of Collective Action*, Harvard University Press, Cambridge: Massachusetts.

[48] RAY.D (1998) *Development Economics*, Princeton, New Jersey: Princeton University Press.

[49] ROSENZWEIG.M.R (1988) "Risk, Implicit Contracts and the Family in Rural Areas of Low Income Countries", *Economic Journal* 98: 1148-70.

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

[51] ROTHSTEIN.F.A (1999) "Declining Odds: Kinship, Women's Employment, and Political Economy in Rural Mexico", *American Anthropologist* 101: 579-93.

[52] SCHIFF.M (2002) "Trade, Migration and Welfare: The Impact of Social Capital", in H. Singer, N. Hatti and R. Tandon (eds.) *Globalization, Technology and Trade in the 21st Century*, Volume 19 of New World Order Series. B.R. Publishing: Delhi.

[53] TAYLOR.J.E, J.ARANGO, G.HUGO, A.KOUAOUCI, D.MASSEY, AND A.PELLEGRINO (1996) "International Migration and Community Development", *Population Index* 62: 317-418.

[54] TEMPLE.J AND P.A.JOHNSON (1998) "Social Capability and Economic Growth", *Quarterly Journal of Economics* 113: 965-90.

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

[56] UDRY.C (1994) 'Risk and Insurance in a Rural Credit Market: An Empirical Investigation in Northern Nigeria', *Review of Economic Studies* 61: 495-526.

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

[58] WEISS.Y AND R.J.WILLIS (1985) "Children as Collective Goods and Divorce Settlements", *Journal of Labor Economics* 3: 268-92.

[59] WINTERS.P, A,DE JANVRY, AND E.SADOULET (2001) "Family and Community Networks in Mexico-U.S. Migration", *Journal of Human Resources* 36: 159-84.

[60] WOOLCOCK.M AND D.NARAYAN (2002) "Social Capital: Implications for Development Theory, Research, and Policy", *World Bank Research Observer* 15: 225-49.

[61] ZELLNER.A (1962) "An Efficient Method of Estimating Seemingly Unrelated Regression Equations and Tests for Aggregation Bias", *Journal of the American Statistical Association* 57: 348-68.

Table 1: The Village Economy

Mean, standard deviation in parentheses

	Village marginality index (standardized)	Village inequality index (Gini coefficient of household's permanent income in the village)	Village size (number of households)
Village Characteristic	-.003 (.994)	8.34 (1.82)	44.7 (29.2)
Minimum Value	-1.97	2.97	6
Maximum Value	2.97	15.8	222

Notes: There is one observation for each of the 504 villages. The village marginality index is 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 the population working in the primary sector, distances from other villages, and health and school infrastructures located in the village. A higher marginality index corresponds to the village being more marginal (poorer). The household welfare 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. An increase in the index implies the households is less poor. The measure of village inequality is the Gini coefficient of the welfare index of all households in the village. This is scaled to lie between 0 and 100. Village size is defined as the number of households in the village. The village marginality index is standardized across all villages. The villages in the sample cover 7 regions.

Table 2: The Share of Households With Extended Family Links in the Village, by Type of Link

Mean, standard deviation in parentheses

	Any Family Link (Connected)	Inter-generational Family Links				Intra-generational Family Links			
		Parents to Son	Parents to Daughter	Son to Parent	Daughter to Parent	Head to Head (Brothers)	Head to Spouse	Spouse to Head	Spouse to Spouse (Sisters)
Fraction of households in the village with such a family link	.781	.164	.085	.178	.087	.452	.285	.262	.230
	(.156)	(.100)	(.067)	(.108)	(.068)	(.186)	(.173)	(.150)	(.164)
Minimum Value	.16	0	0	0	0	0	0	0	0
Maximum Value	1	.621	.318	.600	.417	.939	.872	.939	.939

Notes: There is one observation for each of the 504 villages. The sample is based on couple headed households that can be tracked over the first and third Progresa waves.

