

Forms of Social Capital in Present-Day Soweto: A Factor Analysis of Household Survey Data

Barbara Piazza-Georgi

Econometric Research Southern Africa, and

School of Economic and Business Sciences,

University of the Witwatersrand

ABSTRACT: This paper analyses data on social capital, gathered in Soweto in a 1,500-household survey in 1999 (see ERSA policy papers no. 4 and 15), mainly with factor analytical techniques, in order to reveal any patterns and fit them into meaningful categories in terms of the specific social context, of existing theory, or of new insights.

Seven primary factors and three secondary ones have been identified; the two strongest characteristics are a distinct form of social capital typical of higher-income households, and a distinction between trust in formal institutions and other forms of trust.

1 Background and objective

1.1 Conceptual framework

The definition of capital has expanded over the last century to include human, institutional and social capital.¹ There is an abundance of empirical work on the relationship between various social, institutional and human capital factors and growth indicators,² mostly in the framework of endogenous growth theory. However, a point is raised by Fedderke and Klitgaard (1998) that indicates the need for further empirical study on a different plane: namely, that these factors, identified as playing a role in growth, do so within the framework of what the authors call a “web of associations” among themselves, correlations of various kinds, of various degrees of robustness and with often undetermined causal directions. Thus there is a very real potential for finding spurious associations in such empirical growth studies. It may be useful, therefore, to step back from integrated growth studies, and instead to focus on the details of such “webs of associations”, to throw light on the actual mechanisms by which the different human, institutional and social capital indicators might influence growth, directly or indirectly.³

When testing the “webs of association” theory on the basis of empirical data, the first task is to establish the indicators to use for the web nodes. Detailed indicators, such as are gathered in household surveys, may be too numerous and may yield results too complex for direct analysis; composite indicators inevitably raise methodological and theoretical questions about the rationale behind their composition. This is the case particularly for social capital, which is a complex concept that is still in the process of being satisfactorily defined. It is worthwhile at this point to recapitulate briefly the major categorisations that have been used so far in social capital literature.

Coleman (1990) distinguished six categories of social capital: expectations, information channels, norms and values, authority relations, and two kinds of social organisations. Later literature focused on these and similar categories: e.g. level and intensity of group life; level of interpersonal trust

¹The seminal authors associated with this development are: P. Schumpeter, T. Schultz and G. Becker (human capital); R. Coase and D. North (institutions); R. Coleman, R. Putnam and F. Fukuyama (social capital). A more detailed review and bibliography is given in Piazza-Georgi (2001c).

²Fedderke (1997) gives a useful summary of this work

³This background was set out in Piazza-Georgi (2001c).

and trust in institutions; family/social networking and cohesion. Trust, in particular, was exhaustively described by Fukuyama (1995); social networks and groups were highlighted by Putnam (1993). Uphoff (1999) included all the above categories by dividing them into “cognitive” (the less tangible side, including norms, values and attitudes) and “structural” (local institutions, both formal and informal).

The term “social capital” can also include “formal” institutional capital, which had in fact been recognized and analysed earlier (Coase’s seminal paper was published in 1960) and extensively analysed by the institutional economics school. In recent years has there been a tendency to define both of these as forms of social capital.⁴

Most empirical studies done so far on social capital, have used a deductive approach, gathering empirical data on the basis of pre-accepted categories such as those above, and confirming or rejecting hypotheses about their effects on various socio-economic indicators. For example, the World Bank Tanzania study⁵ derived a single indicator from data on respondents’ associational activity and different forms of trust, and matched it with household income; Knack and Keefer (1997) used indicators of trust and civic norms at country levels, finding positive links with income, education and other social indicators; Paxton (1999) gathered data on two kinds of trust and on associational activity, analysing their changes over time and thence coming to a conclusion about the evolution of social capital. Such studies illustrate well the “webs of association” problem. It could be argued, for example - as indeed is done by Fukuyama - that different kinds of norms and values function by creating different forms of trust and of associational activity, and it is the latter that has the direct influence on income levels, through lowering transaction costs. Then, rather than treating associational activity, norms/values, and trust as separate kinds (nodes) of social capital, as the above studies imply, it would be more appropriate to combine particular kinds of values with particular kinds of trust and of resulting associational activity, into one node of social capital, distinct from another node which might represent *other* kinds of values, trust and associational activities.

Thus, on a theoretical level, there is still work to be done in regrouping the items that constitute the operative categories of social capital. A simplistic though useful division is that between “bonding” and “bridg-

⁴See e.g. Stiglitz, in Dasgupta and Stiglitz (2000)

⁵Narayan and Pritchett (1997)

ing” social capital - i.e. values/norms that bind a community closer together, and values/norms that help to link different communities. Woolcock (1998) developed a more rigorous theoretical categorisation, on the basis of the Polanyi/Granovetter concepts of embeddedness/autonomy⁶ (which are related to bonding/bridging), by adding the interacting dimensions of the macro/micro level. Embeddedness at the micro level (intra-community links) he calls Integration, at the macro level (State-society relationships) he calls Synergy; autonomy at the micro level (inter-community links) is termed Linkage, and at the macro level (institutional competence and coherence) is termed Organisational Integrity. One is thus faced with four separate scales of values, giving 16 social capital profiles if each attribute is rated low/high. Thus different combinations of these four dimensions can account for a range of development outcomes, from “anarchic individualism (the absence of all four dimensions) at one extreme, to beneficent autonomy (the presence of all four) at the other. The same dimension of social capital can thus serve very different developmental purposes when combined with other forms”.⁷

More recently, Fedderke et al.(1999) focused on the dynamic aspect of social capital formation with the use of two dimensions that are related to the ones just mentioned, but defined in functional terms: “transparency” and “rationalisation”. Transparency, loosely related to “bonding” capital, refers to “the clarity and ease of access to social rules, norms and values, whatever form those rules may take, and [the degree to which] rules are consistently reinforced by negative and positive sanction”.⁸ Rationalisation, loosely related to “bridging” capital, refers to increasingly explicit emphasis on *function*, on formally codified norms of a procedural, rather than substantial form, having universal scope (i.e. more explicitly expressing the *rationale* for the rule).⁹ This definition has the merit of focusing on the rationale behind relationships and the logic behind the rules, distinguishing between the simpler, but less flexible, substantive rules or relationships, and those that focus more “upstream”, closer to the ultimate objectives to be attained and easier to apply, or harmonize, across diverse societies - e.g. modern social norms about healthy eating, as against religious rules on periodic fasting; formal insurance institutions, as against community solidarity.

⁶References quoted in Woolcock, p.5.

⁷Woolcock (1998), p.5.

⁸Fedderke et al. (1998), p. 727.

⁹In other terms: rationalisation = how sophisticated the rules are; transparency = how clearly they are understood and how well they are implemented.

While the above categorisations are clearly helpful in advancing our understanding of social capital, they are the fruit of theoretical, not empirical studies. To use them in an empirical study means accepting *a priori* definitions of what constitutes a meaningful category. It seems important therefore, first to use an inductive approach, starting from the data itself and examining the clusters that they form, i.e. groups of variables that exhibit relatively high degrees of correlation. The meanings of these clusters could be expected to match, in part, the categories of social capital that have been hypothesised in previous literature; some clusters might prove to be specific to the particular social environment in which the survey was carried out. Some clusters, on the other hand, may lead to new insights about the nature of social capital, and these would be of particular interest for further study.

A few such studies have been done recently, or are being done at the time of writing. Two comprehensive social capital indices are being developed in the USA.¹⁰ A major attempt was the BARCAS (Barometro de Capital Social) study done in Colombia in 1997-98,¹¹ that found ten dimensions of social capital and calculated their weights within one aggregate social capital index; a study done among five Australian communities identified eight dimensions and a single social capital factor.¹² The 1999 World Bank Conference on Social Capital and Poverty Reduction¹³ attempted, among other things, to draw together and share recent empirical experience in this field and standardise the research methods.

1.2 Objective of this paper

This paper was written in the context of a wider empirical study that the author is undertaking, with the purpose of developing the “webs of association” theory mentioned above, looking at different dimensions of human capital and of social capital. For this purpose, a household survey was conducted in Soweto, South Africa, to collect data on a number of detailed indicators of human and social capital, both at the individual and the household level.

It was decided to preface the main analysis by an inductive exploration of the social capital data, along the lines of the methods outlined in the above section: namely, to examine the patterns of the data and attempt to fit them

¹⁰References in Woolcock and Narayan (2000), p. 241.

