Self-Help Groups and Mutual Assistance:

Evidence from Urban Kenya*

Marcel Fafchamps† Eliana La Ferrara‡

University of Oxford Bocconi University and IGIER

April 2011

Abstract

This paper examines the incomes of individuals who have joined self-help groups in poor neighborhoods of Nairobi. Self-help groups are often advocated as a way of facilitating income pooling. We find that incomes are indeed more correlated among individuals in the same group than among individuals who belong to different groups. Using an original methodology, we test whether this correlation is due to self-selection of similar individuals into the same groups. We find that this correlation is not driven by positive assortative matching. If anything, selection works in the opposite direction: incomes from group activities would be more correlated if individuals were matched at random. These findings are consistent with the idea that self-help groups play a mutual assistance role.

JEL codes: O12 Keywords: mutual insurance; social capital; associations; self-selection

*We thank Kjetil Bjorvatn, Erik Sorenson, an anonymous referee and participants to the 2009 EUDN Conference in Paris and to seminars at Oxford University, the Christian Michelsen Institute in Bergen, and Trinity College Dublin for helpful comments. Ignatius Mayero and Peter Warui provided excellent assistance in conducting the interviews. Fafchamps acknowledges financial support from the Economic and Social Research Council (UK). The work is part of the programme of the ESRC Global Poverty Research Group. La Ferrara acknowledges financial support from the European Research Council grant ERC-2007-StG-208661. The usual disclaimer applies.

†Department of Economics, University of Oxford, Manor Road, Oxford OX1 3UQ. Email: marcel.fafchamps@economics.ox.ac.uk. Fax: +44(0)1865-281447. Tel: +44(0)1865-281446.

‡Department of Economics, Bocconi University, via Sarfatti 25, 20136 Milano, Italy. Email: eliana.laferrara@unibocconi.it

1. Introduction

The risk sharing strategies of poor households have received much attention in the development economics literature. The bulk of the literature has focused on rural households and on income shocks induced by rainfall and illness.[1] Much less attention has been devoted to urban households. Yet, a growing proportion of the population in contemporary developing countries lives in urban settings, and the urban poor also face a considerable degree of income risk. Despite the importance of this topic, we know little about the insurance strategies that the urban poor use to cope with income shocks. This paper seeks to redress this imbalance by studying one particular organizational form that is generally believed to serve this function in urban Kenya, i.e. self-help groups.

The recent empirical literature has turned to explicit risk sharing arrangements such as the funeral societies, studied by Dercon et al. (2006) in rural Ethiopia and Tanzania, and health insurance groups, studied by De Weerdt and Dercon (2006). The formation of risk pooling groups is studied in an experimental setting by Barr, Dekker, and Fafchamps (2009). The authors find no evidence that individuals with similar risk preferences pool risk together, but they find assortative matching on age and gender. Arcand and Fafchamps (2008) examine the formation of multi-purpose associations in rural Senegal and Burkina Faso. They find evidence of assortative matching by wealth and connection to village elites. To our knowledge, no work has been done studying the role of urban self-help groups in helping the poor manage risk.

This study examines the incomes of individuals who have joined informal associations in various poor neighborhoods of Nairobi, the Kenyan capital. Similar groups have previously been studied by Anderson and Baland (2002) with a focus on gender issues, by La Ferrara (2002b)

[1] See, among others, Fafchamps and Lund (2003). Dasgupta (1995) and Fafchamps (2003) provide reviews of this literature.

with a focus on heterogeneity and group decision making, and by Anderson and Francois (2007) with a focus on organization and decision making rules. Our main interest is in understanding the extent to which associations serve to pool income – and hence serve an insurance purpose. Because we do not have data on individuals who did not join associations, we cannot investigate the factors that incite individuals to join associations. But we have recall information on income before and after joining the association.

We find that the incomes of individuals in the same group are more correlated than the incomes of individuals who do not belong to the same group. Using an original methodology, we test whether the correlation in incomes is due to self-selection into groups. We find that this is not the case, i.e. that correlation in total incomes and in incomes from group work is not the result of positive assortative matching. If anything, for group earnings selection works in the opposite direction: incomes would be more correlated if individuals were matched at random. Our paper thus shows that not only incomes are more similar within self-help groups, but that this is not driven by assortative matching into groups. These findings are consistent with the idea that associations play an insurance or redistributive role.

This paper makes several contributions. First, it complements a burgeoning literature on community-based development. Many donors now see local associations as an essential relay for development assistance. Yet little is known on how these associations work in practice. This paper is among the first to test explicitly whether they play a mutual assistance role. From a policy perspective, this is important because it sheds light on the nature of seemingly inefficient arrangements: in our sample, several production groups adopt remuneration schemes that are suboptimal from an incentive point of view, but that are consistent with risk sharing purposes. Secondly, this paper adds to a scant literature on mutual insurance in an urban setting. There is indeed a large literature on risk sharing in rural villages that followed Townsend's (1993) seminal contribution, but very little evidence regarding urban areas of Africa. It should be noticed, though, that aside from focusing on an urban as opposed to a rural setting, our approach differs from the traditional mutual insurance literature in that we do not focus on ex post transfers and consumption smoothing after income shocks, but focus on income smoothing instead. Finally, we make a novel methodological contribution by showing how to correct for group self-selection in the study of group peer effects. To our knowledge, the approach proposed here has never been used before.

The paper is organized as follows. In Section 2 we present our testing strategy. The data and descriptive statistics are presented in Section 3. Section 4 contains empirical results that do not control for self-selection into groups, while the results in Section 5 control for selection on observables and on unobservables. Finally, Section 5 concludes.

2. Conceptual framework and testing strategy

We wish to investigate whether self-help groups serve a mutual assistance role, which we interpret here as income sharing in a broad sense. One form of mutual assistance is when group members explicitly pool incomes. We have some information about formal income pooling rules, such as group rules specifying the sharing of sales proceeds, so we can investigate this issue directly. Group members may pool income informally even when they earn individual incomes within the group and formal pooling rules do not exist. One possibility, observed during fieldwork, is when group members pool customer orders. It is also conceivable that group members share knowledge and tools and learn skills from each other, so that incomes after joining are more alike than before.

Given our data, we cannot disentangle these different sources. But mutual assistance, whatever the source, should result in more similarity in incomes: individual incomes should be more correlated with those of co-group members than with those of non group members. This is what we set to test. For this test to be convincing, however, we need to control for assortative matching: co-group members may have a similar income not because they pool income but because those who join the same group are more alike in terms of income earning potential. The object of this section is to explain how we correct for self-selection.

Let y_{it} be the (log of) income of individual i at time t. For each individual i in the dataset, we observe income in two periods: before joining the self-help group ($t = 0$) and after joining ($t = 1$). Define $g_{ij} = 1$ if i and j belong to the same group, and 0 otherwise. We want to test whether the incomes of individuals joining the same group are more similar than those of individuals belonging to different groups. Let income be written:

$$y_{ikt} = \beta t + \gamma x_{ik} + \rho y_{it-1} + \eta_k + \varepsilon_{ikt} \tag{2.1}$$

where t is time, k indexes the group to which individual i belongs, x_i is a vector of individual characteristics – such as age, gender, and education – thought to affect i's income potential, income before joining the group y_{it-1} proxies for unobserved income earning potential, β, γ and ρ are unknown parameters, η_k is a group random effect and ε_{ikt} is a disturbance term.