Table 3: The Number of Extended Family Links of Households, by Type of Link

Connected Households

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

	Parents	Adult Children		Siblings		All Extended Family Links	Total Links From Household
	Father and Mother	Sons	Daughters	Brothers	Sisters		
From head of household to:	.461			1.39	.840	3.34	
	(.094)	.443	.209	(.340)	(.325)	(1.02)	5.23
	[.837]	(.399)	(1.95)	[1.87]	[1.52]	[3.92]	(1.53)
From spouse of household to:	.250	[1.44]	[.813]	.910	.725	2.54	[5.47]
	(.260)			(1.23)	(.304)	(1.04)	
	[.656]			[1.71]	[1.51]	[3.89]	

Notes: There is one observation per household. The sample is restricted to couple headed households that can be tracked over the first and third Progresa waves that have at least one extended family member present in the village. An adult child defined is to be at least 17 years old. By construction, the number of links to adult sons and daughters is the same from the head and spouse. The total links from the household is the sum of unique extended family links from the household to others in the same village, and does not therefore double count the children of the head and spouse of the household. The decomposition of the standard deviation into that between and within villages takes account of the fact that the number of households differs in each village.

Table 4: Extended Family Network Descriptives

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

	Network Size	Network Size/Number of Households in Village	Average Distance	Degree	Diameter	Standardized Household Welfare Index
Mean	7.76	.167	2.12	2.07	2.45	-.001
SD between villages	(9.65)	(.149)	(.303)	(1.20)	(1.18)	(.566)
SD within villages	[11.3]	[.153]	[.545]	[1.19]	[2.11]	[.521]

Notes: There is one observation per family network so that each network has the same weight irrespective of the number of households within it. The underlying sample of households is based on couple headed households that can be tracked over the first and third Progresa waves. Of the baseline sample of 22553 couple headed households, 17030 (75.5%) of them are within family networks with at least two households. There are 2196 family networks in total. The size of the network is the number of households in the network. Two households that are directly connected are defined to be of distance one to each other. The average distance is the average over all possible pairs of households within the family network. The diameter of the networks is the longest distance between two households that exists in the network. The household welfare 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. The index is calculated relative to a state norm. Hence we standardize the index within each state. An increase in the index implies the households is less poor. The decomposition of the standard deviation into that between and within villages takes account of the fact that the number of family networks differs in each village.

Table 5: Family Network Structures and the Village Economy

OLS regression estimates, robust standard errors in parentheses

Dependent Variable:	Number of Family Networks	Size of the Largest Family Network	Share of Households in the Largest Family Network	Share of Households that are Connected
	(1)	(2)	(3)	(4)
Village marginality index	-.374***	1.46**	.027**	-.007
	(.117)	(.665)	(.011)	(.007)
Village inequality index	.145**	-.558**	-.017**	-.010**
	(.059)	(.283)	(.007)	(.004)
Village size	.034***	.722***	.003***	.001***
	(.007)	(.050)	(.000)	(.000)
Region fixed effects	Yes	Yes	Yes	Yes
Adjusted R-squared	.189	.749	.154	.125
Observations (village level)	504	504	504	504

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. There is one observation for each village, and each dependent variable is constructed using couple headed households that can be tracked over the first and third Progresa waves in each village. The village marginality index is 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 the population working in the primary sector, distances from other villages, and health and school infrastructures located in the village. A higher marginality index corresponds to the village being more marginal (poorer). The household welfare 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. An increase in the index implies the households is less poor. The measure of village inequality is the Gini coefficient of the welfare index of all households in the village. This is scaled to lie between 0 and 100. Village size is defined as the number of households in the village. The village marginality index is standardized across all villages. The villages in the sample cover 7 regions. Of the baseline sample of 22553 couple headed households, 17030 (75.5%) of them are within family networks with at least two households. There are 2196 family networks in total. Robust standard errors are reported.

