Eliciting Poverty Rankings from Urban or Rural Neighbors: Methodology and Empirical Evidence*

Pascaline Dupas† Marcel Fafchamps‡ Deivy Houeix§

This draft: May 14, 2023

Abstract

We introduce a novel approach for eliciting relative poverty rankings that aggregates partial rankings reported independently by multiple neighbors. We first demonstrate that the method works in principle. We then apply it on secondary data from rural Indonesia and to original data from urban Côte d'Ivoire. We find that the aggregation method works well in the rural setting but, in the urban setting, the reconstructed rankings are often incomplete, not always transitive, and they sometimes contain cycles. This disparity suggests that eliciting poverty rankings by aggregating rankings from neighbors may be more difficult in urban settings. We also confirm earlier research showing that poverty rankings elicited from neighbors are correlated with measures of poverty obtained from survey data, albeit not strongly. Our original methodology can be applied to many situations where individuals with incomplete information can only produce a partial ranking of options.

*We thank for excellent comments three anonymous referees and editor Garance Genicot, as well as Christian Ghiglino, Francis Bloch, Gabrielle Demange, Marcin Dziubinski, Rakesh Vohra, Yann Bramoullé, Steven Glazerman, Andrew Crawford, Florian Grosset, Sneha Subramanian, Matthew Olckers, Josh Blumenstock, Gayatri Koolwal, Elliott Collins, Abdul Noury, and the participants to the IPA 2021 Methods and Measurement Conference and to the NYU Abu Dhabi 2023 Conference on the Economics of Networks. We also thank Francis Bloch for inspiring this research. We would like to acknowledge and thank Eva Lestant and Arsène Zongo for excellent field research assistance, and Innovations for Poverty Action (IPA) Côte d'Ivoire for collecting the data. We are grateful to the Stanford King Center on Global Development and to the IPA Research Methods Initiative Grant IPA-M5 for funding this research. Human Subjects Research approval was obtained from the Stanford University IRB (Protocol 42884).

†Stanford University, NBER and CEPR. Address: 579 Jane Stanford Way, Stanford CA 94305. Email: pdupas@stanford.edu

‡Stanford University, NBER, and IZA. Address: 616 Jane Stanford Way, Stanford CA 94305. Email: fafchamp@stanford.edu

§Massachusetts Institute of Technology, Department of Economics. Email: houeix@mit.edu

1 Introduction

Many developmental interventions aim to target the poor (e.g., Ravallion 2000, Ravallion 2009, Ravallion 2015). In some instances, such as when poverty is highly concentrated, community-based or spatial targeting may be sufficient. However, identifying the poor usually requires within-community targeting mechanisms (e.g., Elbers et al. 2007).

Various strategies have been developed to identify the poorest members of communities. In the absence of universal administrative data such as income tax filings, one strategy is to survey individuals or households and rank them on the basis of the information they provide. One famous example is the eligibility assignment of the Progresa Cash Transfer program in Mexico (e.g., Skoufias et al. 1999). In practice, approaches vary in the type of information that is collected: detailed surveys on consumption and income (e.g., Deaton 2019; Grosh and Glewwe 2000); light surveys on poverty indicators (e.g., assets – Elbers and Lanjouw 2003, Elbers and Yin 2007); or answers to subjective well-being questions (e.g., Ravallion and Lokshin 2001, Ravallion and Beegle 2016). These methods all have shortcomings: detailed surveys are expensive and time-consuming; short surveys are thought to be easily manipulable by respondents (e.g., Banerjee et al. 2020); and subjective well-being is often not well correlated with material well-being, either over time or across countries (e.g., Blanchflower and Oswald 2004; Layard 2009). Furthermore, the rankings are affected by measurement error and possible response bias or manipulation, leading to mis-assignment.[1]

Another method is to delegate the targeting decision to the local level. For example, local chiefs in Malawi were tasked with identifying poor households eligible for a large farming input subsidy (e.g., Basurto et al. 2020). However, a key concern with this approach is the potential for local capture or nepotism, as shown in studies such as Alatas et al. 2013. To mitigate this, one can solicit relative rankings from community members themselves – often gathered in a focus group. The focus groups are asked to produce complete relative poverty rankings of a set of individuals or households, typically members of their village or neighborhood (e.g., Alatas et al. 2012). The main advantages of this method are that it is, on the one hand, simpler and cheaper to implement than detailed surveys, and, on the other hand, more transparent than relying on the local elite alone. This approach has been shown to produce reasonable rankings in a rural context (e.g., Trachtman et al. 2022). It has also

[1]The measurement error arises from the fact that respondents have only imperfect knowledge of the answer – e.g., because they do not recall or do not have full information about other household members. This noise leads to errors of assignment – known as type I and type II errors (e.g., Ravallion 2015). Response bias arises when respondents expect a benefit from being assigned to a high or low rank – such as a welfare benefit from being classified as 'below the poverty line'. To the extent that everyone faces the same incentive to bias their survey responses downward or upward, this need not lead to distorted rankings. But it can result in mis-classification of respondents as poor or non-poor (e.g., Ravallion 2008).

been shown to yield valuable information in domains other than poverty rankings, notably entrepreneurial potential (e.g., Hussam et al., 2022) and long-term poverty (e.g., Trachtman et al. 2022). It it, however, vulnerable to local prejudices and views about who is a deserving poor (e.g., Galasso and Ravallion 2005, Ravallion 2008, Alatas et al. 2019). It also assumes that a small number of community members have the necessary information to provide all the requested rankings (e.g., Alatas et al. 2016).

While relying on key informants can produce meaningful rankings in small rural hamlets, it is unclear whether it applies to urban and peri-urban areas with a more mobile population and less dense social networks. One study in an urban setting (Beaman et al. 2021) finds little evidence that individuals can accurately assess whether randomly selected community members are poor. They nonetheless target transfers to the poor modestly better than would be attributable to chance, suggesting that they possess partial but relevant information. If this diffuse information can be combined in a meaningful way, it could be used to derive an aggregate poverty ranking.[2]

In this paper, we propose a novel methodology for aggregating partial rankings and we implement it in two settings: rural Indonesia; and urban Africa. For Indonesia, we rely on data collected by Alatas et al., 2012 in 640 rural communities. For urban Africa, we collect original data in 34 poor neighborhoods of Abidjan, a large metropolis of more than five million inhabitants in Côte d'Ivoire, West Africa. The latter setting is well-suited because poverty measurement is a topical policy issue in the region.[3] In Abidjan, we ask respondents in 34 different neighborhoods to rank up to 14 target households in that neighborhood. Limited overlap between the sets of households ranked by each respondent rules out using the method of Alatas et al., 2012 to construct an aggregate ranking of the target households. We instead develop a methodology to aggregate all the available – but partial – information provided by the respondents.