We only have observations for two time periods: they time of the survey $t = s$, which is the same for all observations; and the time before joining the group $t = b_i$, which varies across individuals. Given that income is in log, equation (2.1) can be rewritten as:

$$y_{iks} = \beta t_i + \rho y_{ib_i} + \gamma x_{ik} + \eta_k + \varepsilon_{iks} \tag{2.2}$$

where $t_i = s - b_i$ is the time individual i has been in the group.[2]

[2]When $\rho = 1$ – which is not the case in our data – equation (2.2) is equivalent to controlling for individual

If members of the same group pool income or help each other smooth income in any way, observations on y_{iks} are not independent within groups: individuals have incomes that are more alike if they belong to the same group. This suggests investigating the presence of mutual assistance by testing whether:

$$E[\varepsilon_{iks}\varepsilon_{jhs}|k = h] = E[\varepsilon_{is}\varepsilon_{js}|g_{ij} = 1] > E[\varepsilon_{is}\varepsilon_{js}|g_{ij} = 0] = E[\varepsilon_{iks}\varepsilon_{jhs}|k \neq h] \qquad (2.3)$$

Equation (2.3) forms the basis of our testing strategy. The test is implemented in a manner reminiscent of a Breusch-Pagan test of heteroskedasticity, or of a test of serial autocorrelation. We begin by estimating equation (2.2) to obtain consistent estimates of the residuals $\widehat{\varepsilon}_{is}$. We then construct all $\widehat{\varepsilon}_{is}\widehat{\varepsilon}_{js}$ pairs, regress them on g_{ij}, and test whether the coefficient of g_{ij} is equal to zero. Since each $\widehat{\varepsilon}_{is}$ residual appears in all $(i, .)$ and all $(., i)$ pairs, observations are not independent. To correct for this, we use the standard error formula proposed for dyadic regressions by Fafchamps and Gubert (2005).[3]

It is possible that correlation in realized incomes – and thus in residuals $\widehat{\varepsilon}_{is}$ – may be due to factors other than group membership. For instance, suppose that, irrespective of group membership, female incomes are more correlated with each other than with those of men. Further suppose that there is assortative matching by gender in the formation of self-help groups, in the sense that women are more likely to belong to a group with other women. This generates a spurious correlation in realized incomes within groups. To allow for this possibility, we introduce a vector of observable characteristics d_{ij} to control for assortative matching. The estimated

fixed effects. In that case, γ should be interpreted as the effect of variable x_i on the growth of income relative to period 0.

[3]Based on Conley (1999), the formula proposed by Fafchamps and Gubert has the drawback that in finite samples it is not guaranteed to yield a positive variance, a point that is discussed in detail by Conley. In practice, however, this is usually not a problem.

model takes the form:

$$\widehat{\varepsilon}_{is}\widehat{\varepsilon}_{js} = \theta g_{ij} + \alpha d_{ij} + u_{ij} \tag{2.4}$$

where θ is our coefficient of interest and d_{ij} is a vector of dyadic controls. Since $\widehat{\varepsilon}_{is}\widehat{\varepsilon}_{js} = \widehat{\varepsilon}_{js}\widehat{\varepsilon}_{is}$ always, control variables d_{ij} must be defined in such a way that the order of subscripts is irrelevant. How this is achieved is discussed in the empirical section.

It is also possible that realized incomes are correlated because of self-selection on unobservables. This would arise, for instance, if individuals with a high income potential team up with other similar individuals when forming groups – and similarly for low income individuals. This introduces a correlation between g_{ij}, our regressor of interest, and the residual u_{ij}. To perfectly address this issue, we would ideally need random assignment to groups. Given the need to have individuals voluntarily join (and stay in) the self-help groups that we study, it was not possible to obtain such data.[4] We therefore try to do the next best thing, which is to control for self-selection directly. To do this we need an instrument, that is, a variable that predicts group co-membership but not correlation in realized incomes. Assuming that such a variable z_{ij} exists, selection on unobservables can be dealt with using one of two methods.

The first correction method is instrumental variables, using z_{ij} as instrument.[5] The second method uses a selection correction procedure that draws on Wooldridge (2002).[6] The advantage of this second method is that it provides information on the nature of the self-selection bias. We

[4]It may be possible to create a laboratory experiment that mimics the structure of self-help groups and in which the researcher randomly assigns individuals to different groups. While this approach would eliminate the potential for self-selection, it would suffer from a number of other shortcomings – most notably that participants may not be representative of the target population and that the laboratory situation is a poor substitute for actual self-help groups, thereby reducing external validity.

[5]Given that the dyadic standard error correction formula is not available for 2SLS, we estimate regression model (2.2) using OLS but include the residuals from the instrumenting regression as additional regressor.

[6]Specifically, this procedure is described in Wooldridge (2002), Chapter 18, p. 631.

first estimate a selection regression of the form:

$$g_{ij} = \phi(\eta w_{ij} + \varphi z_{ij}) \tag{2.5}$$

where w_{ij} is a vector of regressors that affect the likelihood of belonging to the same group, and $\phi(.)$ is the standard probit function. Let $\widehat{\phi}$ denote the predicted density generated by (2.5) and let $\widehat{\Phi}$ be the corresponding predicted cumulative density. We use these to construct two Mills ratios of the form $g_{ij}\frac{\widehat{\phi}}{\widehat{\Phi}}$ and $(1-g_{ij})\frac{\widehat{\phi}}{1-\widehat{\Phi}}$ which are then added to equation (2.4) as control variables for self-selection:

$$\widehat{\varepsilon}_{is}\widehat{\varepsilon}_{js} = \theta g_{ij} + \alpha d_{ij} + \delta_1 g_{ij}\frac{\widehat{\phi}}{\widehat{\Phi}} + \delta_2(1-g_{ij})\frac{\widehat{\phi}}{1-\widehat{\Phi}} + u_{ij} \tag{2.6}$$

Since $g_{ij} = \{0,1\}$, for each observation only one Mills ratio at a time is non-zero.

If there is positive assortative matching on income potential, we expect $\delta_1 > 0$ and $\delta_2 < 0$. Finding that $\delta_1 < 0$ and $\delta_2 > 0$ indicates negative assortative matching, that is, that individuals in the same group are more dissimilar in terms of income potential than if they were matched at random. This may arise, for instance, if self-help groups are organized around energetic individuals who deliberately co-opt into their group weaker members of society as a kind of income guarantee scheme. In this case, we should also find that the estimated parameter θ increases once we control for self-selection.

Before turning to the estimation proper, there is one additional data feature we wish to discuss. For period $t = 1$, income information is disaggregated by source, i.e., we have income earned from the group and income earned outside the group. This information can be used to verify the interpretation of our results as follows.

Suppose that self-help groups (implicitly or explicitly) pool incomes but there is no self-selection into groups. Mutual assistance generates a correlation in realized income residuals $\widehat{\varepsilon}_{is}\widehat{\varepsilon}_{js}$ for income earned from the group. But it should not generate a correlation in income residuals for income earned outside the group. If anything, high potential individuals who earn a low group income because of mutual assistance and redistribution within the group may divert their efforts towards outside income generating activities. As a result we would expect the correlation in non-group income to be, if anything, lower for same group pairs compared to non-same group pairs. Formally, we expect $\theta \leq 0$ for non-group income if groups pool income and exchange skills but there is no self-selection into groups.

In contrast, if there is positive assortative matching into groups on the basis of income potential, we expect correlation in non-group income to be higher for same group pairs compared to non-same group pairs. This means that finding $\theta > 0$ for non-group income provides additional evidence of assortative matching by income in the formation of self-help groups. To investigate these issues, we estimate models (2.4) and (2.6) separately for group and non-group income.

3. The data

3.1. Setting

The data used in our empirical analysis were collected by one of the authors in 1999 in five among the most populated informal settlements of Nairobi: Dandora, Gikomba, Kayole, Korogocho, and Mathare Valley. The majority of Nairobi's population lives in informal settlements, or 'slums', despite the fact that these settlements only account for 6 percent of the city's residential area. As a result, the average population density in the early 1990s was 750 people per hectare (and 250 dwellings), well above the 50 to 180 people per hectare (and 10 to 30 dwellings) in the upper and middle income areas of the city (Alder et al., 1993). After Kibera, which is the biggest slum of Nairobi and is not covered by our analysis, the five locations in our sample account for a large fraction of Nairobi's slum dwellers.