Table 6: Extended Family Links and the Village Economy

Columns 1-8: Seemingly Unrelated Regression Estimates

	Links to Other Households in the Same Village								
	Inter-generational Family Links				Intra-generational Family Links				Network Characteristic
Dependent Variable:	Son to Parent	Daughter to Parent	Parents to Son	Parents to Daughter	Head to Head (Brothers)	Head to Spouse	Spouse to Head	Spouse to Spouse (Sisters)	Local Clustering Coefficient
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Village marginality index	.001	.006**	.017***	.018***	.004	.017***	.021***	.018***	-.024**
	(.003)	(.003)	(.003)	(.003)	(.005)	(.004)	(.004)	(.004)	(.010)
Village inequality index	-.005***	-.004***	-.009***	-.004***	-.021***	-.018***	-.014***	-.017***	.014***
	(.002)	(.001)	(.002)	(.001)	(.002)	(.002)	(.002)	(.002)	(.005)
Village size/100	.022***	.046***	.074***	.063***	.133***	.190***	.178***	.171***	-.183***
	(.007)	(.006)	(.007)	(.005)	(.009)	(.009)	(.009)	(.008)	(.028)
Household controls	Yes								Yes
Region fixed effects	Yes								Yes
Observations (household level)	18979								15358

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. The village marginality index is 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 the population working in the primary sector, distances from other villages, and health and school infrastructures located in the village. A higher marginality index corresponds to the village being more marginal (poorer). The household welfare 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. An increase in the index implies the households is less poor. The measure of village inequality is the Gini coefficient of the welfare index of all households in the village. This is scaled to lie between 0 and 100. Village size is defined as the number of households in the village. The village marginality index is standardized across all villages. The villages in the sample cover 7 regions. The following household level characteristics are also controlled for - the husband's age, years of schooling, whether he speaks an indigenous language, and whether has is currently working, the spouse's age, years of schooling, whether she speaks an indigenous language, and whether she is currently working, the household welfare index, the number of members of the household, whether the household owns its home, own land, has dirt floors, and has any livestock. In the SUR specification, the Breusch-Pagan test of independence rejects the null hypothesis that the error terms are uncorrelated in each regression at the 1% significance level. In Column 9 the dependent variable is the local clustering coefficient of the household, and the sample is restricted to households that are embedded in a family network. The specification is estimated using OLS and robust standard errors clustered by village are calculated.

Table 7: Migration and the Village Economy

Columns 1- 4: Probit Estimates, Marginal Effects Reported

Column 5: OLS Estimates

Robust Standard Errors Reported, Clustered by Village

	Permanent Migration		Seasonal Migration		Fertility
Dependent Variable:	Any member permanently left in five years prior to October 1997?	Sent remittances in last year?	Any seasonal migration from household in last year?	Sent remittances in last year?	Number of children aged less than 16 in the household
	(1)	(2)	(3)	(4)	(5)
Village marginality index	.015	.017**	.008	.000	-.220***
	(.010)	(.007)	(.020)	(.019)	(.035)
Village inequality index	.016***	.011***	-.014	-.011	.077***
	(.005)	(.004)	(.009)	(.009)	(.014)
Village size/100	-.000	.000	.001	.000	.052
	(.000)	(.000)	(.001)	(.000)	(.072)
Mean of dependent variable	.157	.106	.444	.394	2.72
Network controls	Yes	Yes	Yes	Yes	No
Region dummies	Yes	Yes	Yes	Yes	Yes
Observations (network level)	2196	2196	2196	2196	
Observations (household level)					18979

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. In Columns 1 to 4 the unit of analysis the extended family network. In Column 5 the unit of analysis is the household. Robust standard errors are reported allowing for clustering at the village level. The village marginality index is 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 the population working in the primary sector, distances from other villages, and health and school infrastructures located in the village. A higher marginality index corresponds to the village being more marginal (poorer). The household welfare 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. An increase in the index implies the households is less poor. The measure of village inequality is the Gini coefficient of the welfare index of all households in the village. This is scaled to lie between 0 and 100. Village size is defined as the number of households in the village. The village marginality index is standardized across all villages. The villages in the sample cover 7 regions. The following network level characteristics are also controlled for - the size of the network, the average distance between any two households in the network, the degree of the network, the diameter of the network, and the average welfare index across households within the network. In Column 5 the dependent variable is the number of children aged 16 or less resident in the baseline. In this specification we do not control for the total household size.