We then compare the rankings provided by respondents to survey data. For Indonesia and Côte d'Ivoire, target households answered an LSMS-style survey covering incomes, consumption, assets, and household characteristics. A measure of household consumption per capita is then constructed for each target household using this survey data. For Côte

[2]Alatas et al. 2016 note that, even in a rural setting, lack of information leads to partial rankings because respondents are unable or unwilling to rank certain individuals.

[3]In particular, in 2019, Côte d'Ivoire started to roll out its universal health care coverage (CMU - Couverture Medicale Universelle) that provides free access to health care to the poorest members of a community. This is a context where, as we show, poverty levels are highly heterogeneous within neighborhoods, which means that geography-based targeting is insufficient. Under the ongoing government scheme, the poor are identified using a combination of observables (PMT) and community assessment with local leaders. Whether leveraging peer rankings can improve the targeting of the program is an unanswered question in this and similar contexts.

d'Ivoire we also construct two summary statistics often used in practice: a Proxy Means Test (PMT) index of poverty based on survey data on assets and durables;[4] and a Poverty Probability Index (PPI) calculated on answers to a survey module proposed by Innovations for Poverty Action (IPA). We then compare the reported and constructed aggregate rankings from peer-to-peer comparisons to the rankings produced by the PMT and PPI indices as well as by various measures of consumption.

We have three main results. The first result is methodological. We demonstrate that it is, in principle, possible to construct the relative rankings of all households in a neighborhood by combining partial rankings provided by a multitude of individual informants. Using Monte Carlo simulations, we identify potential situation when this theoretical possibility may prove elusive. We also develop a simple way of obtaining confidence intervals for individual ranks.

The second set of results is empirical. We first show that our method can work by applying it to the peer rankings data obtained by Alatas et al. 2012. With this rural dataset, our method yields complete rankings without ties in all villages and these rankings are nearly identical to those produced by the ad hoc rank-averaging method used by the authors. In the urban context of Abidjan, however, we find that the constructed rankings fall short of expectations, due to two critical shortcomings: (1) they are incomplete in all cases – sometimes severely so; and (2) they are not always transitive – i.e., many contain cycles. In spite of this, the rankings obtained from our method outperform rankings obtained by a commonly used algorithm (e.g., Zermelo 1929, Bradley and Terry 1952). These empirical findings highlight the limitations of using peer rankings in high-density neighborhoods: most respondents simply do not know many of the households around them. As a result, there is little information to be harnessed from them. This means that the very areas for which geographical targeting is known to be ineffective – i.e., dense urban neighborhoods – are also areas where peer rankings appear to be of little use, even though people live in close proximity to each other. Using a higher number of observers should in principle yield more precise and complete rankings, albeit at a higher cost.

Our third set of results is that rankings elicited from neighbors are not highly predictive of consumption measures collected from survey data. This applies both to the pairwise rankings reported by individual respondents as well as to the constructed aggregate rankings obtained by combining individual answers. Rankings from the rural Indonesia data are nonetheless better than those from Abidjan at predicting consumption rankings based on individual survey data – although with a relatively low R^2, as already noted by Alatas et al. 2012. We also find that the PMT, PPI, and consumption measures from the survey are only moderately correlated with each other, suggesting the presence of measurement errors in those measures

[4]We apply the same weights as those that are used in the study country to construct the PMT index.

as well. But these observational measures are all more predictive of each other than rankings are of them. For Abidjan, we also investigate whether reported rankings correlate better with the conspicuous consumption expenditures of the target households. They do not.

These results help provide some sense of when and how the method can yield useful information. The individual informants in the Abidjan empirical application were asked to rank 14 households in neighborhoods that often contain more than 200. In a large number of cases, informants did not know the target households and, as a result, the fraction of reported rankings falls far below the number of rankings needed to reliably construct aggregate rankings for each neighborhood. This result stands in contrast with the good performance of our method when it is applied to the rural data of Alatas et al. 2012 where reported rankings are more complete and consistent across observers. This suggests that a successful implementation of our method requires a sufficiently large ratio of informants to target households, and a sufficiently limited geographical area from which the target households and informants are selected.

The particularly low correlation between reported rankings and observational data in Abidjan also suggests that, relative to rural Indonesia, urban and peri-urban areas may experience too much income variation and spatial mobility to allow neighbors to accurately guess each other's relative economic standing. The fact that reported rankings are not even correlated with conspicuous consumption makes us suspect that urban and peri-urban households of Abidjan do not, in general, pay much attention to each other – or at least that they do not compare themselves to others in their neighborhood. This interpretation is consistent with the work of Fafchamps and Shilpi 2008 who find that rural respondents in Nepal gauge their subjective well-being relative to their neighbors while residents of Nepalese towns and cities do not. This could explain why it is possible to ask rural dwellers to rank each other, but not urban residents.

We also investigate the extent to which households bias their rankings when asked to self-rank, a question that has been the focus of recent theoretical work (e.g., Bloch and Olckers 2021a, Bloch and Olckers 2021b). To this effect, a randomly selected half of the Abidjan respondents are asked to rank themselves among the 14 target households; the other half are only asked to rank the targets. We find that self-ranking respondents tend to give themselves a rank that is significatively higher than what others give them. Self-ranking is thus a potential source of bias.