¹¹Sudarsky (1999).

¹²Onyx and Bullen (2000).

¹³Krishna and Shrader (1999).

into meaningful categories - meaningful in terms of the specific social context being surveyed, of existing theory, or of new insights. It should be noted that this analysis had not been anticipated at the time of the survey design, thus the data are in a form that presents some difficulties; it is felt, however, that these do not invalidate such an analysis, and attention will be drawn to them as appropriate.

This paper, then, examines the social capital data of the survey - its largest and most complex group - taking the inductive approach, as explained above. (The human capital data may be similarly examined in a subsequent paper.) In testing the patterns of survey data against existing hypotheses, it is intended to pay particular attention to functional categories, i.e. the simple bonding/bridging function, Woolcock's dimensions, and Fedderke et al.'s transparency/ rationalisation scales. (The latter's high-rationalisation/ low-rationalisation categories will be referred to from now on as "low-rat" and "high-rat".)

Specifically, *the research questions for this paper are: what patterns are visible in the survey data that relate to social capital? Can these patterns be meaningfully fitted into the categories found in recent social capital literature, as outlined in Section 1 above? If not, what other interpretation can they be given?*

1.3 Description of the survey¹⁴

Soweto, an area of approx. 70 sq. km. on the south-western fringes of Johannesburg and containing just under one million inhabitants, was chosen as the survey area. Soweto was developed since the 1930s to serve as a "dormitory city" for Black workers having jobs in the Johannesburg area, and later also used as a resettlement area for Blacks expelled from "white" areas under the apartheid laws. It contains several informal settlements, workers' hostels, and private housing areas, in addition to the typical single-family Council-built houses, many of which contain "backyard shacks" rented out to other households. Since the demise of apartheid (and also before, illegally), it serves as a major destination point of rural-urban migrants. Being still almost 100% Black,¹⁵ Soweto is racially homogeneous, but highly diverse in most other respects (income levels, social origins, ethnic groups, education);

¹⁴This information is set out in more detail in Piazza-Georgi (2001a)

¹⁵"Black" is used in the South African sense, meaning persons of pure African descent, i.e. excluding persons of mixed race (called "Coloured") and of Asian origin.

in particular, it allows an examination of both traditional and modern forms of association and of social capital. It was felt to be important to conduct the study in a racially homogeneous area, since apartheid policies caused significant differences in economic, social and educational conditions among the main racial groups: these would have shown up as different forms of social capital, which indeed they are, but it might have been difficult to extricate the underlying patterns from the unique, situation-determined ones. It was felt that by restricting the study to one racial group, more light could be shed on the patterns that may have a more general validity.

The survey was carried out over a four-week period in July 1999; the interviewers were senior economics students from Vista University, Soweto. It was aimed to capture data from about 1,500 respondents from 1,000 households. The sampling was done on a stratified two-stage cluster basis. The strata were the six main categories of residential areas (various eras of council housing, new private housing, informal settlements and workers' hostels), and sampling clusters were designed to capture some further possible loci of neighbourhood-based social capital (areas with common services or ethnic characteristics.) The respondent could be any household member over 18, not necessarily the head of the household. An additional adult was interviewed in approximately every third household. Interviewers were assigned to areas inhabited, as far as possible, to similar ethnic groups as themselves; they worked at hours that maximized the presence of those who worked full-time.

The questionnaire was designed in three parts. The first part gathered basic demographic data on all members of the household, as well as detailed household expenditure data, the latter to be used both directly (e.g. expenditure on education and on social obligations) and in total, as a proxy for household income. The second part gathered more detailed data on the human and social capital of the individual respondent. The third part was an additional questionnaire, which focused on entrepreneurial practices, skills and attitudes: it was administered only to self-employed and informally employed respondents.¹⁶ After the elimination of the questionnaires that presented problems, the final sample was 1,186 respondents from 908 households, and 200 additional questionnaires. The 908 households had a total of 3,072 adult members, on whom basic demographic, education and employment sta-

¹⁶This third part was more detailed than would have been necessary for the purposes of this study, since it also collected separate data that the World Bank, which funded the survey, needed for its own purposes. The information collected in this third part is set out in Piazza-Georgi (2001b).

tus data were collected and could be compared to household structure and expenditure data.

2 Methodology

2.1 Choice of methodology

The research question was formulated in Part I as follows: what patterns are visible in the survey data that relate to social capital? Can these patterns be meaningfully fitted into the major *a priori* categorizations? Are there - in addition to, or in place of, the former - other categories that seem appropriate to define the different forms of social capital found in the context of this study?

The survey contained about 30 questions, some of them multipart, relating to the social capital endowment of the respondent, resulting in about 100 raw variables. About half are dummy variables, i.e. with only two possible values; most of the others are categorical variables with three to ten possible values.

The raw variables are grouped into indicators of values, social networks, group membership, and of trust. (More details will be given later in this paper, under the four headings.) However, as already mentioned, this is done merely for initial convenience. The purpose of this paper is to “let the data form its own groups”, and to assess the meaning and validity of these groups as potential “web nodes”. These will be the “social capital factors” that can later be examined in conjunction with human capital and other factors, in the core analysis of the “webs of association”. We need to establish:

- Which variables to incorporate into each factor. We have some *a priori* assumptions, based on definitions used in past literature, in particular that of the relationship between values, social networks and trust, as well as that of low-rat and high-rat social capital; but, as already mentioned, the intention is to test these as hypotheses rather than accept them as given.
- What weight to give to each variable within a factor. E.g., once established that the factor “present level of low-rat social capital” incorporates membership in churches, burial societies and charitable groups,

extent of family structure, strength of family and ethnic links, and community-related values - we still have to establish what weight to give to each when constructing the composite indicator (factor) representing the level of low-rat social capital.

Econometric techniques present problems with the kinds of issues that we have here. The data, by its nature, may not exhibit strong correlations, since social capital is only one set among a complex web of links between socioeconomic variables, thus one assumes the existence of important exogenous factors, leading to strong heteroscedasticity; also, the measurement of social capital variables is not precise. The specification problem is particularly complex. We are faced with a large number of variables which are expected to fit into a number of functions (each function being a dimension of social capital), but we do not know a priori which variables combine into which function, and how. *A priori*, the web-of-relationships theory leads us to expect simultaneous-equation models.

What is required to begin with, is a technique to determine complex patterns of correlations among the variables, identifying clusters that seem to belong together and that enable us to set up hypotheses about the dimensions of social capital in the survey area.

Factor analysis is a suitable technique for such a task. It addresses both of the above issues - it reveals patterns of correlations in large sets of variables, and it establishes a set of weights to assign to the variables within each correlation cluster - in the same operation. Section 2 gives an overview of factor analysis.

2.2 Summary of relevant factor analysis techniques

Factor analysis is a statistical technique used in order to disentangle complex interrelationships within large sets of data. It is built upon the matrix of correlation coefficients (R). In a large study with over 100 variables, the correlation matrix has over 5,000 items: in such cases, some simplifying technique is needed to discern the patterns. Factor analysis creates a set of artificial constructs (factors), which are a “condensed statement of the relationships between a set of variables” (Kline, p.5) and operationally defined by the correlations of the different component variables with the factor.

An understanding of factors can be approached either geometrically or algebraically, and in the following sections both will be used.

2.2.1 Basic principles¹⁷

In the traditional mathematical approach, relationships between variables are expressed in a mathematical function $f(X, W, Z)$ connecting Y with X, W and Z . Assuming throughout that the relationship is linear, one expresses such a relationship in the form $Y = \alpha X + \beta W + \gamma Z$.

In simple algebra (simultaneous equations), α, β and γ are known: more precisely, one knows as many sets of values for them (equations) as their number - three in this case - in order for the equations to be solvable. (It is not necessary to know more sets of values, since it is assumed that the relationships are the same.) The task is to find values for Y, X, W and Z that satisfy all equations.

In econometrics, Y, X, W and Z - a large number of sets - are known; the task is to determine the values of α, β and γ , in a situation where there is no one clear solution, since each set of values is distorted by the stochastic error. Thus one looks for the values of α, β and γ , that give the best possible approximation of the observed values of Y, X, W and Z (by OLS or other criteria). The variables are all defined, data are available, and the question is to determine the best function for illustrating the relationship.