Living standards in these communities are extremely low. The lack of property rights, coupled with a very high crime rate, create a climate of extreme insecurity, and the population makes little or no investment in infrastructure and local public goods. Dwellings are made of temporary materials such as mud, wattle and timber off-cuts, and waste disposal is done on the street and in the streams that run through the slums. Most sites have virtually no sewerage systems and lack the basic hygienic conditions.

The inhabitants of the surveyed communities typically work in the informal sector. They may work as hawkers, be occasionally employed for the day, operate small businesses without licenses, or even be involved in illegal activities. Nairobi is well known for its small businesses, locally referred to as "jua kali", which include small manufacturing activities, repair and services.[7] Operating a jua kali microenterprise is not easy. In order to get training, individuals often have to pay their employer for apprenticeship. When they are ready to start their own business, apprentices need to find startup capital. Most of them find this very hard because they have limited if any access to credit and equity markets and lack collateralizable wealth. As a result, it is very common for slum dwellers to pool their capital and form so called "self-help" groups. These self-help groups are the object of our analysis.

Self-help groups operate in the informal sector with income generating activities similar to those of individual jua kali workers. But their organization roughly resembles that of a production cooperative. The degree of formalization of these groups, as well as their stability and the scope of their activities, vary a lot. Some of them serve the formal sector in the city (e.g., one group in our sample produces footballs that are then sold in local stores), others export

[7] In Swahili, "jua kali" means "under the sun". The term was created to describe people who work in the street without permanent structures. Nowadays, however, it is often used to denote the whole informal small business sector in Nairobi.

through the Fair Trade network, others sell goods or services in the slum where they are located.

3.2. Data collection

The sampling strategy employed to select the groups in our dataset was the following. A list of self-help groups active in the areas of Dandora, Gikomba, Kayole, Korogocho, and Mathare Valley was drafted with the help of local community workers and NGO representatives. Only groups that had some kind of income generating activity were considered. These included activities such as crafts-making (e.g., wood carving, basket weaving), tailoring, garbage recycling, informal lending, and the provision of local education and health services. The selection criterion was to pick groups that were representative of the most common activities in these areas, and to ensure sufficient variation across groups in terms of group location, size, type of activity, and age and gender of the members. Neither the residential location nor the ethnicity of group members was known at this stage, and hence were not part of the sampling criterion.

Preliminary meetings were set up with the chairperson and secretary of each group in order to compile a list of all active members of the group. Individual interviews were scheduled with all members to get information on their individual and household characteristics, their income, and their opinions on the management and performance of their group. Interviews were done on the group premises but in a separate room, so that privacy was ensured. Group members were interviewed one at a time. All interviews were face to face and occurred in the presence of one of the authors and one enumerator. All group members were administered demographics, income, and group functioning modules of the questionnaire. An additional module was administered separately to the chairperson, the treasurer, and the secretary of each group. This module contained questions about the group's history, organizational structure, payment schemes, costs, and revenues.

3.3. Descriptive statistics

The sample we shall use consists of 16 groups for a total of 274 individual members.[8] Given the focus on income generation, this sample is not representative of all self-help groups in the informal settlements of Nairobi. But it provides a reasonable representation of small and medium-sized groups with income generating activities that distribute income to their members and have a low degree of formalization.

[Insert Tables 1-2]

Table 1 reports summary statistics on variables of interest at the group level. The groups in our sample differ substantially in terms of size and of the number of years they have been in place. The smallest group has 6 members and the largest has 30, with an average of 14.7 members per group. The average group has been in place for 5.1 years, with a minimum of 9 months and a maximum of 16 years. Recruitment is not always based on skill or productivity: in 5 out of 16 groups, the primary criterion is that the person recruited is in need or is a friend of an existing members.

The main activity of the group varies a lot from group to group. The most common activity (7 groups) is to produce some kind of public good (e.g., garbage recycling, health services). Regarding the organization of production, in 7 groups members specialize in production tasks; in the remaining 9 they don't. Remuneration schemes are also different across groups: in 9 groups internal rules stipulate that all members receive the same pay; in the remaining ones, group rules stipulate that members are either paid in proportion to the amount of goods produced or in proportion to the value of capital contributed to the group. How strictly these rules are enforced is unclear, however. For instance, in the 9 'same pay' groups, individual incomes reported by

[8] From our original list, two groups had to be dropped for which we did not manage to interview 90 percent of the active members. Given that the focus of this paper is on the correlation among member incomes, we drop two groups that did not distribute income to their members and for which income data is consequently missing.

group members vary quite a bit. This may be because formal enforcement tends to be rather poor in the setting under study: only 4 out of 16 groups impose a sanction (typically, a monetary fee or expulsion) if members do not pay group fees; in the remaining cases, individuals are invited to contribute but no specific action is taken.

The majority of groups (11 out of 16) allow members to borrow from group funds. For 9 of the groups in our sample, membership has decreased since the group's foundation. When prompted for an explanation of this fact, several group leaders indicated that often members tended to leave the groups once they achieved sufficient independence to work on their own. Group profits are reinvested in group activities (e.g., to increase the stock of physical capital or buy intermediate inputs) in 7 out of 16 groups; in the remaining ones they are distributed to members.

Turning to individual characteristics (Table 2), the most interesting dimension of individual heterogeneity from our point of view is income.[9] We have information on current income as well as income before joining the group. Current income is further disaggregated into income from group activities and income from work outside the group. Group income is measured after any income pooling and redistribution within the group has taken place. The average monthly income of individuals in our sample was 3076.4 Kenyan Shillings, which is equivalent to 41.5 US dollars at August 1999 prices. The standard deviation of this variable is 2789 KSh. Average income from group activities was 2383.6 KSh (standard deviation 2226.8 KSh), while income from non-group work averaged 661 KSh (standard deviation 2059.2 KSh). These findings suggest that we are looking at a population that is poor and has a very erratic income, facing considerable risk. In fact, when asked directly, 86 percent of the respondents replied that their incomes had been uncertain over the past two years. We also have information on the income

[9] La Ferrara (2002a) studies the relationship between income heterogeneity and group participation in rural Tanzania.

that current group members were earning before joining the group: the mean of this variable is 1346 KSh, i.e. considerably lower than current income. While we do not have the necessary data to investigate this in detail, a likely interpretation is that individuals were less experienced at the time. Furthermore some, many of them women, were not working before joining a group.

The average age of individuals in our sample is 30.3 years, with a standard deviation of 11.2 years. 49 percent of our sample is constituted by women. On average, the members of our groups have 8.5 years of education; 9 percent of the sample has no formal schooling (but may have received adult education), 50 percent has Class 1 to 8, 38 percent has Form 1 to Form 4, and the remaining 3 percent has higher education than Form 4. The most numerous ethnic groups in the sample are the Kikuyus (37 percent) and the Kambas (37 percent); Luos constitute 14 percent and Luhyas 6 percent of the sample. Finally, 57 percent of group members are Catholic, 40 percent belong to another Christian denomination, and 2 percent are Muslim.

4. Correlation in incomes: naive results

The main focus of our analysis is the correlation in the incomes of individuals who belong to the same group. We attempt to disentangle to what extent such correlation reflects peer selection, i.e., ex-ante similar individuals sorting into groups, or group processes that pool individual incomes. Before tackling the problem from an econometric point of view, we present two simple graphs to motivate the analysis.