Table A1: 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]
Number of unique names in the village	15.6	20.1	19.5	20.9
Expected number of same surname matches in population	13.3	11.2	9.92	9.26
	(1.66)	(1.36)	(1.25)	(1.19)
Expected number of same surname matches in the village	7.55	5.31	5.42	4.98
	(.039)	(.036)	(.036)	(.040)

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

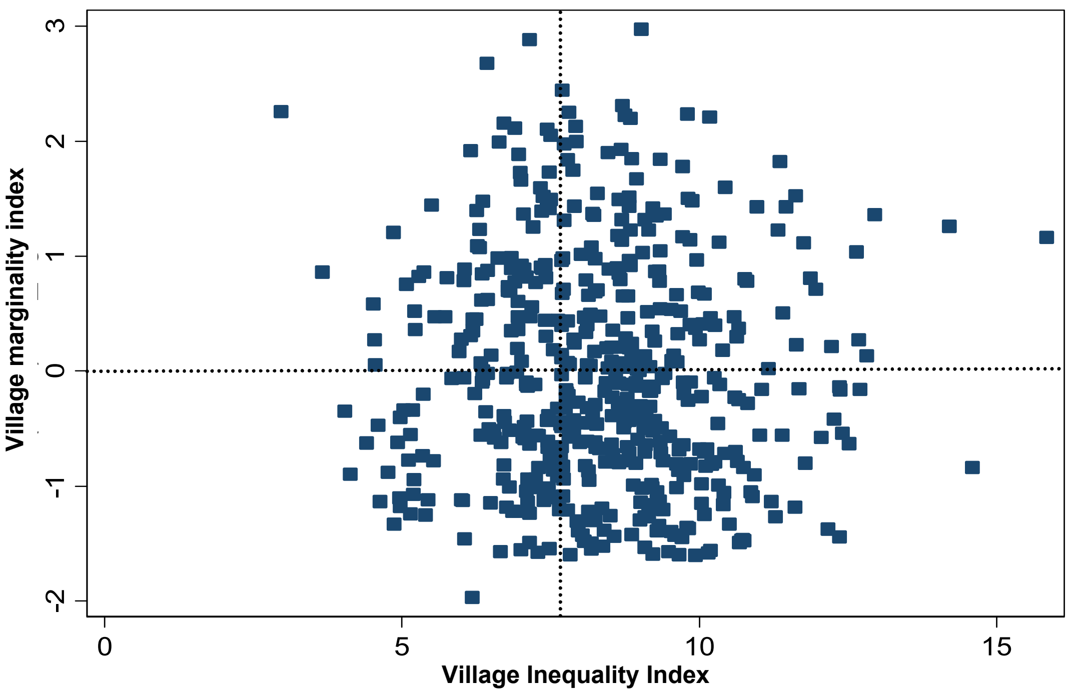
	Outside of the Household (ANY location)				
	Parent	Children Aged 0-16	Adult Children	Siblings	All
From head of household to:	.476	-	1.23	3.27	4.97
	(.035)		(.089)	(.116)	(.014)
From spouse of household to:	.669	-	1.23	3.50	5.39
	(.039)		(.089)	(.113)	(.148)

	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)