While ultimately ineffective in our Abidjan setting, the novel methodology we propose in this paper applies to other situations in which individuals have specific information that allows them to produce a partial ranking of options. Examples includes: farmers experimenting with new crops and techniques; workers observing co-workers; and consumers trying new products. In all these cases, individual economic agents have specific information that enables them to correctly rank some of the available options, but not all. One solution to this aggregation problem is to take the average of the ranks given by different observers. This approach, however, does not give all options equal weights, since the sets of ranked options differ across observers in ways that are not random. Averaging ranks is therefore likely to produce biases whenever there is insufficient overlap in ranked sets across observers. It also penalizes options only known by a few agents, such as those that have only been newly introduced. This in turns generates inertia and discourages innovation. Our method overcomes these problems.[5]

The remainder of the paper is organized as follows. Section 2 demonstrates the main methodological contribution of this paper. Section 4 explains the experimental design and data collection. Section 5 describes the empirical rankings obtained and the subsequent directed graph of relative rankings. Section 6 investigates whether rankings are informative. Section 7 examines the self-rank randomized treatment. Section 8 looks at characteristics that predict the propensity to rank others.

2 Methodology

Our objective is to construct an aggregate ordinal ranking from a multitude of partial ordinal rankings obtained from respondents. Formally, consider a set S of n individuals ranked in the *order* of their income:

$$y_1 < y_2 < ... < y_n$$

where, for this presentation, we assume that all inequalities are strict. This true ranking can be represented as an $n \times n$ matrix R with $r_{ij} = 1$ if $y_i < y_j$. For instance, for $y_1 < y_2 < y_3 < y_4$, matrix R is:

$$R \equiv [r_{ij}] = \begin{bmatrix} . & 1 & 1 & 1 \\ 0 & . & 1 & 1 \\ 0 & 0 & . & 1 \\ 0 & 0 & 0 & . \end{bmatrix} \tag{1}$$

where r_{ii} is defined as missing (i.e., $r_{ij} = .$). The total rank t_i of individual i is simply the sum of its row +1:

$$t_i = \sum_j r_{ij} + 1$$

[5]Truthful reporting is necessary for our method to yield correct rankings, but this is also true of elicitation approaches – i.e., stars or likes. The method does *not*, however, apply to situations in which observers are known to have different preferences or rankings and the purpose of aggregation is to average these preferences.

where the sum is taken over non-missing values. The richest person, individual 4, has rank 1; the second richest, individual 3, has rank 2, and so on. The poorest individual is the one with the largest number of individuals richer than he/she is, i.e., individual 1 in this case.

Missing information is easily accommodated in this matrix representation. Say $y_1 < y_2 < y_3 < y_4$ but respondent a does not know y_3. We have:

$$R_a = \begin{bmatrix} . & 1 & . & 1 \\ 0 & . & . & 1 \\ . & . & . & . \\ 0 & 0 & . & . \end{bmatrix} \tag{2}$$

from which we see that a missing person shows up in R_a as a missing row and column. In this case, the rank of individual 3 is unknown from respondent a. The matrix representation can also accommodate disconnected rankings, e.g., let respondent b report $y_1 < y_2$ and $y_3 < y_4$. We then have:

$$R_b = \begin{bmatrix} . & 1 & . & . \\ 0 & . & . & . \\ . & . & . & 1 \\ . & . & 0 & . \end{bmatrix} \tag{3}$$

Disconnected rankings imply that R_b is a block diagonal matrix.

Equipped with this notation, we can combine rankings across respondents to obtain the union of the individual rankings, and then use this union to construct an aggregate ranking. Formally, let us posit that the number of respondents is $m \leq n$ and that each respondent i observes the incomes of k_i individuals in set S. The rankings reported by the respondent k are represented by ranking matrix R_k. For now, we assume that respondents only report true rankings and report them truthfully without error. Given this assumption, their rankings on a particular ij pair always agree. We show how the pairwise rankings of the individuals in set S can potentially be recovered using the union of the ranking matrices, which we denote:

$$\overline{R} \equiv \cup_k (R_k) \tag{4}$$

To illustrate the intuition behind the method, suppose one respondent reports that $y_1 < y_2$, another that $y_2 < y_3$ and a third that $y_3 < y_4$. The union of the respective ranking matrices yields:

$$\overline{R} = \begin{bmatrix} . & 1 & . & . \\ 0 & . & 1 & . \\ . & 0 & . & 1 \\ . & . & 0 & . \end{bmatrix} \tag{5}$$

from which we immediately see that $\overline{R} = [\overline{r}_{ij}]$ does not contain all the relevant information: by combining the three reports, we have $y_1 < y_2 < y_3 < y_4$, and yet $\overline{R}$ in equation (5) does not look like R in (1). To recover matrix R from $\overline{R}$ we use results from network analysis. By replacing each missing value by 0, ranking matrix $\overline{R}$ can be turned into the adjacency matrix of a directed network where a link from i to j means that i 'looks up to' j, i.e., has lower income than j. Let this adjacency matrix be denoted:

$$A \equiv \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{6}$$

Asking whether there is a sequence of inequalities such that $y_1 < y_4$ is equivalent to asking whether there is a directed path from 1 to 4 in the directed network represented by A. It is well known that all paths in a network can be recovered by taking n successive (integer) powers of the adjacency matrix. Paths of length 1 are given by the non-zero elements in matrix A itself. A path of length $1 < l \leq n$ exists between nodes i and j if element a_{ij} in A^l is greater than 0. The shortest path between i and j is the smallest positive integer l at which $a_{ij} > 0$ in A^l. Let us now define $\hat{r}_{ij} = 1$ if there is a directed path, of any length, between i and j, and let us define the matrix:

$$\hat{R} \equiv [\hat{r}_{ij}] \tag{7}$$

Applying this method to matrix A in equation (6) yields $\hat{R} = R$ in equation (1). This demonstrates that, in this particular example, it is possible to recover the aggregate ordinal ranking from the union of ranking matrices $\overline{R}$.

The above intuition can be generalized as follows:

Definition An adjacency matrix $\overline{R}$ is said to be fully rankable if it is upper-triangular and has 1's all along the upper off-diagonal. Examples: Matrix A in expression 6 is fully rankable. Matrices R_a and R_b in expressions 2 and 3 are upper-diagonal but not fully rankable.

Lemma If there are (A1) there is no reporting error, (A2) each respondent k truthfully reports his or her known ordinal rankings R_k, (A3) there are no ties in the true ranks, and (A4) each ij pair of nodes is ranked by at least one observer k, then the adjacency matrix $\overline{R}$ constructed by union of the ranking matrices R_k over all k is fully rankable.