However, in factor analysis we address a situation that is one step more complex. Again we have a large number of values for Y, X, W and Z , but we are looking for a relationship not between Y and X, W, Z , but between Y and *a number of different functions of X, W, Z* . For example, we may know the time spent on self-education by each of a sample of n persons ($Y_{1...n}$), and a large array of social data $(X, W, Z)_{1...n}$ on each of these persons, but what we are interested in are some unknown "social capital characteristics", $F_{l...j}(X, W, Z)$, which, if identified, would give a simpler, though less precise explanation for variations in Y , resting on j elements instead of the full number of the original variables. This is of practical advantage if one is able to identify, say, five factors, which account for a useful part of the variation in about 50 variables; and of considerable theoretical interest, if the factors thus identified highlight previously unknown relationships between the original variables.

In other words, we have a multivariate composite array, of which the intermediate variables (artificial constructs, to be called factors) are unknown, or merely hypothesised and impossible to measure directly. What factor analysis does is to *confirm the existence of these variables (factors) and define*

¹⁷This section draws heavily from Rummel

them, by detecting patterns of correlations in the primary, known variables. From a set of observations W, X, Y, Z , it extracts a set of relationships:

$$\begin{aligned} Y_1 &= \alpha_{11}F_1 + \alpha_{12}F_2 + \alpha_{13}F_3 \cdots \cdots \alpha_{1m}F_m \\ Y_2 &= \alpha_{21}F_1 + \alpha_{22}F_2 + \alpha_{23}F_3 \cdots \cdots \alpha_{2m}F_m \\ Y_3 &= \alpha_{31}F_1 + \alpha_{32}F_2 + \alpha_{33}F_3 \cdots \cdots \alpha_{3m}F_m \\ &\cdots \\ &\cdots \\ Y_n &= \alpha_{n1}F_1 + \alpha_{n2}F_2 + \alpha_{n3}F_3 \cdots \cdots \alpha_{nm}F_m \end{aligned}$$

where: Y, X, W, Z = variables with n known observations

α_{ij} = a constant, representing the *loading* on that factor

$F_{1\dots m}$ = m functions $f()$ of variables X, W, Z (usually a large number of variables, of which each function might only use a few, i.e. many of the coefficients may be = 0). *These functions are the factors.* They are assumed in practice, as for much econometric work, to be linear; variables may be included in their log forms, so as better to capture non-linear relationships.¹⁸

In practice, factor analysis is done with matrix algebra methods, processing the correlation matrix to extract characteristic vectors, which are rescaled to give the principal factors. Each vector then represents one factor; each item in the vector is the *factor loading* of the corresponding variable on that factor (α_{ij} from the above system of equations) i.e. the correlation of that variable with the factor. Absolute values for the loadings that are higher than .3 or .5, depending on the nature of the data, are considered to be substantial; as in the case of equation coefficients, negative values have the same importance as positive ones, except that they indicate negative relationships. The eigenvalue associated with each factor, being the sum of the squares of the loadings, is a useful immediate indicator of the usefulness of that factor: the higher the eigenvalue, the more of the total variance of its component variables is explained by that factor.

Factor analysis methods extract the factors in order of importance, i.e. from the highest eigenvalue down. Thus, the most salient commonality of the variables is identified first and extracted from the matrix (“partialled out”), then the residual is processed again to find the second most salient

¹⁸This was done in this study for some of the variables: household expenditure, in particular, proved to be better related to other variables in its natural log form.

commonality that is uncorrelated to the first, and so on. Principal components analysis extracts as many factors as there are variables: thus the sum of the eigenvalues is equal to the number of the variables, and their average is 1. Therefore, an eigenvalue of 1 or less indicates that the factor does not explain more of the variance of its component variables, than one of the single variables would: as we shall see below, this is the first criterion for the retention or non-retention of a factor. The decision on the number of factors to retain is a crucial one, since the raw method, yielding as many factors as the original variables, does not yet permit one to achieve the main aim of factor analysis, which is to reduce the number of variables to be analysed. Once the reduction of factors has been effected, a value called “commonality” can be calculated for each variable; it represents the degree to which a particular variable is predictable in terms of the chosen factors. STATA, the software used in this study, calculates the inverse instead, called “uniqueness”: 60% uniqueness is the equivalent of 40% commonality, both meaning that the retained factors explain 40% of the variation of a variable, while the other 60% would be lost if the variable was replaced by the retained factors.

2.2.2 Rotating the factors

As a consequence of the algebra used for extracting the factors of a set of data, the first factor will generally have high positive loadings on a number of variables, while the subsequent ones will have smaller, bipolar loadings. This does not in fact reflect reality, but is merely an artefact of the algebraic method used. To get interpretable loadings, one must first eliminate this artificial set of high loadings; this is done by *rotating* the factors. This process can be visualised geometrically by realising that the first factor extracted will “fall through the centre of gravity” of all the clusters, obscuring the distinction between them. Rotation changes the position of the factor so as to go through the centre of gravity of the most distinct cluster, making it possible to calculate the real variance of each variable of the cluster from it.

The subsequent factor, as mentioned above, is calculated so as to be orthogonal (uncorrelated) to the first, and so on for each factor. The basic kind of rotation (called VARIMAX) maintains this orthogonality; one can however also use other kinds of rotation that do away with this constraint, thus calculating factors that may be correlated to each other. One such method of rotation, called PROMAX, will be used in this analysis in preference to VARIMAX, since the nature of social capital data is such that one cannot *a*

priori eliminate the possibility of correlated factors.

It can be demonstrated mathematically (Kline, among others, does so) that rotation of factors only changes the loadings on each variable (to a more easily interpretable form), but does not change the eigenvalue, i.e. the *sum of the squared factor loadings, which gives the proportion of the variance of the variables that is explained by the factor* - which is the purpose of the exercise.

2.2.3 Choosing the method of factor analysis and interpreting the results

The basic method of factor analysis, as described above, is called principal components (PC) analysis. There are other variations of this technique, the most important one being principal factor (PF) analysis, although in large samples such as the present one, the differences are small. The PF method of factor analysis has the advantage of calculating, for each variable, the proportion of the variance that is common with the other variables included in a particular factor - i.e. the variance due to the factors in which that variable plays a part - and including only that proportion, rather than the total variance, in the analysis. The PF method is superior to PC for detecting the structure of the data, although it has the disadvantage of being less easy to interpret in terms of the proportion of variances being explained and the significance of the eigenvalues. It is nevertheless generally preferred, and was the method of choice for this study, although PC analysis was also used to obtain information from the eigenvalues.

Factor analysis brings out a number of factors in order of relevance, but the decision has to be made about the cut-off point below which one considers them unimportant. This is, by nature, an arbitrary decision, as texts are careful to point out;¹⁹ however, some guidelines are commonly applied and give good results in practice. A classic one is to disregard any factors which yield an eigenvalue of less than 1 in PC analysis, i.e. which extract less than the equivalent of one original variable; however, this may still leave too many factors. (In the present study, ten factors have eigenvalues >1 under the PC method, yet only five have been retained, the weakest having an eigenvalue of 1.6.) A more tailored solution is to use the “scree test”,²⁰ based on graphing

¹⁹E.g. Kline (1994), p. 43; StatSoft (undated), p.5.

²⁰Scree (in geology): the loose gravel that collects at the bottom of a steep slope.

the eigenvalues from the highest to the lowest, and identifying a point, at a low eigenvalue, at which the slope of the resulting line undergoes a visually obvious change from steeper to less steep. At that point the eigenvalues of the succeeding factors become less distinct from one another, implying that the factors may just represent “scree”.

A very important question is how to interpret the loadings of the variables in each factor. Examples used in textbooks show loadings of .5 and above as the substantial ones; this implies that the factor explains .5², or 25%, of the variance of that variable. That standard may be too rigorous for a study involving social capital, since the latter is generally not a primary factor, but it is known to be overshadowed by stronger social and economic influences. We know that social capital is composed of many variables - Coleman (1990) distinguishes 12 forms - and we are still in the process of defining them; the influence of each of these may be relatively minor. Thus, knowing that the correlations between social capital factors are not expected to account for all, or even the bulk, of variances in socio-economic data, factors explaining 10% of a variance are still definitely of interest: accordingly, loadings of .3 (explaining 9% of the variance) have been regarded as substantial.