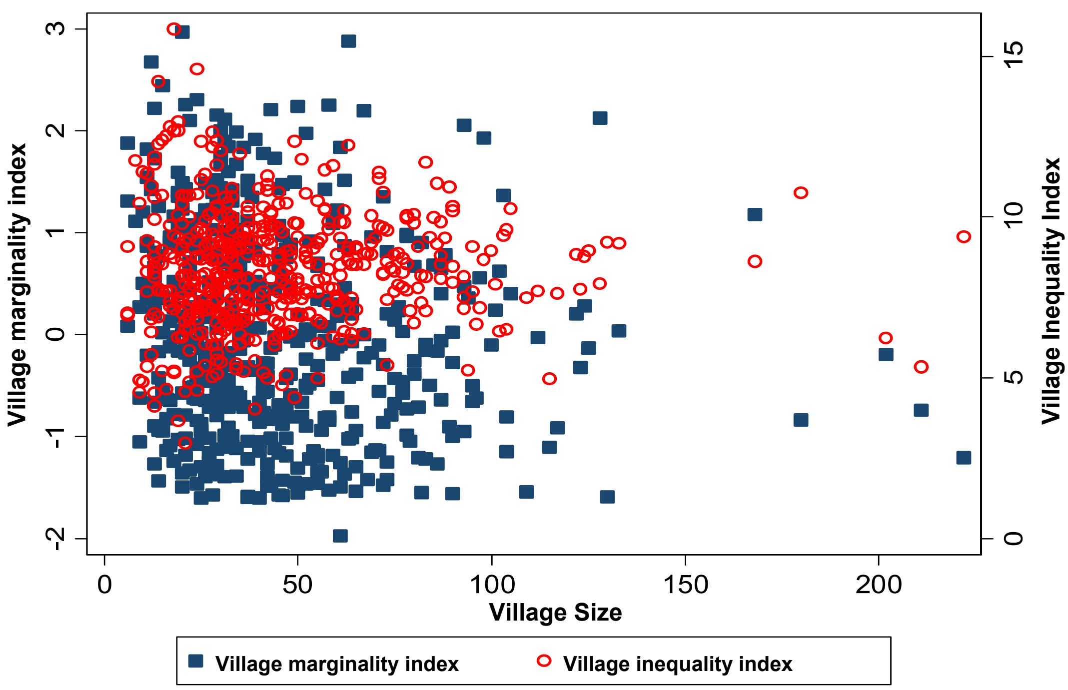
Notes: The sample is taken from the first wave of the Mexican Family Life Survey, 2001. Standard errors are clustered by village. We restrict this sample to the seven Mexican states that are also covered in the *Progresa* evaluation data, and to couple headed households, in locations with less than 2500 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).