[Insert Figures 1-2]

Figure 1 plots mean incomes before and after joining the group for current group members. Each observation corresponds to one of the groups in our sample (hence the 16 data points in the graphs). On the horizontal axis we show the average total income that members of a given group were earning before joining it. On the vertical axis we show the average total income of the same group members at the time of the interview, including group and non-group income. The solid line is the 45° line. Figure 1 shows that incomes before and after are positively correlated, and that for all groups except two, mean income increased after joining the group.

Figure 2 looks at income dispersion. It shows the ratio of the standard deviation of income to mean income, before (horizontal axis) and after (vertical axis) joining the group. Again, the solid line is the 45° line. The Figure suggests a remarkable decrease of the dispersion in incomes after joining the group compared to initial levels. In fact, for 12 out of 16 groups the ratio of standard deviation to mean income is lower at the time of the survey than it was before the respondents joined their respective groups. This constitutes prima facie evidence of income correlation within groups.

Econometric analysis provides evidence pointing in the same direction. In Panel A of Table 3 we regress individual incomes on a constant plus a set of group dummies. We use three different dependent variables: the log of total current income (column 1); the log of income earned in the group (column 2); and the log of income earned outside the group (column 3).[10] For each regression we report the F statistic and the p-value of the test of joint significance of the group fixed effects. For all three measures of income we reject the null of equal intercepts across groups at the 1 percent level. This is suggestive of a common source of income variation for individuals who belong to the same group.

[Insert Table 3]

[10] Original income values were rescaled by dividing them by 100. To avoid losing observations with zero income, we added 1 (i.e., the equivalent of 100 Shillings) before taking natural logs.

In Panel B of Table 3 we explore this more directly using a dyadic regression framework. We begin by estimating a simple model of the form:

$$\frac{(y_{is} - \overline{y_s})(y_{js} - \overline{y_s})}{\sigma^2_{y_s}} = \theta g_{ij} + u_{ij} \tag{4.1}$$

where $\overline{y_s}$ is average income at the time of the survey. Regressors include a constant and a dummy g_{ij} equal to 1 if i and j belong to the same self-help group. $E[(y_{is} - \overline{y_s})(y_{js} - \overline{y_s})/\sigma^2_s]$ is the correlation in the (log of) income y_{is} of any two pairs of individuals (i, j). Coefficient θ captures how much more correlated incomes are when individuals i and j are in the same group. This regression does not account for the possibility of endogenous sorting of individuals into groups, an issue we revisit below. Standard errors are corrected for dyadic correlation across errors, as suggested by Fafchamps and Gubert (2007).

The coefficient of g_{ij} in column 1 is around .6 and is significant at the 1 percent level. This means that the total incomes of members of the same group are more correlated with each other than those of members of different groups. The effect is larger in magnitude when we consider income earned within the group (column 2), with an estimated coefficient of .92, significant at the 1 percent level. Incomes earned outside the group are also significantly more correlated for individuals who belong to the same group (column 3), although the estimated coefficient is smaller, at around .2.

It is unclear is whether this correlation is due to mutual assistance or to self-selection into groups, i.e., the sorting of individuals with similar characteristics (hence similar earning profiles) into the same group. In the next section we explicitly acknowledge the possibility of sorting and use different ways of controlling for selection.

5. Controlling for selection into groups

5.1. Selection on observables

Part of the correlation in incomes may be due to the fact that individuals with similar observable characteristics (e.g., gender, age, education, ethnicity, etc.) join the same group. For example, gender may affect income potential – women in Kenya are typically expected to take responsibility for parenting and housework, giving them less time to pursue income generating activities. Ethnicity and religion may affect access to different occupations, thereby affecting income potential. These characteristics are all observable at the time individuals join the group. Individuals may also differ in ways that are not directly observable but are reflected in the income they earned before joining the group, and are thus indirectly observable.

To control for selection on observables, we start by netting out the effect of these observables on log income. To this effect, we estimate equation (2.2), that is, we regress individual income after group formation on a set of observable covariates thought to affect income potential, and on income before joining the group to proxy for other individual characteristics. We then obtain estimates of residuals $\widehat{\varepsilon}_{is}$ and reestimate the dyadic regression using as dependent variable the correlation in these residuals $\widehat{\varepsilon}_{is}\widehat{\varepsilon}_{js}/\sigma^2_{\varepsilon_s}$ instead of the correlation in actual (log) incomes $(y_{is} - \overline{y_s})(y_{js} - \overline{y_s})/\sigma^2_{y_s}$.

[Insert Table 4]

Table 4 shows the first step of this process, i.e. the income regressions we use to generate the residuals. As before, model (2.2) is estimated for three dependent variables: the log of total current income; the log of income earned in the group; and the log of income earned outside the group. Among individual covariates x_i we include gender, age and its square, years of education, ethnicity and a set of dummies for the area (slum) of Nairobi where the individual lives. The log of income earned by the individual before joining the group is a proxy for other factors affecting the income generating potential of individual i. The number of years the individual has been a member of the group serves as a proxy for experience. For internal consistency, equation (2.2) is estimated using group random effects to explicitly acknowledge the possible correlation in residuals within groups.

Several interesting patterns emerge from Table 4. The first is that income earned before joining the group is a predictor of total income, and especially of non-group income. But it is not a good predictor of income earned within the group. This suggests that earnings from group work tend to be levelled and do not reflect ex ante differences in income generating potential. Non-group incomes, however, still reflect these differences. Since we regard income before joining the group as a proxy for income-generating ability, our results suggest the existence of a skill premium in non-group work, but not in group work.

A second result, in line with the first one, is that group income does not increase with years of education (columns 1 and 2) while non-group income does (column 3). For non-group income, one additional year of education is associated with an income increase of 5.6 percent. Given the low average level of education in the sample, these low returns may reflect the low skilled types of occupations in which our respondents engage.[11] Taken together, these results imply that the incomes distributed by self-help groups do not reflect differences in skill or education that are otherwise valued in the market.

Turning to other regressors, as is often found in African data women have significantly lower incomes than men. Age does not seem to be correlated with individual earnings. But experience in the group (proxied by the number of years an individual has been member of the group) displays puzzling coefficients: a negative and significant coefficient in the regression

[11] Fafchamps, Söderbom, and Benhassine (2009) show that in much of sub-Saharan Africa returns to schooling are convex, with very low returns to primary education.

whose dependent variable is income from the group, and a positive and significant coefficient in the regression of income from outside activities. One possible explanation is that the self-help groups we are studying are not a satisfactory form of long-term employment. Individuals who stay longer in these groups tend to complement relatively low incomes from group activities with higher incomes from other jobs. As for ethnicity, none of the coefficients is statistically significant, with the exception of the Luo dummy in the total income regression, which has a (marginally significant) negative coefficient.

[Insert Table 5]

We use the estimates in Table 4 to predict the residuals for all three types of incomes, and estimate the dyadic model (2.4) using $\widehat{\varepsilon}_{is}\widehat{\varepsilon}_{js}/\sigma^2_{\varepsilon_s}$ as dependent variable. The results are shown in columns 1, 4 and 7 of Table 5. The coefficient of the "Same group" dummy g_{ij} is positive and significant at the 1 percent level for total income and income from the group (estimated coefficients of .3 and 1.02, respectively), while it is close to zero (and not significant) for income earned outside the group. If we compare the magnitude of these coefficients with those in Panel B of Table 3, we see that the impact of g_{ij} is reduced for non-group earnings and for total income, while it is if anything increased in the case of group earnings. This suggests that correlation in observable characteristics is almost entirely responsible for the observed correlation in (raw) incomes earned outside the group, while it does not explain the correlation in incomes from group work. We next explore to what extent this correlation may be driven by correlation in unobserved individual characteristics.