The lower limit for taking a variable into consideration, in this study, is around .1 (indicating that the factor accounts for roughly 1% of the variable's variance). The reason for this is that, in a sample with the size of this one, it is around that point that *statistical significance* needs to be checked. Factor analysis does not include a direct calculation of the statistical significance of loadings, but one can extract it from the relevant portion of the correlation table of the factors with their component variables.²¹ In the present sample, it has been found that loadings around .1 may already be statistically insignificant at the 5%LS (there is no precise cut-off point, since it depends on the size of each pairwise sample being correlated). Thus in this study, loadings of .1 or more are considered *a priori*, but the weakest are checked for statistical significance by going back to the correlation table, and if insignificant, they are discarded.

Although this survey was originally designed for regression analysis only, factor analysis can be used with good results on many parts of it, since the sample is large. On a sample of about 1,000, such as we have, even dummy variables will show good results under factor analysis, although the relation-

²¹Kline (1994), p.52.

ships will be weaker than with numerical data.²² The problem is the same as that of using categorical variables in a regression: a multichoice question with say, seven possible choices - whether only one choice can be made or several - needs to be converted into a set of six dummy variables for factor analysis. There is a number of such sets in this particular questionnaire, leading to a fairly large number of dummies. As will be seen later, in some cases various forms of manipulation were used to reduce the data into discrete numerical variables rather than such sets of dummies.

3 Preliminary work on social capital variables: derivation of composite variables and sub-factors

As mentioned before, this study gathered social capital data of three categories: (1) values, (2) social networks and group membership, and (3) trust. Before attempting to extract the overall patterns of interrelationships, it was found necessary to do some preliminary work, simplifying the data, extracting preliminary information, and/or adapting it to a form more suitable for factor analytic techniques. The sections below summarise the different processes that were applied to the three groups of data.

3.1 Values variables

The “values” dimension of social capital - expectations, obligations and norms - was the backbone of Coleman’s early analysis (together with the use of social relations as information channels, to which we will return in the next section, on social networks). Its study goes back much further, with roots stretching back to Max Weber with his concept of the “Protestant work ethic”:²³ one can regard the latter as the first definition of what today we would call social capital. “Values” is a generic term that can include all of Uphoff’s “cognitive” items - values, attitudes, norms. Trust may also be thought of as a value, but it is best treated separately, as explained below.

The survey had four questions on values: (1) agree/disagree to six statements which showed up the respondent’s attitude to wealth, community cohesion, and risk-taking; (2) a question about the minimum price that the

²²Kline (1994), p.126.

²³Weber, M.: “The Protestant Ethic and the Spirit of Capitalism”, 1904-05.

respondent would require for an additional hour of work, measuring the relative value attributed to work and leisure; (3) an open question on the respondent's interpretation of the word "Ubuntu";²⁴ and (4) to rank, in order of importance, the following values: generosity, honesty, cleverness in business, family loyalty.

The *first question* contained six sentences, designed in pairs (but mixed up on the questionnaire) to elicit the respondent's attitude to wealth, to risk-taking, and to community values: four levels of agreement/disagreement were possible to each sentence. The responses to each pair were added up in such a way as to result in 0 or 1 not only if they expressed neutrality, but also if they were mutually contradictory; but to attain maximum values of +4 or -4 if they were mutually consistent in expressing a strongly positive, or negative, attitude on that particular point. The question thus yields three variables, with values from -4 to +4: these are given the names ATTRICH, ATTRISK and ATTCOMM.

The *second question* yields one variable, with values ranging from 1 to 7, representing increasing amounts of money needed to induce the respondent to work an additional hour. (1 = SAR1, about US\$ 0.15; 7 = more than SAR 30, about US\$ 4.) It is labelled MARGW. It should be noted that this value is not the same as the reservation wage: it is a marginal value, that attributed to an extra hour of work in the context of the respondent's present use of time, not the minimum hourly wage that would entice him/her into a full-time job. This variable is a difficult one to interpret, since it does not represent a value as such, but indicates it in a particular context - e.g. the response given by an unemployed person living in a low-income household indicates a very different value from an identical response coming from an already fully employed person living in a high-income household. This variable is expected to be more useful in forthcoming regression analyses, than in this paper; nevertheless, it was decided to include it, in case it shows any unexpected behaviour.

The *third question* needed processing of some kind, since it was an open question. It was decided, for the purposes of this first analysis, to identify responses that reflected non-norm interpretations of the word *ubuntu*. Normally, the word refers to community values, positive interpersonal rela-

²⁴"Ubuntu" is a widespread traditional concept among the Black population of Southern Africa, revived and widely used today in politics and business life, as a summary of African social values (see Mbigi & Maree 1995), often in contrast with Western individualism. It is defined by the phrase: "A person is a person through other persons".

tionships. Some responses, however, had a conservative/authoritarian slant, referring e.g. to norms, tradition, obedience; these were labelled UBAUTH. Others had an individualistic slant, referring to personal standards (e.g. honesty, sense of humour, self-respect); these were labelled UBIND. About half the responses reflected one of these slants, and were therefore assigned a value of 1 on one of these two dummy categories, which are mutually exclusive (i.e. the same respondent cannot score 1 on both). The other half are not given further note, on the basis that they represent a normal, therefore neutral, interpretation of the concept.

The *fourth question* (to rank, in order of importance, the following values: generosity, honesty, cleverness in business, family loyalty) presented a problem for factor analysis. It yields four variables, each with a value of 3 (if mentioned first) to 0 (if mentioned fourth): but they are perfectly correlated, i.e. the fourth can be predicted perfectly from the other three. This makes these variables unsuitable for factor analysis, either in conjunction with just the other values variables, or even with the full range of social capital variables. In order to be able to incorporate this group of variables, albeit imperfectly, into a factor analysis of the social capital variables, a single scale needed to be created out of the four variables, either by taking just one variable into account, or by combining two. All three possible pairwise combinations, as well as each of the four original variables, were tried out in a series of trial factor analyses. One pairwise combination showed stronger correlations with other social capital variables, i.e. resulted in factors with higher loadings, than any of the others. This combination was: honesty and cleverness in business, as against generosity and family values. One could interpret these two pairs as “individualistic” versus “communitarian” values.²⁵ The resulting variable, that sums the scores on the value attributed to honesty and to cleverness in business, was called INDIVVAL. The conclusion is that the most useful way of characterising a respondent’s value system on the basis of the data of this survey, is to measure the degree to which “individualistic” values are prized more than “community-oriented” ones.

There are therefore seven variables in the “values” group, that are to be introduced into the general analysis: ATTRICH, ATTRISK, ATTCOMM, MARGW, UBAUTH, UBIND, and INDIVVAL.

²⁵There may be some uncertainty as to whether honesty is in fact an individualistic value, since it does refer to relationship with others and the sense of reciprocity; perhaps one can say that it is “indirectly” communitarian, while the other two are directly so.

3.2 Social networks variables

Social networks are another major category of social capital, anticipated in Coleman (1988) in their role as information channels, and expanded in Putnam (1993) and subsequent work. Uphoff's taxonomy distinguishes them from other forms, by calling them "structural" aspects of social capital. The World Bank Tanzania study²⁶ based its analysis on group participation (measures of social cohesion and trust were also included); it was also one of the early works where the risk-sharing purpose of social networks was recognized, as distinct from the information-carrying function.

For this study, the aim was to capture both the risk-sharing and the information-carrying roles. For the latter, it was decided to focus particularly on one form of information-carrying, namely the job-search role of social networks. The reasoning was that employment is the fundamental economic issue in an urbanised environment such as Soweto; and as Wittenberg (1999) demonstrates, there are links in South Africa between social networks and the rate of unemployment.²⁷

On a general level (i.e. with possible relevance to any role of social networks) there was a detailed set of questions on *group membership*, i.e. the organised groups to which the respondent belongs; these were one of the resources mentioned in the questions on job search and safety-nets, but they might also have other roles. Further, a question was asked about neighbourhood networking.

There were thus four questions on this aspect of social capital:

- *To assess group membership.* This was a fairly lengthy question. Each respondent could mention up to six groups of which he/she was a member (but none mentioned more than three), and was then asked a number of questions about each; this resulted in about 12 variables per group, covering various characteristics of the group and the respondent's level of participation in it.
- *To assess social safety nets:* "If the main income source of the household were to fail, where would you find help?" (10 choices, including an open answer and "I don't know"; respondents chose up to three.)

²⁶Narayan and Pritchett (1997).

²⁷He focuses in particular on household structure, i.e. the presence of employed persons in the household who provide what he calls an "insider network".