Figure 1: The Village Economy

A: Village Marginality and Inequality

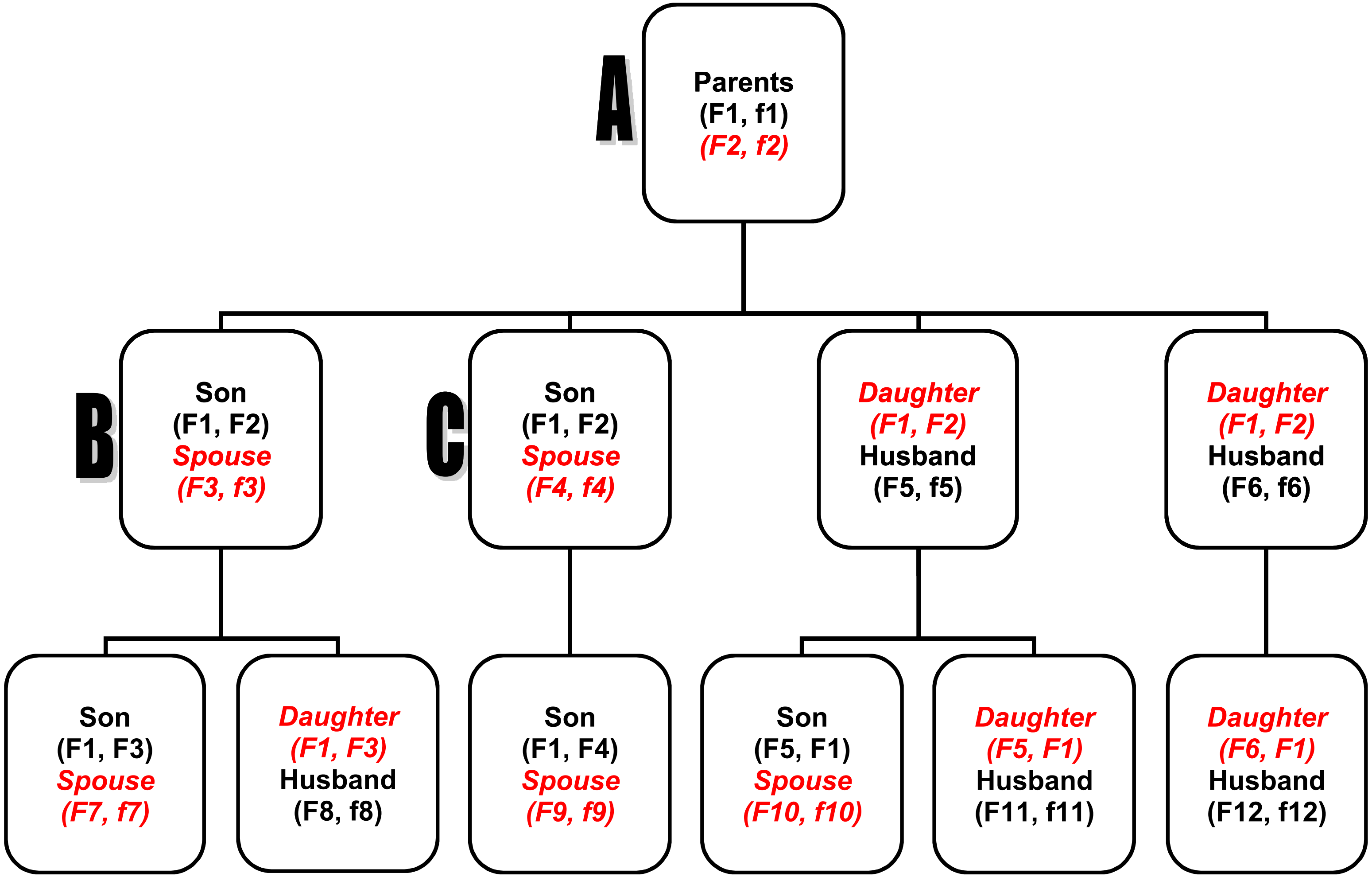


B: Village Marginality, Inequality, and Village Size



Notes: In each figure, there is one observation at the village level. The village marginality index is 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 the population working in the primary sector, distances from other villages, and health and school infrastructures located in the village. A higher marginality index corresponds to the village being more marginal (poorer). The household welfare 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. An increase in the index implies the households is less poor. The measure of village inequality is the Gini coefficient of the welfare index of all households in the village. This is scaled to lie between 0 and 100. Village size is defined as the number of households in the village. The village marginality index is standardized across all villages.