5.2. Selection on unobservables

To address self-selection into groups based on unobservables, we start by estimating equation (2.5) to identify the factors that affect the likelihood that any two individuals belong to the same group. The dependent variable is g_{ij}, i.e, it is equal to 1 if i and j are members of the same group, and 0 otherwise. Since g_{ij} is symmetric by construction (i.e., $g_{ij} = g_{ji}$), regressors w_{ij} must be constructed in such a way that $w_{ij} = w_{ji}$. To achieve this, we follow Fafchamps and Gubert (2007) and construct regressors of the form $|w_i - w_j|$ and $(w_i + w_j)$ where w_i and w_j are characteristics of i and j. A negative coefficient on a $|w_i - w_j|$ regressor indicates positive assortative matching: the more dissimilar i and j, the less likely they are to be in the same group. A positive coefficient on a $(w_i + w_j)$ regressor indicates that this characteristic is associated with membership in larger group. The individual characteristics included are those appearing in Table 4, namely, age, gender, education, (log of) pre-group income, ethnicity (dummies for Kamba, Luhya, and Luo, with Kikuyu being the omitted category), and religion (dummy for Catholic).

As we discussed in Section 2, in order to separately identify the effects of group membership on income correlation from that of self-selection into groups, we need an instrument z_{ij}, that is, a variable that affects group membership but not income correlation once we control for self-selection. To serve this purpose, we select distance between place of residence: presumably, individuals who live close to each other find it easier to join the same self-help group. Since we only have limited information on the place of residence of each individual in the sample, we construct a distance variable that takes value 0 if two individuals live in the same block within a slum, 1 if they live in the same slum but not in the same block, and 2 if they live in different slums. This is the best we can do given the data.

To be a valid instrument, this variable must satisfy inclusion and exclusion restrictions. As we will see, the variable satisfies the inclusion restriction: it is a strong predictor of group membership. Regarding the exclusion restriction, one important concern is the possibility that individuals who reside near each other may hear of similar work opportunities, a point emphasized for instance by Topa (2001). To capture this possibility, residence fixed effects have been included in the income regression (2.2) so that residuals $\widehat{\varepsilon}_{is}$ are already purged of the average correlation in group incomes due to residential proximity. Given this, it is not unreasonable to assume that the correlation in $\widehat{\varepsilon}_{is}$ residuals between two individuals does not depend on residential proximity once we control for group self-selection. This is not as strong an assumption as it seems because the areas we surveyed are extremely comparable in terms of living standards and income generation opportunities. The main effect of living in different parts of these slums is that of increasing the transportation costs incurred to participate in group income generating activities (which often occur daily), not that of giving access to different labor markets.

[Insert Table 6]

The estimates of the selection equation are reported in Table 6. Standard errors are corrected for dyadic dependence between residuals. Coefficient estimates suggest strong assortative matching by age, gender, ethnicity, and religion. Years of education is not significant and pre-group income is only marginally significant,[12] suggesting that the groups in our sample are fairly heterogeneous with respect to ex ante earning ability. The coefficients on the 'sum' variables indicate that women, Catholics and relatively more educated individuals tend to belong to smaller groups, while members of the Kamba ethnic group belong to larger groups. Distance between the places of residence significantly decreases the likelihood of belonging to the same group. This variable is significant at the 1 percent level. This satisfies the inclusion restriction.

Equipped with these results, we revisit the income correlation regressions and correct for selection bias. As explained in Section 2, we employ two alternative methodologies: Mills ratios and instrumental variables. In the first approach, the selection equation is estimated using Probit and then used to construct Mills ratios that are used as control variables. In the IV approach, the instrumenting regression is estimated using OLS and the residual from the this

[12] Not significant in a logit regression with similarly corrected standard errors.

regression is used as control. This amounts to using instrumental variables. The covariates are the same as those reported in Table 6.

Results are presented in Table 5. In columns 2, 5 and 8 we present estimated coefficients using Mills ratios as controls. In column 3, 6 and 9 we include the residual from the instrumenting regression as controls. Columns 2 and 3 report the results for the correlation in total incomes. Using either method we find that the coefficient of g_{ij} is positive and significant. It is also not significantly different from the estimate reported in column 1. This suggests that our methodology does not reveal the presence of self-selection into groups above and beyond that captured by selection on observable characteristics. This is also confirmed by the fact that neither the coefficients on the two Mills ratios, nor that of the residual, are significantly different from zero.

When we focus on earnings from the group, on the other hand, the results are different. In columns 5 and 6 we find that membership in the same group increases the correlation in group incomes by a factor of 1.3 to 1.9, respectively. These point estimates are higher than the estimate in column 4. This suggests that if individuals had been randomly assigned into groups, the group incomes that they currently earn would have displayed higher within-group correlation than they actually do. Put differently, if people had been randomly assigned to groups, the group income they earn would have been even more alike. This is also the interpretation we get from the coefficients of the two Mills ratio control variables, and it suggests that selection into groups follows dissimilar matching: weaker people join groups with stronger people. Because group incomes are at least partly pooled, this enables weaker people to do better than they otherwise would.

When we take as dependent variable the correlation in incomes earned outside the group (columns 7-9), none of the estimated coefficient on "Same group" is significantly different from zero. To the extent that income outside the group can be considered a proxy for income growth potential, the coefficients of the two Mills ratios suggest positive assortative matching in group formation: individuals with similar unobserved income potential join together. However, this result is not robust: when we use the correction method based on instrumental variables (column 9), no evidence of endogenous group formation appears.

Overall, the results in table 5 suggest that group formation in the setting under study involves an element of mutual assistance: incomes of group members are more correlated than that of non-group members, suggesting some form of explicit or implicit income sharing. When looking at earnings from group activities we find evidence that self-help groups bring together dissimilar individuals, possibly for mutual insurance purposes.

5.3. Robustness

Formal income pooling

Among the groups in our sample, there is a fair amount of heterogeneity in the remuneration schemes adopted. In 9 out of 16 groups, internal rules stipulate that members are paid the same amount regardless of the amount they produce or the hours of work they put in. In the remaining 7 groups, incomes received within the group are supposed to vary depending on the amount of work done or of the capital contributed to the groups itself. In what follows we test whether our results are affected when we take this group heterogeneity into account. Formally, we augment the specification (2.4) with one regressor that captures whether group rules explicitly require that each individual is paid the same amount. The dummy variable "Same group and same pay" takes value 1 when two individuals belong to the same group and this group adopts an equal pay rule, and takes value 0 otherwise.

One important caveat is that the choice of remuneration schemes may not be exogenous, hence our estimated coefficient on "Same group and same pay" may suffer from endogeneity bias. While we have a suitable instrument for "Same group", we do not have one for the choice of remuneration scheme. For this reason, the coefficient on "Same group and same pay" should not be given a causal interpretation. The purpose of including this variable among the controls is purely to test whether the higher correlation we find in group incomes is mechanically driven by the formal remuneration rules chosen by the groups.

[Insert Table 7]

Estimation results are displayed in Table 7. As before, regressions in columns 1, 4 and 7 only control for selection based on observable characteristics, while the remaining columns control for selection on unobservables using Mills ratios and instrumental variables. A first notable result is that in all specifications the coefficients of the variable "Same group" are very similar to those of table 5, indicating that our results are robust.

As for the variable "Same group and same pay", it has positive and significant coefficients in the regressions involving earnings outside the group (columns 7-9), insignificant in the regressions for group income (columns 4-6) and negative and marginally significant with total income (columns 1-3).[13] A possible interpretation is that the same pay rule may act as a disciplining device on individuals with otherwise similar income potential, inducing them to dedicate similar efforts to activities outside the group, which would generate higher correlation in non-group incomes. However, as we discussed above, this coefficient should not be interpreted in a causal way.

From the point of view of our analysis, the main message from table 7 is that the estimated effect of belonging to the same group on correlation in incomes is robust.