Proof. [A] Show that if there is no misreporting and no ties in true rankings, matrix R must be upper-triangular. This is because, by assumptions (A1) and (A2), matrix $\overline{R}$ can only contain true rankings and, in the absence of ties (A3), true rankings form an upper-triangular adjacency matrix. [B] Show that if each pair is ranked by at least one observer, then $\overline{R}$ is fully rankable: A contrario, suppose pair ij is not ranked by any observer. This means that elements ij and ji must be 0 in all R_k matrices. It follows that elements ij and ji of matrix $\overline{R}$ must also be 0 since $\overline{R} \equiv \cup_k (R_k)$.

Proposition 1 Provided that (A1) there is no reporting error, (A2) each respondent k truthfully reports his or her known ordinal rankings R_k, and (A4) the adjacency matrix A is fully rankable, then the true ordinal ranking R can be recovered from the union of all reported rankings R_k using the method described above.

Proof. If assumptions A1 and A2 are satisfied, the ranks reported in the upper-triangle of matrix $\overline{R}$ must correspond to true ranks. Since $\overline{R}$ is fully rankable, there can be no ties. This means that the true rank matrix R made of all pairwise rank comparisons is an upper triangular adjacency matrix with all elements equal to 1 in the upper triangle. Iterating on adjacency matrix $\overline{R}$ as discussed in the text identifies all the paths leading from each of its node. Since in a fully rankable matrix each upper off-diagonal element $r_{i,i+1} = 1$, a simple iterative argument implies that this process yields an upper triangle made only of 1's in at most $n - 1$ steps.

There are many circumstances where the above approach cannot identify the complete ranking. This arises, for instance, when $\overline{R} = R_a$ or R_b since neither of these matrices are fully rankable. In matrix R_a, there is no information at all on y_3, which means individual 3 can never be ranked relative to the others. In matrix R_b, individuals $\{1, 2\}$ and $\{3, 4\}$ belong to two distinct components[6] and thus cannot be compared to each other. It is also possible that aggregating responses results in a partial ranking. For instance, if respondent a reports $y_1 < y_2$ and respondent b reports $y_1 < y_3$, we do not know whether y_2 is greater or smaller than y_3.[7] In that case, the network has a fork at y_1, with one arrow pointing to

[6]A component of a network is a connected sub-network that is not connected to any node outside of it.

[7]It is even possible that aggregating responses results in non-transitive rankings, i.e., cycles in the ranking

y_2 and another pointing to y_3. It is also possible to have two arrows pointing to the same node, e.g., if $y_1 < y_3$ and $y_2 < y_3$. In all these cases, matrix $\overline{R}$ is not fully rankable since we lack information on certain ij pairs. In spite of this, partial rankings can nonetheless be recovered using the iterative method proposed here.

The ranking information recovered from respondents' reports can be represented in two ways. First, we produce graphs of the recovered rankings in each studied location. In order to convey all available information in an intuitive and compact format, we construct, for each studied location, the minimally connected graph $\hat{M}$ that spans $\hat{R}$. This graph is obtained by eliminating as many links from A as possible while ensuring that $\hat{M}$ still produces the same pairwise rankings $\hat{R}$ as A. For example, for the case where $y_1 < y_2 < y_3 < y_4$, $\hat{M} = A$ in equation (6): we only need three directed links to define all the relative rankings, we can drop the $\{1, 3\}$ link, the $\{1, 4\}$ link, etc, without changing the information about relative rankings that are presented in the graph.

Second, we also compute summary statistics that convey approximate information about the relative ranking of each individual in a compact fashion. These are r_i^{up}, the number of individuals who rank higher than i, and r_i^{down}, the number of individuals who ranked lower than i.[8] If we have complete ranking information on n individuals, then $r_i^{down} = n - r_i^{up} - 1$. When ranking information is incomplete, that is, when adjacency matrix A is not fully rankable, this is typically not the case. For individuals who are unranked, for instance, we have $r_i^{down} = r_i^{up} = 0$. The difference $P_i \equiv r_i^{down} - r_i^{up}$ nonetheless remains informative about the *relative position* of individual i in the constructed network since it counts how many people can be ranked as poorer than i and subtracts how many can be ranked richer.[9] These rankings are most instructive for individuals located within the same component but not across components since, by definition, individuals from different components cannot be compared to each other. Even in such cases, however, individuals with a high index P_i have a higher expected true rank, and vice versa for those with a large negative P_i.

Reporting errors This approach can be generalized to allow for reporting errors. The key intuition is that, if reporting mistakes are uncorrelated across respondents, averaging r_ij rankings across respondents should converge to the true rankings as the number of reported rankings on pair ij tends to infinity.

To formalize this idea, let the income of individual i that is observed by respondent k be

matrix. We discuss this possibility below when we introduce reporting errors.

[8]Variable r_i^{up} is the row sum of matrix $\hat{R}$ for individual i,while r_i^{down} is the column sum of matrix $\hat{R}$ for that same individual.

[9]If we add n to P_i and divide the result by 2, in the full information case, this modified $P_i = r_i$. In this case, unranked individuals receive a middle rank of $n/2$.

given by:

$$y_i^k = y_i + e_i^k \tag{8}$$

where y_i is the true value and e_i^k is an i.i.d. observation error with mean 0 and variance σ_e^2. Respondent k reports $y_i < y_j$ iff $y_i^k < y_j^k$, which implies that:

$$Prob(y_i^k < y_j^k) = Prob(y_i - y_j < e_i^k - e_j^k) \tag{9}$$

With these assumptions, it follows that, for a given σ_e^2, reported ranks on a pair ij are more likely to be correct if y_i and y_j are more different. Put differently, respondents are more likely to correctly rank pairs with dissimilar incomes than those with similar incomes. This means that ranking errors are higher when they matter less for ordinal rankings.