- *To assess the role of social networks in job search*: respondents chose one out of eight possible job-search strategies that (a) got the respondent his last job, (b) would be used now if he was job-hunting. The choices included non-network strategies, informal networks and specific employment networks.
- *Neighbourhood networks*. “Do the people of this neighbourhood get together to solve some problems? Sometimes? Often?”

The variables resulting from three of these questions were submitted to various forms of data processing, which will be described in the three sections below. The fourth question, referring to neighbourhood networks, was simple and the resulting variable (GETTOG, values from 1 to 3) could be used directly.

- **“Group” variables**. This question went into details, to permit further analysis on group membership at a later stage. For this first stage, it was felt that it was important to capture the kinds of groups that respondents chose to belong to; however, it was desirable also to capture the level of the individual’s participation. It was therefore decided to generate dummy variables for membership in each category of group, and then transform these into aggregate “social capital values”. There are 8 possible categories: church or religious group (60% of all the groups mentioned); political party (2%); burial society (12%); other type of stokvel (7%); other economic group (defined as benefiting its members, e.g. cooperative) (1%); community group (defined as benefiting the community as a whole, e.g. school committee) (1%); cultural/ sports group (2%); other (<0.5%, therefore discarded).

Thus seven variables were created (VCHUR, VPOLP, VBURS, VSTOK, VECON, VCOMM and VCULT); the value of each was created as follows:

1. a basic value of 1 was first assigned if the respondent was a member of such a group; it was doubled if the respondent had mentioned it first out of two or three groups;
2. to this base of 1 or 2 was added the log of the time (categorical variable ranging from 1 to 4) spent weekly by the respondent on

the activities of that group: thus a value of up to 1.61 (corresponding to 4 hours or more a week) was added to the base value if the respondent spent more than 1 hour a week in this group. (The log value of 1 hour a week is 0.) It was decided to use logs, in order to compress the figures and to reflect the diminishing returns to scale of additional time spent on a group.

This manipulation captures, in an elementary way, the level of the individual's participation in that group. The resulting variables can take on 11 possible values: 0 (if the respondent is not a member of such a group), 1 (if the group is mentioned alone, or second or third, and no more than one hour a week is spent on it), or 9 possible variables at unequal intervals to a maximum of 3.61 (if the group is the first to be mentioned, and the respondent spends more than 4 hours a week on it.). This discontinuity poses no problems for factor analysis, and gives better results than two simpler scorings that were tried.

Thus, the group membership aspect of social capital is thus captured by *seven variables* ranging from 0 to 3.61, with names starting with V (VCHUR, VSTOK etc.). As a point of interest, the variables were factor-analysed by themselves, but with weak results. The only pattern worth noting, but not enough to justify using the derived factors, is a correlation between church members and burial society members.

- **“Help” variables.** (This was the name given to the variables relating to informal social safety-nets, since the question was phrased with the word “help” as its key.)

This question was asked in order to assess the importance of social networks as a risk-sharing mechanism. Respondents were asked: “If the main source of income of the household were to fail, where would you find help?” They could choose up to three of the following: (1) extended family; (2) friends; (3) one of the respondent's groups; (4) social welfare, including unemployment insurance; (5) credit from banks or moneylenders; (6) dispersion into other households (implying reliance on extended family, since households in Soweto are family-based); (7) informal employment; (8) existing savings; (9) other (specify what); (10) don't know.

Answers were coded into ten dummy variables (HELPFAM, HELPFR, HELPGR, HELPSOCW, HELPCRED, HELPDISP, HELPINF, HELPSAV,

HELPOTH, NOHELP. “Don’t know” implies the absence of a source of help, so it was called NOHELP.)

Since choosing three out of ten presented possibilities for patterns of commonalities, an exploratory factor analysis was done on these ten “help” variables, with the aim of reducing them to a smaller number.²⁸ The PF analysis results in as many factors as variables, with five having eigenvalues >0 (none >1); the scree test is doubtful. On balance, it was felt that the results are not strong enough to warrant using them in the general factor analysis of social capital.²⁹ Instead, it was decided to maintain the original ten variables.

- **“Job-seeking” variables.** This question was designed to assess the respondent’s social capital from the point of view of its information-carrying role. As mentioned earlier, finding employment is perhaps the most fundamental economic act in an urbanised setting such as Soweto, and one for which the individual’s access to social capital is highly relevant, as indicated by Wittenberg (1999).

The question was: (a) “If you were looking for a job today, what do you think would be your best chance of finding a job?”; and (b) “How did you get your present/last job?” Both parts of the question could be answered with one out of eight choices: (1) mass media, (2) school or employment centre, (3) family or friends, (4) schoolmates or colleagues, (5) one of your groups, (6) self-employment, (7) going door-to-door, (8) other.

No factor analysis can be done on these data as they stand, since only one choice was possible per respondent, thus no commonalities can be examined for patterns. Moreover, this structure gives problems even for a joint factor analysis with the other social capital variables, since it dilutes the commonalities. It was decided to extract one aggregate variable: by combining those who mention family/ friends, schoolmates/ colleagues, and groups. This variable represents those who use informal social networks for job-seeking. (It excludes those who mention mass media, employment centres, self-employment or going door-to-door.) The resulting dummy variable is called JOBSN; it should be emphasised that it does not include those who

²⁸Factor analysis can be done on dummy variables, if the sample is larger than a few hundred - see Kline, p.126.

²⁹Factor analysis was tried both with these factors and with the individual variables, as well as with some simple aggregations of the variables; factor eigenvalues were highest when using the individual variables, and lowest when using these factors.

use institutionalised information sources (mass media, employment centres), although the latter could also be defined as forms of social capital.

- **Neighbourhood networks.** There was one question on this aspect: “Do people in this neighbourhood get together to solve some problems (clean up street, joint security, buy in bulk, etc.?” Three choices were given: often, occasionally, never. The resulting variable is labelled GETTOG. It is important to note that it is a negative value, i.e. the higher it is, the less neighbourhood cooperation it indicates (“Yes” was coded 1, “never” was coded 3).

3.3 Trust variables

As mentioned in the first part of this paper, trust has been recognized as a key element of social capital from the beginning: it is mentioned by Coleman, but was studied in particular by F. Fukuyama. He emphasises the sequence: values → trust → social organisation → economic well-being, although he recognizes also the feedback influence of social organisation on trust and of cultural values. He highlights both family-based trust and wider community-based trust, including trust in the nation-State, in different cultures.³⁰

In this context, different forms of trust can be seen as the *outcome* of different kinds of values, and as the *source* of different kinds of social networks and associational activity. On the other hand, one can also hypothesize trust as the outcome of all other categories of social capital: associational activity could well be seen as a trust-building activity, and trust as the direct transaction-cost-reducing item in the production function. In order to do justice to these hypotheses, it is appropriate not to aggregate measures of trust *a priori* with other values, but let them form separate variables, that can be examined for their relationship with those relating to values and social organisation.

In addition, for the purposes of this study, trust needs to be broken down into different subcategories, in order to bring out possible differences in the patterns of association with other categories of social capital.

In this survey, respondents were asked about their level of trust in various individuals and institutions. Following Fukuyama, it was hypothesised that trust can be divided into (a) family-based trust, (b) community-based trust,

³⁰Fukuyama (1995).

and (c) trust in institutions. Taking this as a basis, respondents were asked about their level of trust in the following:

1. extended family;
2. neighbours;
3. your local school;
4. members of your groups;
5. your professional/working contacts;
6. local police station;
7. local government/community leaders;
8. central/provincial government;
9. people who earn their living the same way as you;
10. people in the same income category as you.

Respondents could rate their level of trust from 1 (lowest) to 5 (highest). This gave 10 trust variables, each with values ranging from 1 to 5. The mean of all answers was 3.18, indicating relatively little positive bias (i.e. tendency to give positive answers). On the other hand, there was quite a marked internal positive correlation between the responses, i.e. higher trust in any one direction is correlated with higher trust in other directions, thus indicating either a strong influence of individual character (i.e. some individuals are generally more trusting/less trusting than others) or a strong influence of some outside characteristic on an individual's general level of trusting. (It might be worth noting that the highest mean level of trust was expressed in one's groups (3.9) and family (3.7); the lowest level of trust was towards neighbours (2.7) and those in the same income category (2.6).³¹)

This group of data is particularly well-suited for factor analysis, since the variables have a good level of differentiation (1 to 5) and there are clear patterns of commonality between them. Thus, a preliminary factor analysis was done on these ten variables.