[13] The different sign of "Same group and same pay" in the "Total income" regressions is likely explained by the negative correlation between group and non-group income.

Group size

We also investigated whether our results are affected by the possibility that there is less pooling in large groups, as some models of localized peer effects may suggest. We found that they do not: if we regress the correlation in residuals on group size dummies, we find no pattern. If we regress the correlation in residuals on an interaction term between "Same group" and group size, the coefficient of the interaction term is either not significant or appears with the wrong sign. Again, this exercise should be interpreted with caution, as the group size variable is potentially endogenous, e.g., groups that pool income may be more successful and lead more people to join or less people leave. Still the pattern of estimated coefficients suggests that our results do not depend on group size.

Specification of the income regression

Finally, we experimented with a different specification of the income equation (2.2), omitting the group random effects and estimating a simple OLS model.[14] This affects our results significantly: more of the variation in incomes is ascribed to individual characteristics, and less to within-group correlation. In particular, when we estimate our dyadic regressions controlling for selection on observables, only the correlation in group earnings remains significantly higher for members of the same group. When we control for selection on unobservables, the coefficient on "Same group" is either insignificant or negative; it becomes positive for those groups that adopt equal remuneration schemes.

Two considerations make us prefer estimating our income regression with a random effects model rather than OLS. The first is theoretical: when we estimate the correlation model, we

[14] We also examined whether our findings in Tables 5 and 7 are robust to omitting income before joining the group from equation (2.2). Results are broadly similar, but the coefficient of the "Same group" dummy is smaller and significance occasionally drops below the 10% level, depending on the model. This is not surprising: omitting y_{ib_i} from (2.2) leads to less precisely estimated ε_{iks}, and thus more noise in the dyadic dependent variable $\widehat{\varepsilon}_{is}\widehat{\varepsilon}_{js}$.

test the assumption that residuals are correlated within groups. Given this, it is natural to also estimate the income regression allowing for residuals to be correlated within groups. In fact OLS may over-ascribe to individual characteristics income correlation that is in fact due to group correlation of incomes combined with assortative matching. Income random effect estimates allow for group correlation after controlling for individual characteristics, and is thus more appropriate for our analysis. The second consideration builds on a specification test. When we compare the random effect and OLS models using a likelihood ratio test, we obtain a p-value of 0.002, which makes us select random effects over OLS.[15]

Other issues

Because the choice of group activity is endogenous to each group, we have not controlled for it. It is possible that group activity affects the mean and variance of group incomes and is one possible channel by which peer effects manifest themselves. This begs the question of how much of the correlation in incomes within group arises because of the activities different groups choose to undertake. We would have liked to investigate this question by including activity fixed effects in our analysis. Unfortunately this is not feasible because the small number of groups in our dataset undertake a wide range of different activities. Although this shortcoming of the data limits the interpretation we can give to the results, it does not invalidate the methodological approach per se.

6. Conclusions

We have examined whether self-help groups in the informal settlements of Nairobi, Kenya, serve a mutual assistance role and found evidence that they do. Lacking suitable exogenous variation,

[15] Specifically, we compare maximum likelihood implementations of the random effect and OLS models. The resulting likelihood ratio statistic is 8.151. The p-value is corrected for the fact that it is a boundary-value likelihood-ratio test.

we have not sought to disentangle deliberate income sharing, through the explicit or implicit pooling of risk, from peer effects operating through the choice of group activity and the diffusion of knowledge and skills. Our results should thus be construed as evidence that either or both are present.

Our empirical focus has been on distinguishing mutual assistance from self-selection into groups. Indeed, the incomes of group members may be correlated because individuals with a higher income potential group together. We deal with this possibility in two ways.

First, we decompose income into a component that can be predicted by individual characteristics – including pre-group income – and an individual-specific unpredicted component. We then examine whether the unpredicted component of income is more correlated among group members than non-group members. This approach controls for selection on observables.

Second, we seek to correct for selection on unobservables by using distance between the place of residence as instrument. Residential proximity can reasonably be expected to reduce the cost of joining the same self-help group. But since the different residential areas included in the study are similar in terms of structure and income opportunities, residential proximity should not affect income once we include dummies for each residential area in the income regression. To control for self-selection on unobservables, we use instrumental variables as well as Mills ratios.

Both approaches yield similar results, indicating that total incomes and group earnings are significantly more correlated among members of the same group (not so the incomes earned outside the group). When focusing on group earnings, our two correction methods provide evidence that, if anything, self-help groups are formed between individuals with dissimilar income potential: if groups were formed randomly, the incomes of group members would be even more correlated than they currently are. This provides further evidence that the studied groups serve a mutual assistance function, seeking to help disadvantaged members of poor urban neighborhoods by incorporating them in groups with less disadvantaged individuals.

Our findings are to be compared with the existing literature on whether groups and informal links are formed with the purpose of maximizing economic gains. Using an experimental approach, Barr, Dekker and Fafchamps (2009) find no evidence of assortative matching by risk preferences in the formation of risk pooling groups, in spite of the fact that, given the structure of the game, theory predicts such groupings. Using data from the Philippines, Fafchamps and Gubert (2007) find no evidence that risk sharing networks are formed in a way that would maximize the gains from pooling income risk. De Weerdt and Fafchamps (2008) similarly find little or no evidence that risk sharing links in a Tanzanian village maximize the gains from pooling health risk. In all these cases, the formation of groups and links is strongly influenced by social and geographical proximity, but little by economic motives. Our findings are complementary in the sense that they suggest that mutual assistance groups are formed with a redistributive motive in mind. These issues deserve further research.

References

[1] Abraham, A., J-M. Baland, and J-P. Platteau (1998), "Groupes Informels de Solidarité dans un Bidonville du Tiers-Monde: Le Cas de Kibera, Nairobi (Kenya)", Non-Marchand, 2, 29-52.

[2] Alder, G. et al. (1993), Nairobi's Informal Settlements: An Inventory, Nairobi: Matrix Development Consultants.

[3] Alesina ,A. and E. La Ferrara (2000), "Participation in Heterogeneous Communities", Quarterly Journal of Economics, 115, 847-904.

[4] Anderson, S. and J.-M. Baland (2002), “The Economics of Roscas and Intra-Household Resource Allocation”, Quarterly Journal of Economics, 117(3), 963-95.

[5] Anderson, S. and P. Francois (2007), “Formalizing Informal Institutions: Theory and Evidence from a Kenyan Slum”, in E. Helpman (ed.), Institutions and Economic Growth, Harvard University Press.

[6] Arcand, J-L. and M. Fafchamps (2008), “Matching in Community-Based Organizations”, mimeo, Oxford University.

[7] Barr, A., M. Dekker, and M. Fafchamps (2009), “Risk Sharing Relations and Enforcement Mechanisms”, Department of Economics, Oxford University, Oxford (mimeograph)

[8] Conley, T. (1999), “GMM Estimation with Cross Sectional Dependence”, Journal of Econometrics, 92(1), 1-45.

[9] Dasgupta, P. (1995), An Inquiry Into Well-being and Destitution, Oxford University Press.

[10] Dercon, S., J. De Weerdt, T. Bold and A. Pankurst (2006), “Group-Based Funeral Insurance in Ethiopia and Tanzania”, World Development, 34(4), 685-703.

[11] De Weerdt, J. and S. Dercon (2006), “Risk-sharing Networks and Insurance Against Illness”, Journal of Development Economics, 81(2), 337-56.

[12] Fafchamps, M. (2003), Risk, Poverty and Rural Development, Elgar

[13] Fafchamps, M. and S. Lund (2003), “Risk-Sharing Networks in Rural Philippines”, Journal of Development Economics, 71(2), 261-87.