Secondly, if we observe multiple reports r_{ij}^k on a particular ij pair, the effect of reporting mistakes can be minimized by taking the median of these reports across all k respondents when constructing $\overline{R}$. Since each individual report r_{ij}^k is either 0 or 1, the median is 1 if the proportion of 1 reports is more than half, and 1 otherwise. [10] It follows that matrix $\overline{R}$ continues to be only made of 0's and 1's, as in the case without reporting errors. We also note that, by definition, $r_{ji}^k = 1 - r_{ij}^k$, which implies that the median of $r_{ij}^k = 0$ when the median of $r_{ij}^k = 1$ and vice-versa. It follows that $\overline{R}$ can be transformed into matrix A by the same procedure as outlined earlier, and an estimate of the true ranking matrix $\hat{R}$ can be calculated by iteration in the same way as in the case without reporting errors. This matrix, however, need not be equal to the true ranking matrix R due to small sample variation. But it should asymptotically converge to R as the number of reports on *each* ij pair increases to infinity. This is a more stringent condition than requiring that sample size goes to infinity. This key result can be summarized as follows:

Proposition 2 Provided that reporting errors are i.i.d. as shown in equations (8) and (9), $lim\hat{R} = R$ as the number of reports on each ij pair goes to infinity.

Proof. Apply the law of large numbers to each element of the averaging matrix $\overline{R}$.

[10] In practice, we also take into account the difference in the reported ranks of i and j that was reported by observer k. For instance, if k ranked i is fifth position and j in first position, we set $r_{ij}^k = -4$, meaning that i is four positions below j in k's rankings. We then calculate the average of r_{ij}^k over all observers and set $\overline{r}_{ij} = 0$ if the mean is less than 0 and $\overline{r}_{ij} = 1$ if is larger than 0. The reason for doing so is to increase power: from the perspective of the researcher, the expected income gap between each consecutive pair of individuals ranked by k is the same – i.e., the expected gap is the same for i and $i+1$, $i+1$ and $i+2$, etc. Hence a difference of four ranks between i and j means that the expected income difference between i and j is four times larger than the expected difference between i and $i+1$. Including this information in the estimation should speed up convergence to the true pairwise rankings as the number of observers increases.

Confidence intervals Reporting mistakes can be particularly damaging for constructed aggregate rankings in small samples. In particular, it is possible for (directed) cycles to arise in the minimally connected graph $\hat{M}$ – and the corresponding matrix $\hat{R}$. In other words, it is possible for inferred rankings not to be transitive. To illustrate, let the true ranking be $y_1 < y_2 < y_3 < y_4$, but one respondent mistakenly reports $y_4 < y_1$. The minimally connected graph $\hat{M}$ that spans $\hat{R}$ is now a directed circle going from y_1 to y_4 – and then back to y_1. It follows that $r_i^{down} - r_i^{up} = 0$ for all individuals in this directed circle: they cannot be ranked globally. We expect this to arise when respondents are least able to rank two households by income, e.g., when their incomes are relatively similar. Pairwise comparisons do, however, remain informative: in our example, it is only the pairwise comparison $y_4 < y_1$ that is incorrect; all the others correspond to the true ranking. For this reason, in the empirical part of the paper we conduct our analysis both in terms of constructed aggregate rankings $\hat{r}_{ij}$ and in terms of pairwise rankings $\overline{r}_{ij}$.

To investigate the precision of our estimates, we rely on a bootstrapping approach in which we keep the number observers constant in each location. To construct a synthetic sample of observers, we randomly sample with replacement from the available observers in that location. Since each observer is itself randomly selected and can thus be considered as i.i.d., each draw of observers for an EA mimics alternative random samples of observers. This randomization inference approach is similar in spirit to the wild bootstrap in that it allows for correlation in rankings within observer (e.g., Cameron et al. 2008). By sampling only observers from the same EA, we also control for possible correlation in rankings across observers within an EA. We then apply our network matrix methodology to each of the bootstrapped samples. Since Krackhardt, 2022, randomization inference is the preferred approach to inference in network data such as ours.

Monte Carlo analysis In Appendix B we use simulations to examine how precisely the reconstructed rank index P_i approximates true ranks r_i. We do this under different scenarios regarding the information available to respondent observers k, namely: the number of target households they know – and thus can provide a report on; and the precision of the information they have about other households' incomes, which directed affects reporting errors.

We find that when respondents observe a large enough proportion of local households and have accurate information about their income, a lot of information on true ranks r_i can be recovered from the reconstructed rank index – in many cases, P_i accurately ranks all or nearly all target households. As could be expected, the method starts to fail when observers only known a very small proportion of the target households. This is because, in this case, there is insufficient overlap across the pairwise rankings provided by different observers. As a result, the minimally connected set $\hat{M}$ does not constitute a giant component that contains a majority of target households and, hence, many household pairs cannot be ranked relative to each other.

Allowing for observation error in the information available to observers unsurprisingly leads to a deterioration of the precision of reconstructed ranks P_i. This arises because of the creation of branches and cycles in the reconstructed ranking matrix $\hat{M}$.[11] While averaging reported ranks $r_i j^k$ can in principle attenuate the effect of observation error, it is not sufficient, in our simulations, to stop the deterioration. Averaging observers' reports does nonetheless provide valuable information on specific pairwise rankings even in the presence of cycles. We make use of this feature in our analysis.

The HodgeRank Algorithm Bloch and Olckers 2021b have proposed an alternative method for aggregating pairwise comparisons into a single measure. Their method relies on local network information and does not seek to construct the complete ranking matrix $\hat{R}$ and its minimally connected set $\hat{M}$. In this method, the reported rank $Y_{ij}^k = 1$ if k ranks i above j, $Y_{ij}^k = -1$ if k ranks i below j, and 0 if k does not rank i relative to j. The researcher then computes the average of the reported ranks as:

$$Y_{ij} \equiv \frac{1}{N} \sum_k Y_{ij}^k$$

where N is the total number of respondents. The relative 'score' or ranking of individual i is then obtained by applying the HodgeRank algorithm of Jiang et al. 2011 to minimize the squared difference between the scores and the aggregated rankings Y_{ij}. It is the solution to:

$$\min[\sum_{\{i,j\}\in N} ((s_i - s_j) - Y_{ij})] \tag{10}$$

In practice, the scores are the coefficients of individual-specific dummies d_m in a least-square regression of Y_{ij} on $d_m = 1$ if individual $m = i$, $d_m = -1$ if $m = j$, and 0 otherwise.