³¹These data are set out in more detail in Piazza-Georgi (2001a).

Two of the variables - trust in those who earn their living the same way as the respondent, and trust in those whose income level is similar to the respondent's - presented such a strong correlation that it was realised that they constituted a bloated variable,³² thus they were first combined into one factor, *representing trust in those of similar economic status*. This left a total of nine variables. They were factor-analysed separately, with good results, namely three well-identifiable factors with acceptable eigenvalues:³³ however,

³²A bloated variable is a group of variables (always very strongly correlated) that measure practically the same thing, thus the commonality between them is meaningless: e.g. the number of household members and the number of beds in the house.

³³The three factors are as follows:

Factor 1: “Trust in formal institutions” (Eigenvalue=2.04)

Variable Loading

Trust in local government .78

Trust in central government .75

Trust in the police .60

Trust in local school .38.

This appears as quite a strong factor, which brings together the variables that had been hypothesised as representing *trust in institutions* (see previous page), and adding a secondary element of trust in schools, justifying the above hypothesis that this can be an element of institutional trust.

Factor 2: “Trust in specialised social links” (Eigenvalue=.71)

Variable Loading

Trust in those of same economic status .44

Trust in working contacts .33

Trust in co-group members .32

Trust in police -.23.

This factor brings together trust in those with whom one is linked in the economic sphere and in the sphere of voluntary association, with a negative loading on trust in some forms of local authority. Going back to the data, it was found that this trust factor is correlated almost exclusively with *burial society membership*, not membership of stokvels, cooperatives or any other kinds of associations. This factor recalls both Woolcock's micro-level “linkage” and Fedderke et al.'s “high-rat” social capital, as it seems to denote both trust in non-traditional community links and trust in “specialised”, goal-oriented social links.

Factor 3: “Community-based trust” (Eigenvalue=.28)

Variable Loading

Trust in neighbours .49

Trust in extended family .44

Trust in local school .40

Trust in co-group members .23

This indicator seems to combine Fukuyama's family trust with *local community-based trust*; perhaps one could equate it better with “bonding” trust, or Fedderke et al.'s “low-

it was established that if the main analysis was run with the nine individual trust variables, the same three trust factors reappeared robustly (in first, third and sixth place). Thus it was not felt necessary to use the result of the separate factor analysis of the trust variables.

4 Interpretation of the factor analysis.

4.1 First-level principal factors analysis

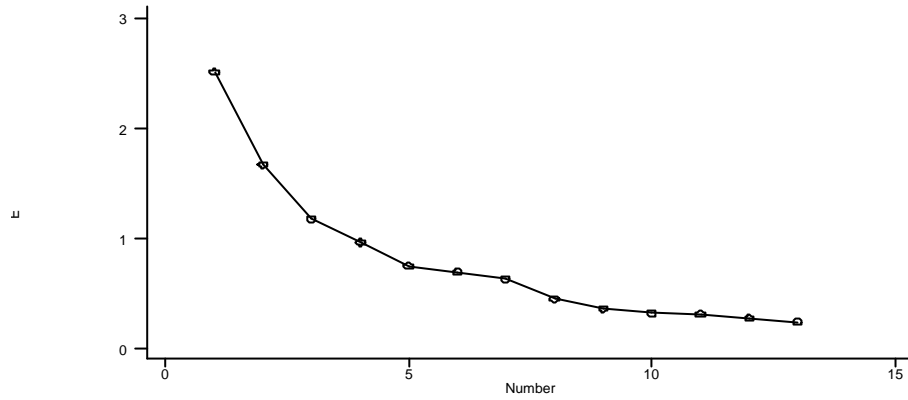
The variables obtained, or retained, as described in the sections above, were then run through the main factor analysis together. To repeat the research question (see Part I) once again: can the numerous information items relating to social capital be meaningfully fitted into the low-rat/high rat categorization? Or are there other meaningful categories that define the different forms of social capital in the context of present-day Soweto?

The printout of the analysis is attached. The scree test shows the clearest changes of slope at 5 and 7 factors. It was decided to interpret seven factors. Promax rotations were used, on the assumption that the social capital factors to be identified would not necessarily be orthogonal (uncorrelated) to each other.

In the lists below, negative loadings are emphasised by listing the variables *in italics*; the asterisks indicate that that particular loading is the strongest, either negatively or positively, of all the loadings on that variable. Loadings below $|.20|$ have been listed *only* if they can be asterisked.

rat" social capital. Note that this is weaker than the other two factors.

Figure 1: Scree test of principal factor analysis of social capital variables



4.1.1 Interpretation of seven-factor PROMAX rotation

Factor 1: Trust in formal institutions (eigenvalue: 2.50)³⁴

Trust in local government	.78*
Trust in government	.78*
Trust in the police	.60*
Trust in local school	.28
Trust in professional/working contacts	.27
<i>Trust in those of same economic status</i>	-.21*
<i>Individualistic interpretation of "Ubuntu"</i>	-.10*
<i>Marginal value of leisure</i>	-.07*
<i>Attitude to wealth</i>	-.06*
<i>Attitude to risk</i>	-.06*

This factor is almost completely a reflection of the first trust factor, trust in formal institutions (see footnote 33). It appears that this pattern, the strongest among the trust factors, is in fact the strongest of *all* the social

³⁴The eigenvalues are derived from PF analysis, thus lower than those that would result from PC analysis.

capital variables. (The presence of some additional trust variables, with weaker loadings, can be explained by the internal positive correlation of the trust variables, which has already been remarked upon; one should note, however, the negative loading on TRIEFA.) What is interesting is the additional information, on the other social capital variables with which this factor is associated, albeit weakly: an understanding of the term “Ubuntu” that tends towards the individualistic, a willingness to do additional work for relatively low wages, and last but not least, a negative attitude to rich people and to risk-taking.

This factor is conceptually linked with Woolcock’s “linkage” dimension. It is also worth noting that, on correlating this factor with demographic variables, no link can be seen with education or income, but there is a slight positive correlation with age.

Factor 2: High-material-means (eigenvalue 1.66)

Safety-net based on savings	.65*
Safety-net based on credit sources	.55*
Attitude to wealth	.34*
<i>Trust in one’s groups</i>	-.33*
<i>Trust in one’s family</i>	-.30
Membership in cultural/sports societies	.30
<i>Network-based job-search strategy</i>	-.29*
Individualistic values	.27*
Authoritarian views on the meaning of “ubuntu”	.26*
Stokvel membership	.20*
Marginal value of leisure	.21
<i>Burial society membership</i>	-.17*
<i>Safety-net based on friends</i>	-.16*
<i>Attitude to community values</i>	-.16*
<i>Neighbourhood solidarity</i>	-.16*
<i>Safety-net based on family</i>	-.10*

This factor reflects differences between the wealthier and the less wealthy in Soweto. It is highly correlated with high household income ($R = .46$), and consequently with education (.29). It centers around a safety net based on financial resources (either savings or credit). Some other loadings are to be expected: positive attitudes towards wealth and individualistic values (but note the authoritarian interpretations of the term “ubuntu”); and the high marginal value given to one’s time. One notices also a tendency to stokvel membership, away from the traditional burial societies; and for affiliating

with “modern” cultural and sports groups.

The fact that job-search strategies tend not to be network-based may not be an indicator of a lack of networking, but may simply reflect the different nature of the job-market for persons of higher education and means. However, a number of other points indicate a lack of certain kinds of social capital: negative loadings on family and group trust, on neighbourhood solidarity and on community values; and on friends and family-based safety-nets. Interesting, on the other hand, is the positive loading on neighbourhood solidarity (as mentioned earlier in this paper, that value is inversed - a lower number means a higher value: since there is a strong correlation between this factor and living in the wealthier, private-housing areas of Soweto, this loading probably indicates neighbourhood cooperation in higher-income areas for security - similarly to what happens in the wealthier “white” areas of Johannesburg

It seems clear from the above that material means are linked with a decrease of social capital, particularly the traditional “bonding” forms (family, community values). The “high-rat”, “bridging” forms of social capital are more strongly present. Note also that this factor has the second-strongest correlation ($R = .13$) with being an entrepreneur.