[14] Fafchamps, M. and Gubert (2007), “The Formation of Risk Sharing Networks”, Journal of Development Economics, 83(2), 326-50.

[15] Fafchamps, M., M. Söderbom, and N. Benhassine (2009), "Job Sorting in African Labor Markets", Journal of African Economies (forthcoming).

[16] Grootaert, C. (1999), "Social Capital, Household Welfare, and Poverty in Indonesia", Local Level Institutions Working Paper No.6, The World Bank.

[17] Karlan, D. (2007), "Social Connections and Group Banking", Economic Journal, 117, F52-F84.

[18] La Ferrara, E. (2002a), "Inequality and Group Participation: Theory and Evidence from Rural Tanzania", Journal of Public Economics, 85(2), 235-273.

[19] _ _ _ _ (2002b), "Self-Help Groups and Income Generation in the Informal Settlements of Nairobi", Journal of African Economies, 11(1), 61-89.

[20] Miguel, Edward, and Mary Kay Gugerty (2005), "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya", Journal of Public Economics, 89(11-12), 2325-68.

[21] Otieno Okumu, C. (1999), "Promotion of Micro Enterprise Activities among the Youth as a Practical Alternative to Formal Employment: Focusing on Youths in Korogocho Slums, Nairobi", mimeo, Catholic University of Eastern Africa.

[22] Topa, G. (2001), "Social Interactions, Local Spillovers and Unemployment", Review of Economic Studies, 68(2), 261-95.

[23] Townsend, R. (1993), "Risk and insurance in village India", Econometrica, 62, 539-91.

[24] Wooldridge, J. (2002), Econometric Analysis of Cross Section and Panel Data, MIT Press.

Table 1: Descriptive statistics, group level

	Obs.	Mean.	Std. Dev.
Group size	16	14.75	7.27
Group duration (in years)	16	5.10	4.89
Group rules (number of groups)		*Yes*	*No*
Members recruitment: need or friendship	16	5	11
Group produces public good	16	7	9
Members specialize in production tasks	16	7	9
Members receive same pay	16	9	7
Group profits re-invested in group	16	7	9
Members can borrow from group funds	16	11	5
Membership decreased since foundation	16	9	7
Sanctions if miss contributions to group	16	4	12

Source: authors' calculations

Table 2: Descriptive statistics, individual level

	Obs.	Mean.	Std. Dev.
Income			
Total	276	3076	2789
From group	281	2384	2227
From outside group	276	661	2059
Before joining group	271	1346	2012
Age	282	30.3	11.2
Female	282	49%	
Education (years)	281	8.5	3.5
No formal school	281	9%	
Class 1-8	281	50%	
Form 1-4	281	38%	
Advanced educ	281	3%	
Kikuyu	282	37%	
Luhya	282	6%	
Luo	282	14%	
Kamba	282	37%	
Catholic	282	57%	
Other Christian	282	40%	
Muslim	282	2%	

Source: authors' calculations

Table 3: Correlation in incomes within group

Panel A: Group fixed effects, individual level regressions

Dependent variable:	*Total income (ln)*		*Income from group (ln)*		*Other income (ln)*	
	[1]		[2]		[3]	
	coef	*t*	*coef*	*t*	*coef*	*t*
group 1	-0.364**	-2.444	-0.312***	-2.825	-0.670*	-1.709
group 2	-0.244*	-1.879	-0.125	-1.133	-0.900***	-2.629
group 3	-0.011	-0.087	-0.001	-0.012	-0.336	-0.779
group 4	-0.762***	-3.832	-0.791***	-5.380	-0.328	-0.732
group 5	0.255**	2.002	0.375***	3.509	-0.900***	-2.629
group 6	-1.859***	-5.207	-2.335***	-21.530	0.022	0.038
group 7	-1.704***	-6.587	-2.153***	-17.898	-0.007	-0.015
group 8	0.447***	3.418	0.546***	4.970	-0.701**	-1.968
group 9	-0.254**	-2.110	-0.134	-1.364	-0.900***	-2.629
group 10	-0.929***	-4.535	-1.353***	-13.437	0.328	0.702
group 11	-0.917***	-3.174	-3.100***	-32.212	1.578***	3.610
group 12	-1.436***	-3.599	-2.239***	-13.123	0.507	0.855
group 13	-0.008	-0.033	-0.073	-0.499	0.061	0.099
group 14	-0.460***	-3.109	-0.536***	-6.214	-0.027	-0.050
group 15	-0.035	-0.199	-0.337*	-1.953	0.554	1.194
Constant	3.429***	30.953	3.309***	38.310	0.900***	2.629
No. Obs.	276		276		276	
Adj. R-sq.	0.51		0.88		0.24	

Panel B: Correlation among income pairs, dyadic regressions

Dependent variable: Correlation in	*Total income*		*Income from group*		*Other income*	
	[1]		[3]		[5]	
	coef	*t*	*coef*	*t*	*coef*	*t*
Same group	0.572***	6.297	0.919***	8.183	0.197***	3.522
Constant	-0.056***	-5.688	-0.088***	-6.596	-0.022***	-6.917
No. Obs.	74,256		74,256		74,256	

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%

t-values in Panel A are based on standard errors adjusted for heteroskedasticity using White's correction.

t-values in Panel B are based on standard errors corrected for clustering at the group level.

Table 4: Income regressions, individual level

Dependent variable:	*Total income (ln)*	*Income from group (ln)*	*Other income (ln)*
	[1]	[2]	[3]
	coef	*coef*	*coef*
Ln(earnings before)	0.083***	0.020	0.153***
	(2.727)	(0.974)	(2.898)
Years of grp membership	-0.001	-0.032*	0.084**
	(-0.066)	(-1.884)	(2.302)
Female	-0.395***	-0.187**	-0.278
	(-3.399)	(-2.178)	(-1.466)
Age	0.015	0.016	0.024
	(0.641)	(0.988)	(0.625)
Age squared	-0.000	-0.000	-0.000
	(-0.838)	(-1.388)	(-0.634)
Years of education	0.010	-0.012	0.056**
	(0.643)	(-1.165)	(2.086)
Kamba	0.006	-0.022	-0.183
	(0.041)	(-0.203)	(-0.752)
Luhya	0.252	0.052	0.075
	(1.392)	(0.422)	(0.242)
Luo	-0.244*	-0.140	-0.259
	(-1.652)	(-1.316)	(-1.058)
Constant	2.956***	2.708***	-0.942
	(6.420)	(7.890)	(-1.239)
Area of residence fixed effects	YES	YES	YES
No. Obs.	273	273	273

Notes: t-statistics in parenthesis

All t-values based on standard errors corrected for clustering at the group level.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 5: Correcting for selection bias, dyadic regressions

Dependent variable: Correlation in	*Total income (residual)*			*Income from group (residual)*			*Other income (residual)*		
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Same group	0.301***	0.310**	0.283*	1.024***	1.314***	1.889***	0.050	-0.020	-0.025
	(3.857)	(2.539)	(1.952)	(7.498)	(6.298)	(6.493)	(1.643)	(-0.619)	(-0.025)
Mills1		-0.015			-0.211*			0.060*	
		(-0.235)			(-1.774)			(1.785)	
Mills2		-0.065			0.512***			-0.029***	
		(-0.726)			(3.055)			(-3.893)	
Residuals			0.023			-1.178***			0.103
			(0.159)			(-4.149)			(0.103)
Constant	-0.013	-0.006	-0.011	-0.062**	-0.119***	-0.141***	-0.008	-0.005	-0.001
	(-0.766)	(-0.251)	(-0.508)	(-2.404)	(-3.282)	(-3.978)	(-0.008)	(-0.005)	(-0.001)
No. Obs.	74,256	74,256	74,256	74,256	74,256	74,256	74,256	74,256	74,256