Jiang et al. 2011 show that the residuals of the least-squares problem corresponds to the cycles in the directed graph. Based on this, they propose the following cycle ratio measure as an indicator of the importance of cycles in the graph:[12]

[11] The only exception is when the proportion of observed households is so small that the minimally connected set $\hat{M}$ is very sparse. Because the reconstructed matrix has few links, observation error does not lead to the creation of cycles. For this reason, in this case the (poor) performance of P_i remains insensitive to observation error.

[12] The reader may recognize that the cycle ratio is equal to $1 - R^2$, and is thus also a measure of the inadequacy of fit. This is because the more the graph is dominated by cycles, the less discriminating the individual HodgeRank dummies $\hat{s_i}$ and $\hat{s_j}$ are, and the less well they predict pairwise rankings Y_{ij}.

$$\text{Cycle ratio} = \frac{\sum_{\{i,j\}\in N}((\hat{s}_i - \hat{s}_j) - Y_{ij})^2}{\sum_{\{i,j\}\in N}(Y_{ij})^2} \tag{11}$$

The higher the cycle ratio, the more cycles dominate the network and the less informative the graph is about aggregate rankings – irrespective of the information contained in pairwise rankings. The cycle ratio, however, only considers the ranked nodes and ignores unranked ones. It also does not correct for the presence of multiple (disconnected) components. We illustrate this limitation in Section 5.

The Newman Algorithm We also examine how our proposed method performs relative to an algorithm commonly used to rank alternatives when only some pairs have been ranked relative to each other. The Zermelo 1929 algorithm has been used extensively to rank individuals, teams, or objects in a variety of contexts that include sports, chess, and consumer choices. The Bradley and Terry 1952 model provides a multinomial logit likelihood foundation for this algorithm. Here we use an improvement to the algorithm introduced by Newman 2022 because of its simplicity and speed.

Suppose we have pairwise rankings over many ij pairs from a variety of observers. Let w_{ij} be the number of observers who rank i above j. Let π_i and π_j denote the (unknown) income of i and j based on these observations. The likelihood of observing the w_{ij} realizations based on a vector of incomes π_i can be written:

$$P(\{w_{ij}\} \,|\, \{\pi_i\}) = \prod_{ij} \left(\frac{\pi_i}{\pi_i + \pi_j}\right)^{w_{ij}} \tag{12}$$

While the log-likelihood of the above expression has no close-form solution, Zermelo 1929 has shown that it can be solved by simple iteration on the vector of π_i's. Newman 2022 proposes to rewrite this algorithm as:

$$\pi_i' = \frac{\sum_j w_{ij}\pi_j/(\pi_i + \pi_j)}{\sum_j w_{ij}/(\pi_i + \pi_j)} \tag{13}$$

where π_i' denotes the revised guess and π_i the previous guess. Newman 2022 proves that this algorithm converges. He also offers a modification of this algorithm to allow for ties, that is, pairs ranked equal by some observers. In both cases, the result of this algorithm is a unique aggregate ranking in the form of a series of perfectly ranked estimates of the π_i's.

3 First application: Rural Indonesia

We start by testing the performance of our method with consumption ranking data from Alatas et al. 2012. In this study, the authors collect rankings of, on average, 8.7 randomly selected rural households in each of 640 villages.[13] Each household is ranked separately by 7.5 observers on average.[14] These observers are the other sample households living in the same village.

Apart from the fact that observers do not include themselves in their reported ranking, there is near universal overlap in the set of households on which each observer provide a ranking in a given village.[15] This feature allows Alatas et al. (2012) to construct a unique ranking index $\tilde{r}_i$ for each sample household by averaging reported rankings within villages. Since the number of reported ranks on each household is large – i.e., 7.5 on average – this serves to reduce observation error. In addition, there is considerable agreement in reported rankings across observers: the standard deviation of constructed ranks from their individual average is small (1.24) relative to the standard deviation of ranks themselves, which is 2.25. Alatas et al. (2012) then compare average rankings to consumption expenditure data collected on 5,352 of the ranked households. They find that average rankings are only poorly correlated with reported consumption.

We apply our network matrix method to the same data in order to obtain estimates of P_i for each sampled household in each village. From this value, we construct an estimated rank $\hat{r}_i$ by sorting observed households according to P_i in each village. We then compare these estimates to similarly constructed ranks obtained by sorting observed households according to their ranking average $\tilde{r}_i$. Both measures are normalized (i.e., divided) by the total number of ranked households in the village to eliminate sample size effects. We find a correlation of 0.95 between the two measures.[16] This indicates that our method is capable of recovering ranking information that is very similar to that obtained by averaging ranks across observers, even though we only use the pairwise comparisons implied by the observers' reported rankings.

We then compare (similarly normalized) household rankings based on reported consumption to those obtained from $\tilde{r}_i$ and P_i. For the Alatas et al. (2012) rankings, we find that

[13] In most villages (94%) the authors rank 9 households. In 32 villages they rank 8, 4 villages they rank 7, and 1 village they rank 6 households.

[14] Of 5711 households ranked in the study, 74.9% are ranked by 8 observers, 12.9% by 7, 4.8% by 6, 3% by 5, and the rest by 4 or fewer observers – except for 17 households that are ranked by 9 observers.

[15] If individual observers only reported on a subset of households in their village, their reported ranks would not be comparable since they would not include the same individuals. In addition, this could also indicate self-selection bias.

[16] This figure is for the subsample of 5,352 households with consumption data. On the full sample of 5,711 households, the correlation is 0.94.

rankings constructed from $\tilde{r}_i$ predict consumption ranks with a coefficient of 0.3804 and a R^2 of 0.145. For the rankings obtained by our matrix method, the corresponding regression coefficient is 0.3807 and the R^2 has the same value of 0.145. This shows that, in this case, the two methods, on their own, perform equally well in terms of predicting consumption. We then regress consumption ranks on *both* estimated ranks to see which one is best at predicting relative consumption. We find that the two rank variables get nearly identical coefficients and t-value while the R^2 rises to 0.1487. This demonstrates that, in these data, both measures contain equally valuable information on consumption rankings. From this we conclude that, with high-quality data such as that gathered by Alatas et al. (2012), our proposed method produces results that are just as good as averaging, even though they are not identical.