Factor 3: “Bridging” social capital, trust in high-rat links (eigenvalue 1.17)

Attitude to community values	.43*
Trust in persons of same economic status	.42*
Trust on one’s groups	.42*
Trust in professional/working contacts	.41*
Neighbourhood solidarity	.30*
<i>Household dependent on remittances</i>	-.26*
<i>Church membership</i>	-.24*
Safety-net based on social welfare	.23*
Membership in economic groups	.17*
<i>Trust in the police</i>	-.15*
<i>Individualistic values</i>	-.13*
<i>Safety-net based on credit sources</i>	-.10*

This factor is anchored in high value given to community values and neighbourhood solidarity (note also negative INDIVVAL), combined with the second trust factor (see footnote 33), i.e. trust in non-institutional but “modern” social links (persons of same economic status, groups, professional contacts, negative trust in police). Safety-nets tend to be based on social welfare. Note the financially independent households, and weak group mem-

bership overall.

In contrast to the previous factor, this seems to be a factor that reflects conscious social capital, as shown by the strong positive loading on community values; but note the strong negative loading on neighbourhood solidarity, which must be seen in the context of the fact that, unlike the previous factor, this one has no correlation with living in rich areas.

One notes that this factor is linked with younger and more educated persons ($R = -.11$ and $.12$ respectively), although the previous factor partialled out the bulk of the wealth effect. This factor seems to fit in well with hypotheses about “bridging” forms of social capital, and particularly with Fedderke et al.’s rationalisation dimension, as will be discussed in the conclusions.

Factor 4: safety-nets (eigenvalue: .96)

<i>No place to turn for help</i>	-72*
Family-based safety-net	.52*
Reliance on self-employment as a safety-net	.32*
Marginal value of leisure	.25*
Attitude to wealth	.22
Friends are a safety-net	.21*
Safety-net based on joining other households	.13*
<i>Trust in professional/work contacts</i>	-.10*
<i>Trust in neighbours</i>	-.06*

This factor is centered so strongly around the different forms of safety-net that it can be assumed to reflect simply the correlation between the safety-net variables, caused by survey design. The absence of a place to turn for help is of course negatively linked to all the other possible sources of help. It is interesting, however, to note the link between having positive sources of help in case of need, a high marginal value of one’s time, and a positive attitude to wealth; but accompanied by a negative loading on some forms of trust.

Factor 5: entrepreneurship (eigenvalue: .74)

Groups are a safety-net	.42*
Authoritarian views on “ubuntu”	-.29*
Other forms of safety-net	-.25*
Network-based job-seeking strategy	-.25
Marginal value of leisure	.22
Safety-net based on social welfare	.21
Attitude to risk	.21*
Safety-net based on joining another household	-.15*
Membership of cultural/sport group	-.20
Individualistic interpretation of “Ubuntu”	.19
Political party membership	.19*
Neighbourhood solidarity	-.18*

This factor has been labelled “entrepreneurship” because it has the highest correlation (.16, admittedly not a high value) with self-employed respondents; there is also a .12 correlation with household income. It has the highest loading on positive attitudes to risk-taking, and the strongest individualistic and anti-authoritarian slant on the respondent’s interpretation of the concept “ubuntu”. The other components are not the *a priori* expected ones: in particular, the marginal value of leisure tends to be high rather than low, and there is no reliance on self-employment as a safety-net.

The strongest loading is on the tendency to rely on one’s groups or on social welfare as safety net; this factor tends to be related with membership in a political party, in an economic group, a burial society or a stokvel, but not a church or community group. It is also noteworthy that there is no loading, positive or negative, on any form of trust. As a whole one notices that the loadings indicate a strong tendency to “bridging” forms of social capital rather than the more traditional “bonding” types.

Factor 6: Kinship-based social capital (eigenvalue: .69)

Trust in neighbours	.50*
Trust in family	.49*
Trust in school	.33*
Safety-net based on family	.25
<i>Safety-net based on social welfare</i>	-.24*
Individualistic views on meaning of “ubuntu”	.21*
Member of an economic group	-.17*
<i>Member of a stokvel</i>	-.14*
<i>Member of a political party</i>	-.12*
<i>Safety-net based on one’s groups</i>	-.11*
<i>Trust in central government</i>	-.07*

This factor is clearly linked to traditional, kinship-based social capital. It incorporates the third trust factor (see footnote 33), “low-rat trust”, a pattern reinforced by the other loadings: safety-net based on family rather than social welfare, and a tendency not to belong to “modern” groups such as a political party, a stokvel or an economic groups. Note, however, that this factor has the strongest positive loading on individualistic interpretations of “ubuntu”. Nevertheless, it clearly represents “bonding”, low-rat social capital.

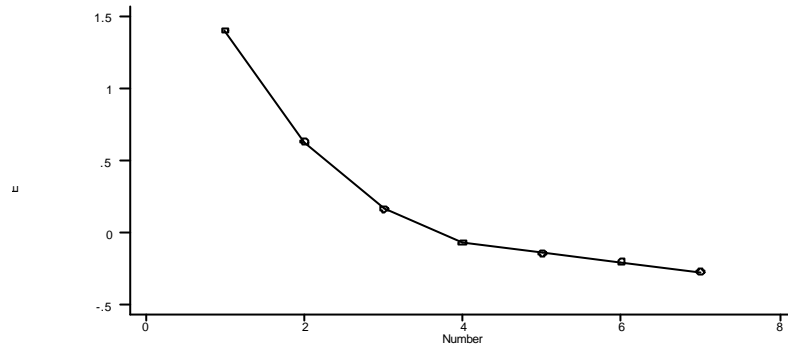
There is a positive correlation (.12) with age, and linked to that, a negative one (-.14) with education; this seems consistent with this kind of social capital.

Factor 7: dependency (eigenvalue: .64)

<i>Reliance on self-employment as a safety-net</i>	-.68*
No place to turn for help	.30*
Church membership	.24*
Attitude to wealth	.22
Burial society membership	.19*
Household depends on remittances	-.13*
<i>Trust in local government</i>	-.10*
Other sources of help	.10*

The last factor is not easy to interpret. Noting that the demographic profile is similar to the previous factor (higher age, lower education), but that this factor lacks any form of trust, one might guess that this is a form of negative, i.e. weak social capital. It is labelled “dependency” because of the first three components, which suggest a sense of fatalism, and the maximal loading on dependency on remittances. Burial society membership is

Figure 2: Scree test of PF analysis of the seven promax-rotated factors from previous analysis



generally positively correlated with church membership, and this is reflected in this factor.

4.2 Second-level principal factor analysis

Since the factors were rotated by the promax method, they are not necessarily orthogonal to each other; thus, there may be relationships between them which it is worth analysing further. This can be done by *subjecting the factors themselves to a further round of factor analysis*, as done by Sudarsky (1999) and Onyx and Bullen (2000). Such an analysis reveals the commonalities between the factors themselves, showing up those that tend to be associated with one another in a respondent, and those that tend to be mutually exclusive. In this analysis, three second-level patterns³⁵ are found with eigenvalues >0 , as shown in the scree test below.

The three patterns appear as follows:

Pattern 1: traditional social capital (eigenvalue 1.41)

Factor 1	kinship-based social capital	.62
Factor 2	trust informal institutions	.52
Factor 3	<i>high material means</i>	-.52

³⁵The term “patterns” is used in this section for these second-level factors, to distinguish them from the seven primary factors.

This pattern confirms the hypothesis of the substitution effect between the kinship-based, “bonding” kind of social capital, and the existence of material means. It is interesting that it is strongly associated with trust in formal institutions: in Woolcock’s terminology, this pattern appears to link micro-level integration with micro-macro linkage, leaving out micro-level synergy (bridging social capital); and this form of social capital is negatively linked to high material resources.

Pattern 2: safety-nets and dependency (eigenvalue.63)

Factor 7 dependency .65

Factor 4 safety-nets .61

This pattern is more difficult to interpret, as it draws together the two factors that had themselves posed an interpretation problem. It contrasts with the first and third patterns, which show a clear “profile” even though they include some surprises; further analysis may be needed to obtain more clarity on it. One can already note, however, that the pattern gives some sense of a lack of proactivity (membership only in churches and burial societies, high marginal value of leisure, general reliance on others); and it is interesting that this trait should go together with a general lack of trust (no positive loadings, some negative).

Pattern 3: “bridging” and entrepreneurship (eigenvalue.17)

Factor 3 “bridging” and hi-rat social capital .29

Factor 5 entrepreneurship .27

The surprise here is that this pattern should be the weakest, since it draws together what one might suppose to be a distinct and identifiable form of social capital, the “modernized” form. Note again that factor 3 pulled together all the hi-rat trust factors, while factor 5 pulled together other forms of “modernized” social capital but no forms of trust.