Notes: t-statistics in parenthesis

All t-values based on standard errors corrected for clustering at the group level.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6: Selection equation, dyadic regression

Dependent variable = 1 if individuals i and j belong to same group				
			(for reference only)	
	Probit		*Logit*	
	[1]		[2]	
	coef	*t*	*coef*	*t*
Absolute difference in:				
Age	-0.038 ***	-6.49	-0.074***	-6.51
Gender	-0.775 ***	-5.92	-1.451***	-5.89
Years of education	0.009	0.79	0.020	0.86
Log of pre-group income	-0.036 *	-1.75	-0.057	-1.54
Luhya dummy	-0.271	-1.46	-0.498*	-1.68
Luo dummy	-0.363 ***	-3.07	-0.788***	-3.14
Kamba dummy	-1.295 ***	-8.13	-2.488***	-8.44
Catholic dummy	-0.095 **	-2.19	-0.170**	-2.31
Sum of:				
Age	0.005	1.24	0.008	1.14
Gender	-0.174 ***	-2.58	-0.361***	-2.66
Years of education	-0.017 *	-1.70	-0.034*	-1.68
Log of pre-group income	0.002	0.09	0.007	0.18
Luhya dummy	0.175	1.01	0.233	0.90
Luo dummy	0.051	0.77	0.119	0.82
Kamba dummy	0.661 ***	7.16	1.159***	7.21
Catholic dummy	-0.117 *	-1.92	-0.231*	-1.95
Distance between residence	-0.809 ***	-13.66	-1.487***	-12.24
Constant	0.683	1.89	1.526**	2.25
No. Obs.	74,256		74,256	

Note: * significant at 10%; ** significant at 5%; *** significant at 1%

Table 7: Formal income pooling, dyadic regressions

Dependent variable: Correlation in	*Total income (residual)*			*Income from group (residual)*			*Other income (residual)*		
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Same group	0.352***	0.319***	0.314**	1.026***	1.311***	1.869***	-0.005	-0.029	-0.056
	(3.952)	(2.637)	(2.177)	(7.417)	(6.239)	(6.679)	(-0.597)	(-0.988)	(-0.056)
Same group and same pay	-0.221*	-0.229*	-0.228*	-0.007	0.062	0.153	0.239**	0.233**	0.229**
	(-1.717)	(-1.817)	(-1.702)	(-0.023)	(0.196)	(0.466)	(2.152)	(2.065)	(2.095)
Mills1		0.025			-0.222**			0.020	
		(0.440)			(-2.039)			(0.681)	
Mills2		-0.065			0.512***			-0.029***	
		(-0.726)			(3.055)			(-3.893)	
Residuals			0.054			-1.198***			0.072
			(0.356)			(-4.010)			(0.072)
Constant	-0.013	-0.006	-0.009	-0.062**	-0.119***	-0.143***	-0.008	-0.005	-0.003
	(-0.766)	(-0.251)	(-0.403)	(-2.404)	(-3.282)	(-3.828)	(-0.008)	(-0.005)	(-0.003)
No. Obs.	74,256	74,256	74,256	74,256	74,256	74,256	74,256	74,256	74,256

Notes: t-statistics in parenthesis

All t-values based on standard errors corrected for clustering at the group level.

* significant at 10%; ** significant at 5%; *** significant at 1%

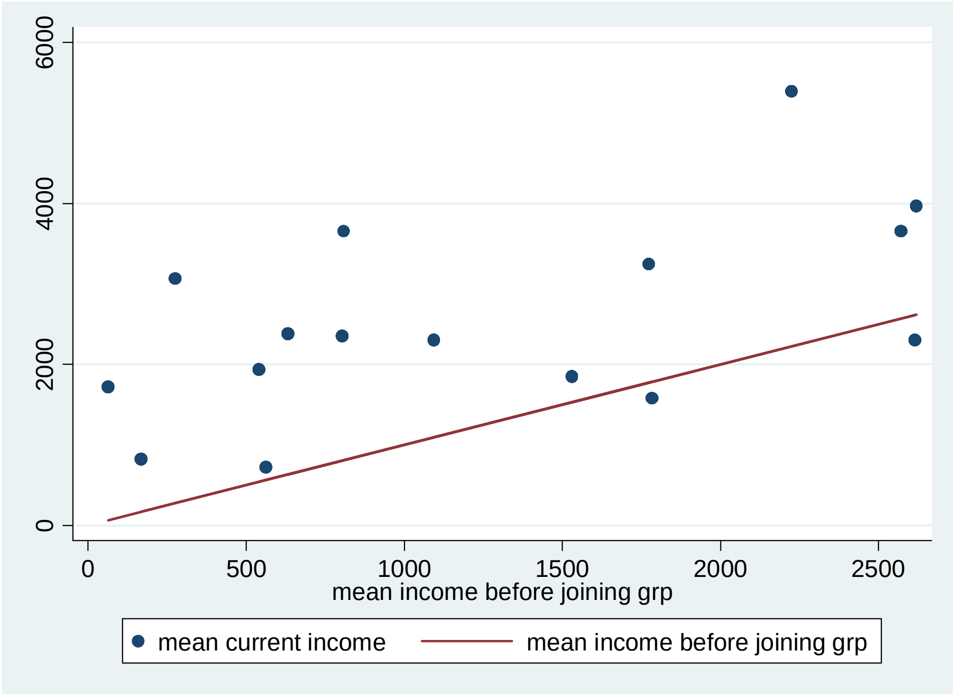

Figure 1:
Mean of total income before and after joining the group

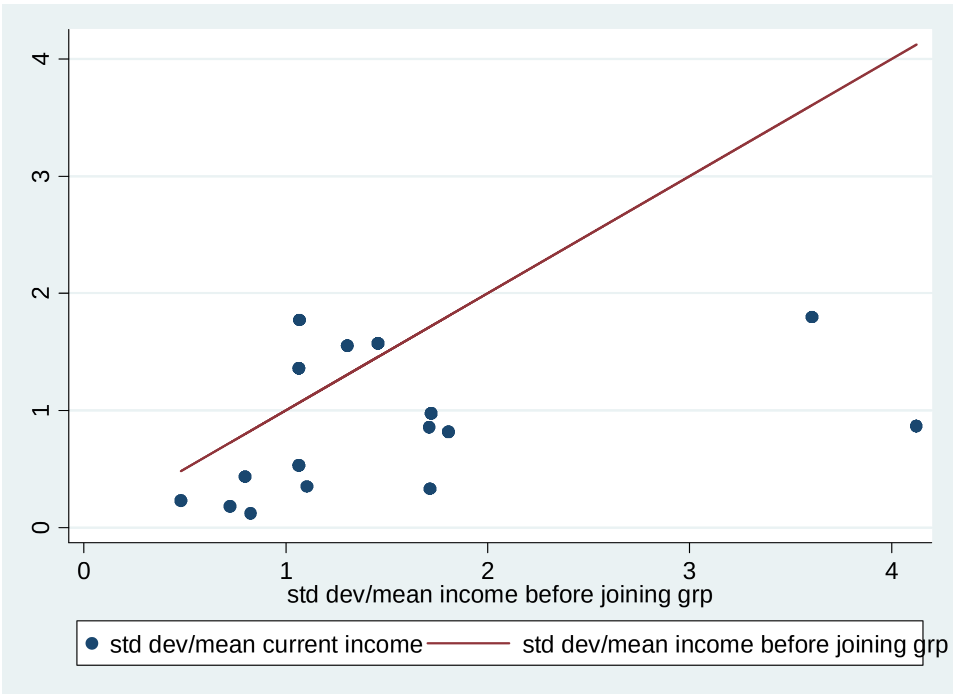

Figure 2:
Ratio of std dev to mean income before and after joining the group