4 Second application: Urban Côte d'Ivoire

We now apply our proposed method to original data that we collected in and around the large city of Abidjan in Côte d'Ivoire. As we shall see, this urban setting created conditions very different from those encountered by Alatas et al. (2012) in rural Indonesia: unlike in their setting where all sampled households know each other, our observers are much less knowledgeable about their neighbors. As a result, there is much less overlap in the set of households that are ranked by individual observers in a particular location. This means that averaging rankings to construct a $\tilde{r}_i$ index is not possible in our data. But our network matrix approach still applies.

4.1 Sampling frame

The sample for this peer ranking exercise was directly embedded in a data collection effort conducted under the African Urban Development Research Initiative (AUDRI) at Stanford University. The main objective of AUDRI is to generate representative data of urban and peri-urban populations in the Greater Abidjan, the capital city of Côte d'Ivoire. We use the National Statistical Institute (INS)'s enumerations areas (EAs) as a sampling frame. In 2014, EAs were defined as follows: (i) in urban areas, an EA includes exactly 200 households, (ii) in rural areas, an EA includes all the households living in a village, which can be more or less than 200.

We used the EA geographic delimitation as described in the 2014 database to infer the total rural and urban population. About 83% of the sampling frame live in urban areas in 2014. The AUDRI sample over-samples areas in the process of urbanizing, with 84 "semi-rural" EAs (peri-urban villages) and 622 urban EAs across 16 sub-districts around the capital city of Abidjan. These correspond to the yellow areas in Figure 1. For the ranking study, we selected 20 "semi-rural" EAs and 20 urban EAs among those in the AUDRI sample. [17] The 20 urban EAs were randomly selected among EAs (a) in the two most populated municipalities in Abidjan and (b) defined as "slums" according to the 2014 census.[18] These study areas are named *ranking areas* hereafter.

4.2 Household Sampling and Data

Listing and Individual Surveys As part of a broader data collection initiative (the AUDRI project), we conducted both a listing and an individual survey from a subset of the listing sample. First, the household listing survey was conducted in July-August 2019 in all 706 EAs. In the *ranking areas*, for the sake of this study, enumerators were instructed to list 14 *consecutive* households (i.e., instead of adopting a jumping right-hand rule as in the AUDRI project, enumerators were asked to survey all the 14 first consecutive households to their right, thus surveying households living close to each other and considered as neighbors). They started counting households from the centroid (barycentre) of the EA and moved in circles of increasing radii around the centroid, knocking on doors. The listing survey collected information about each member of the household and basic dwelling characteristics and asset ownership. Only one member of the household (above 18 years old) was surveyed and asked about other members.

From the listing survey, 70% of households in each EA were sampled. Among the sampled households, we randomly selected one adult for what we call "the Individual Survey". The Individual Survey was a 4-hour long questionnaire, which included a wide range of topics about the individuals' labor activities, transport habits, health conditions, and public service access. The Individual Survey was conducted between December 2019 and early March 2020. Out of the 142 individuals selected for the individual survey in *ranking areas*, 119 individuals were surveyed, representing a completion rate of 84%. Reasons for non-completion include re-locations and long-term travel, non-availability, appointment refusal, and insufficient (working) information to join or reach out to the respondent (non-working cellphone numbers, GPS position, and home directives). A detailed description of the collected survey data is given in Dupas et al. 2021.

In sum, the sample in the 34 *ranking areas* contains 207 households, 88 coming from the

[17]Six villages could not be reached by the team of surveyors because the village chief did not allow the study to enter. So we ended up surveying 20 urban EAs and 14 rural EAs.

[18]Slums were defined based on the definition of UN-Habitat 2006, i.e., areas lacking access to improved water, improved sanitation, sufficient living area, durable housing, and secure tenure.

listing survey only, 119 coming from the individual survey.[19] All 207 households responded to the questions necessary to construct a PMT score.

Ranking Survey The ranking exercise was administered in the ranking areas in early March 2020, typically a few weeks after the Individual Survey. To increase the sample size of 207 households discussed before (coming from the listing and the individual survey), we selected additional households on the spot, living close to the 14 neighbors (e.g., the absent ones at the time of the survey). We also added a ranking survey with key informants living in the 34 neighborhoods.

Each respondent was asked to rank up to 14 other households in their neighborhood in terms of their material well-being. Half of the sample was also asked to include themselves in the ranking. Respondents were free to include anyone they knew by name and enumerators were instructed to identify – and confirm with the respondent – those names that appeared in the listing exercise or the survey, so that they could be matched with the data we collected on them. To avoid identification errors, we drop from the ranking analysis those households that could not be matched by the respondent with someone in the listing or survey.[20]

The ranking exercise was done with a total of 507 respondents, of four types (see Table 1):

1. *Target households* (N=119): These respondents are taken from the 70% listed households selected for the *Individual Survey*. We re-visited them for the ranking exercise a few weeks after the Individual Survey to collect rankings.

2. *Listed households* (N=88): These respondents are taken from the *listing survey* that were not selected for the Individual Survey. In order to administer the ranking survey with these households, we contacted the head and scheduled an appointment, then surveyed a household member available at home at the time of the enumerators' visit.[21] The survey includes a subset of the modules in the Individual Survey, notably consumption information, and the rankings. These first two groups of individuals are identified with an A letter in the graphical analysis.

[19] Dwellings with no one at home at the time of the first knock were included in the count and, if sampled for listing, revisited later in the day or in the next few days to attempt to conduct the listing survey. Thus, listed households considered "absent" are households for whom no member could be surveyed during the listing. Note that in four EAs, due to miscommunication in the field, dwelling closed on the first visit were neither counted nor listed. Thus, surveyed households live quite far apart (a few blocks away) from each other in these four areas. We control for this case in the analysis when possible.

[20] We did attempt to include names that were mentioned by more than one respondent in an EA, in an effort to improve the quality of reconstructed rankings. But there too few of them and we could not be sure these names identified the same households.

[21] Note that we did not randomly select the respondent for this group.

3. *Additional Households selected on the spot* (N=230): These are households who were absent during the listing. We attempted to visit them again and, when we found someone available, administered the ranking questionnaire. We also asked for additional information (e.g., PPI questions and consumption module) since these households were not surveyed before. We surveyed a household member found at home at the time of the enumerators' visits in the EA. These households are identified with a B letter in the graphical analysis.