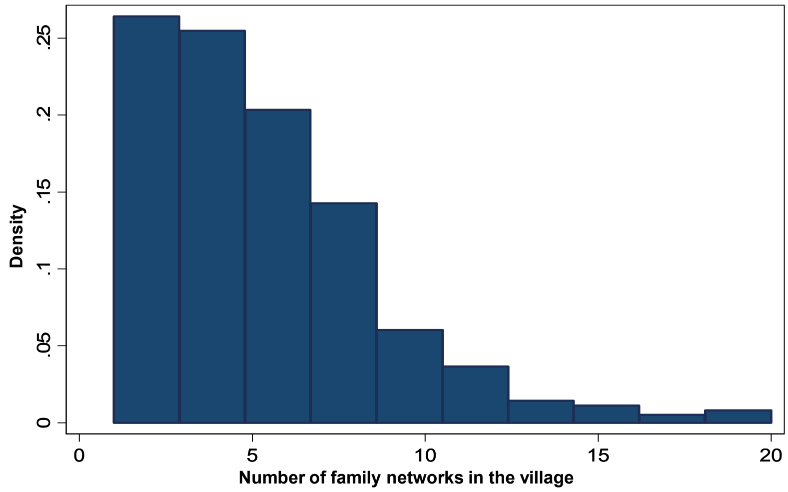
Figure 2: Family Tree



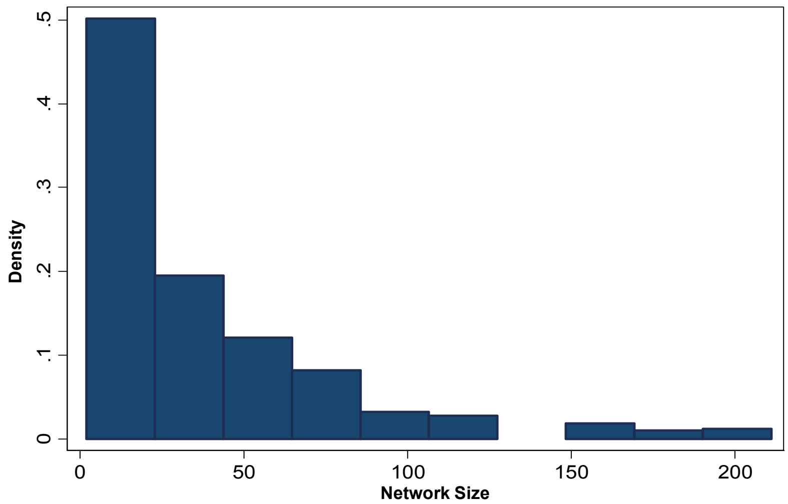
Notes: We use the convention that the head's surnames are written in standard (black) font, and those of his wife are written in (red) italics. Paternal surnames are indicated in upper case (F1, *F2*) and maternal surnames are indicated in lower case (f1, *f2*). First names are not shown as they are 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 3: Family Network Descriptives

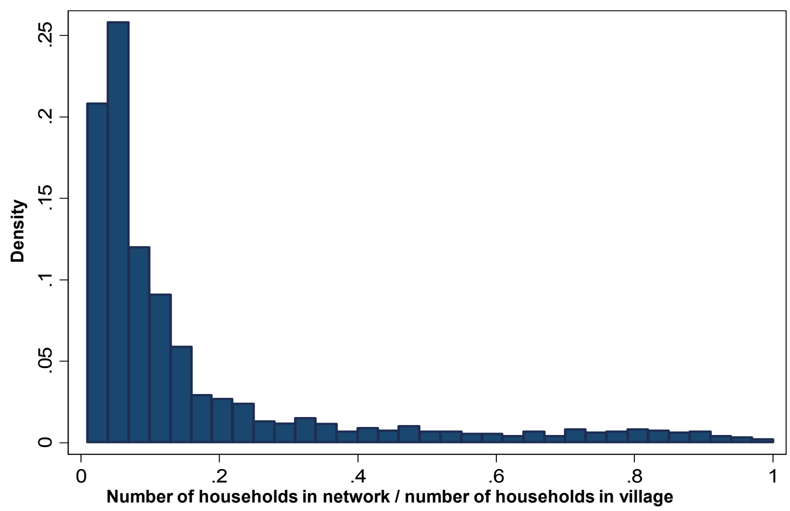
A. Number of Family Networks in the Village



B. Size Distribution of Family Networks



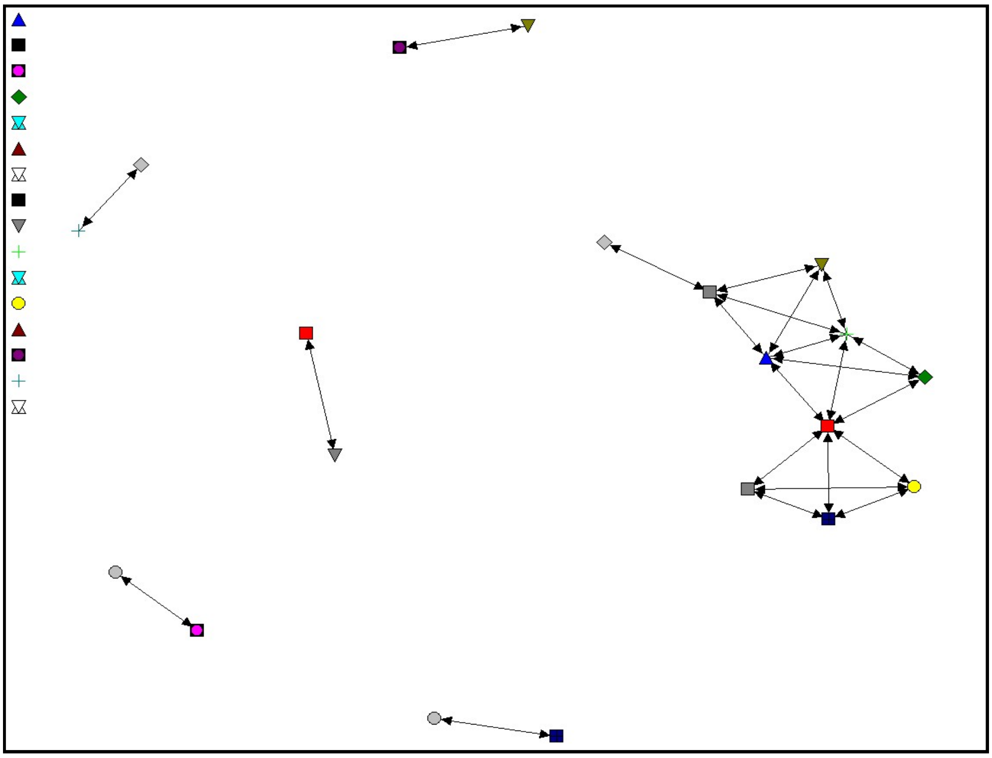
C. Family Network Size as a Ratio of Number of Households in the Village