5 Conclusions

Most factors and patterns are relatively weak as regards loadings, but that is to be expected from the subject itself, and due to the less than ideal structure of the original data.

The data indicates that, in the context of present-day Soweto at least, social capital can be expressed in about seven primary factors, that combine into three distinct patterns. The primary factors have two robust characteristics, which reappear regardless of the different ways of defining, analysing

and rotating the data:

1. a distinction between trust in formal institutions and other forms of trust; and
2. a distinct form of social capital correlated with high-income households and high-education respondents.

Other characteristics are less robust, and the final choice in the definition of the variables and in the rotation of the factors, was determined by the objective of bringing these out as clearly and intelligibly as possible.

The objective of the paper, as set out in the first section, was to examine the patterns of the data and attempt to fit them into meaningful categories - meaningful in terms of (a) existing theory, paying particular attention to functional categories, i.e. the simple bonding/bridging function, Woolcock's dimensions, and Fedderke et al.'s "low-rat" and "high-rat" categories; (b) the specific social context being surveyed, or (c) new widely-applicable insights.

Some of the main points that have arisen from the analysis will now be repeated and re-focused, in order to bring out their significance in the context of the objective of the study.

- the distinction between trust in formal institutions and other forms of trust, runs counter to some recent thinking.³⁶ The separate factor that thus arises, reflects, on the other hand, Woolcock's "synergy" dimension, i.e. the embeddedness of State-society relationships.
- The second factor, that distinguishes high-income, high-education persons, is not easily matched to any existing social capital theory. It could point towards the main hypothesis of the study of which this paper is a part: the possible substitution effect between certain aspects of human and social capital. It was noted, for example, that this factor seems negatively correlated with traditional kinship-based forms of social capital but neutral towards other forms. It is also interesting that these persons, who tend to live in more expensive areas, show the strongest positive loading on neighbourhood cooperation, which is contrary to the assumption of neighbourliness as a "popular" value.

³⁶Cf. first section: reference to Stiglitz's essay in Dasgupta and Serageldin (2000).

- The third factor seems to be best explainable in terms of Fedderke et al.'s "rationalisation" dimension. As a reminder, this dimension focuses on the procedural rather than the substantial, and the universally applicable rather than the particular. Insofar as one assumes material well-being (income growth) to be a key objective of social links - as is implied in the very definition "social capital"- social links with a direct economic interest are "high-rat". That is what we find in this factor: trust in persons of the same economic status, professional/working contacts, and in one's groups; a positive loading on membership in economic groups, and negative on church membership; social welfare considered as the primary safety-net.
- The sixth factor, kinship-based social capital, reflects the basic traditional form of social capital for which the concepts of "embeddedness" and "bonding" were coined, and which Woolcock called the micro-level integration dimension. The fact that it is negatively correlated with education (though interestingly, not with household income), is a further pointer towards the main hypothesis of the study of which this paper is a part.
- The fourth, fifth and seventh primary factors, taken by themselves, are not clear enough to use them to contradict or support any previous theory. Particular attention was paid to identify entrepreneurial aspirations. The fifth factor, unclear as it is, merits attention in this regard, since it is distinctly correlated with self-employed persons. More analysis will be done on this point.
- At the second-level analysis, it is interesting that the strongest pattern combines the "trust in formal institutions", that reflects Woolcock's synergy" dimension, and the "traditional" kinship-based factor that can be taken to reflect Woolcock's micro-level embeddedness or "integration" dimension. This may indicate a tendency for embedded social capital at the micro and macro level to reinforce each other, i.e. to be present in the same persons. Whether this pattern is a context-specific one or more universally applicable, remains to be determined by further research.
- Again at the second level, the middle pattern is confusing but potentially very interesting: it may indicate the existence of "passive" social

attitudes and links that have not been picked up so far in theory. It remains to be seen, however, whether this pattern is a robust one and whether it applies more widely, or whether it is a local pattern linked to the difficult past of the majority of the members of this community.

- Lastly, the third pattern, although weak, suggests interesting perspectives for the further study of “high-rat” forms of social capital, drawing together the “high-rat” factor and the more undefined fifth factor, which is correlated with entrepreneurship. It is a surprise that this pattern should be the weakest, since it draws together what one might suppose to be a distinct and identifiable form of social capital, the “modernized” form. Note again that factor 3 pulled together all the hi-rat trust factors, while factor 5 pulled together other forms of “modernized” social capital but no forms of trust.

References

- Coleman, James S. (1988): “Social Capital in the Creation of Human Capital”, *American Journal of Sociology*, 94 (Supplement), S95-S120.
- Coleman, James S. (1990): *Foundations of Social Theory*, Cambridge, Mass. The Belknap Press of Harvard University Press.
- Dasgupta, P. and Serageldin, I., eds. (2000): *Social Capital: a multifaceted perspective*. World Bank, 2000.
- Fedderke, J. (1997): “Political and Social Dimensions of Economic Growth”, *Theoria*, 89, 1-43.
- Fedderke, J. and Klitgaard, R. (1998): “Economic Growth and Social Indicators”, *Economic Development and Cultural Change*, 46(3), 455-490.
- Fedderke, J., De Kadt, R. and Luiz, J. (1999): “Growth and Social Capital: a Critical Reflection”, *Theory and Society* (forthcoming).
- Fukuyama, Francis (1995): *Trust: the Social Virtues and the Creation of Prosperity*, Hamish Hamilton, London.
- Kline, Paul (1994): *An Easy Guide to Factor Analysis*; Routledge, London.

Knack, S. and P. Keefer (1997): "Doea social capital have an economic payoff?", *Quarterly Journal of Economics*, November 1997, p.1251-1288.

Krishna, A. and E. Shrader (1999): "Social Capital Assessment Tool", paper prepared for the Conference on Social Capital and Poverty Reduction, June 1999.

Mbigi, L. with Maree, J., 1995: *Ubuntu: The Spirit of African Transformation Management*. Knowledge Resources (Johannesburg).

Narayan, Deepa and Pritchett, Lant (1997): "Cents and Sociability: Household Income and Social Capital in Rural Tanzania", World Bank Economic Growth Project research paper.

Onyx, J. and Paul Bullen (2000): "Measuring Social Capital in Five Communities", *Journal of Applied Behavioral Science*, vol.36 no.1, March 2000 p.23-42

Paxton, Pamela (1999): "Is social capital declining in the United Sates? A multiple indicator assessment", *American Journal of Sociology*, July 1999, p. 88-127.

Piazza-Georgi, B. (2001a): *Human and Social Capital in Soweto in 1999*. ERSA policy paper no.4, University of the Witwatersrand, Johannesburg.

Piazza-Georgi, B. (2001b): *A profile of the self-employed in Soweto in mid-1999, with gender disaggregation*. ERSA policy paper (forthcoming).

Piazza-Georgi, B. (2001c) : "The role of human and social capital in growth: extending our understanding" - *Cambridge Journal of Economics* (forthcoming).

Putnam, Robert D. et al. (1993): *Making Democracy Work: Civic Traditions in Modern Italy*, Princeton, NJ: Princeton University Press.

Rummel, R.J.(undated) : "Understanding Factor Analysis" (www.mega.nu:8080/ampp/rummel/ufa.htm) - a summary of his *Applied Factor Analysis* (1970) (Evanston, Ill., NUP)

STATSOFT on-line textbook (undated): “Principal Components and Factor Analysis” (www.statsoftinc.com/textbook)

Sudarsky, John (1999): “Colombia’s Social Capital - the National Measurement with BARCAS”. World Bank, www.worldbank.org/poverty/scapital/library/engsudarsky.pdf

Uphoff, N. (1999): “Giving Theoretical and Operational Content to Social Capital”, Cornell University (draft).

Wittenberg, M. (1999): “Job Search and Household Structure in an Era of Mass Unemployment: a semi-parametric analysis of the South African labour market”, ERSA Working Paper no.3, Univ. of the Witwatersrand, South Africa

Woolcock, M. (1998): “Social Capital and Economic Development: Towards a Theoretical Synthesis and Policy Framework”, *Theory and Society*, vol. 27 no. 2, pp. 151-208

Woolcock, M. and Deepa Narayan (2000): “Social Capital: Implications for Development Theory, Research, and Policy”; *The World Bank Research Observer*, vol. 15, no.2 (August 2000), pp.225-49.