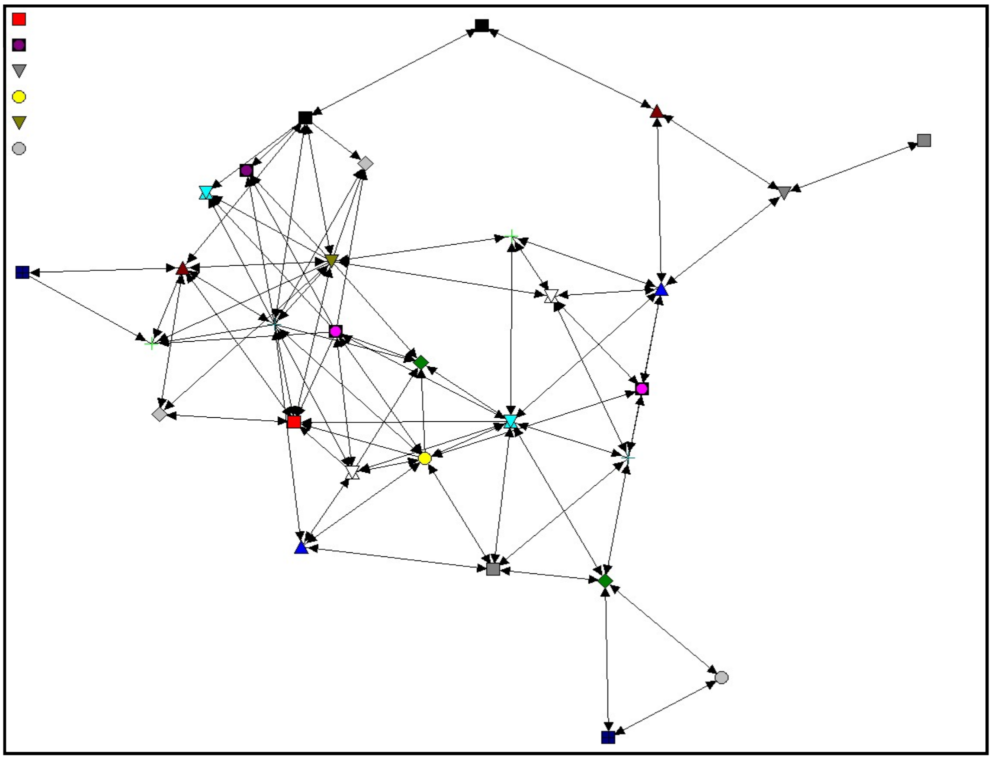
Notes: Each figure is constructed from those family networks with at least two households in them. Of the baseline sample of 22553 households that can be tracked over the first and third waves of Progresa, 17030 (75.5%) of them are within family networks with at least two households. These can be single or couple headed households. There are 2196 family networks in total.

Figure 4: 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.