4. *Key Informants* (N=70): We visited key informants in each ranking area. Since "chiefs" are rare or nonexistent in Abidjan's urban areas, we decided to survey traders around the surveyed dwellings in both rural and urban areas for consistency. Those individuals are identified with a C letter in the graphical analysis.

This protocol was used to generate a sufficient overlap between individuals' rankings while keeping the number of surveyed respondents small enough to be cost-effective compared to a full survey approach to poverty targeting. While we have detailed information on the target households and some information on listed households, we collected much less data on additional households and key informants since the goal was primarily to have them ranked target households. The share of household heads surveyed is relatively similar across groups – except for the key informants (39-47% as opposed to 26% for key informants).

4.3 Poverty measures

We collected several poverty measures, described below. For all measures, a lower value indicates *greater* poverty.

Consumption / expenditures We focus on three main household consumption measures collected on most of the sample:[22] (1) *Value of food consumption in the last week* before the survey: we used a typical consumption module to collect recall information on the value of household consumption of cereals, pulses, spices, milk products, meat, bread/pasta, vegetables, fruits, drinks, alcohol, and other consumables[23]. (2) *Value of conspicuous/social consumption in the last month* before the survey: we asked specific questions about non-food expenses, such as communication, beauty products, entertainment (concert, bar, cinema, games), and charitable contributions. (3) *Spending on durables in the last 12 months* before

[22]The information is missing for 13 respondents who could not be reached at the time of the individual survey or did not recall their past consumption

[23]Note that the consumption module was administered earlier for respondents in the individual survey (about a month before), and we pulled all data together for consistency.

the survey: these include expenses for clothing, shoes, furniture, school fees. Consumption information was collected on most respondents either at the time of the individual survey or right before the ranking activity.[24]

Proxy Means Test (PMT) We re-compute the PMT score built by the government of Côte d'Ivoire in 2015. The weights imputed to each household characteristic are the coefficients from a regression run by the government of Côte d'Ivoire on survey data collected in 2015 as part of the living standards measurement study. The regression predicted the log(food consumption per capita) using about 25 predictors (assets, house characteristics, etc.). The PMT score was performed on two separate models, urban and rural. We use on these same weights to build a PMT poverty index, depending on whether we surveyed a respondent in a rural or urban areas, whose distribution for our sample is shown in Figure 2. We also cross-validate the methodology used by the Government of Côte d'Ivoire with our data and compare the fit of the regression in predicting log(food consumption per capita). We obtain a similar fit and relatively similar coefficients in terms of magnitude (Table A2).[25] We cannot construct this index on additional households and key informants since, as described above, they were not administered the full questionnaire.

Poverty Probability Index (PPI) We also use an index developed by Innovations for Poverty Action (IPA), tailored to the Ivorian context. The PPI was created in April 2018 using Côte d'Ivoire's 2015 living standards measurement study. The index relies on ten questions (geographic location, household characteristics, living conditions), and we built a specific PPI score from these ten questions (see in Table A1) for the entire sample of households. We observe in Figure 2 that our sample of households is widely distributed across their poverty status and slightly wealthier than the average household in the country. This fact is consistent given the urban sample studied here, living in or around the capital city Abidjan. In Table A2, Column (6) and (7), we report the fit from the PPI regression, i.e., regressing log(food consumption per capita) on the variables used to build the PPI index. We obtain a reasonably large R^2.

Table 2 shows summary statistics from the various measures, separately for urban and peri-urban (rural) EAs. Figure 3 shows how the measures of poverty are correlated with

[24] A few respondents answered that they did not know when asked about a particular good consumed (typically 1-3% of the sample in a given consumption question). In such a case, we replace the answer by the average in the enumeration area to preserve the sample size.

[25] Note that the number of observations is limited for PMT since most households surveyed for rankings did not respond to the detailed Individual survey, including all PMT-related questions, as described in Section 4.

each other. In particular, the PPI index is positively correlated with the PMT score, food consumption, conspicuous consumption, and expenditure on durables.

4.4 The ranking exercise

A total of 507 respondents were asked to rank neighboring households based on neediness. Respondents were told that the study's goal was to understand the lives of the communities around Abidjan and the extent to which people interact and know their neighbors. We asked respondents to name all of their neighbors, and the surveyors classified each of them on the tablets using a pre-loaded household list from the listing[26]. We then asked the respondents questions about each listed neighbor. Finally, we asked respondents to rank *all* the neighbors listed from the poorest to the richest. We did not ask them to rank within each possible pair, but instead to put the list in order. They could not leave anyone out of the ranking.

Respondents were told that their ordinal rankings (and any other survey information we collected from them) would *not* be used to provide anyone with any gift or anything else. As such, there was no incentive for participants to be strategic in their rankings and reporting. All respondents were told that they could do no better than telling the truth.

We randomly varied whether respondents were asked to rank their own household relative to others. This was meant to estimate the extent to which adding selves improves the accuracy of the constructed rankings by adding more ranked pairs, or whether it distorts reported rankings, either because of mis-perception of one's own position Hvidberg et al., 2020 or because of manipulation in peer rankings even absent any direct financial incentives to do so (Bloch and Olckers 2021a). Within each neighborhood, around half of the respondents were asked to rank their household; the others were not.

We collected specific data to understand how people form their rankings and how they perceive poverty more generally. We asked three questions about poverty's perceptions, i.e., (i) the perceived poverty level of their household; (ii) how they regard their households compared to others in the neighborhood; and (iii) how they think others perceive the respondent's household. This data is summarized in Table 3. While 29% of respondents surveyed consider their households as poor, 21% consider themselves poorer than their neighbors, and 21% think that other people would classify their households as poor. Interestingly, 53% of the respondents who report their household as poor do not think that others regard them as poor.

We then asked respondents to tell us the criteria they used to classify households and to define poverty "in their own words". The vast majority of respondents refer to poverty

[26]Respondents had to list at least 5 neighbors and were told to list as many neighbors as possible (maximum of 14)

as "food deprivation" (80%); some mention "unresolved health problems" (43%). To rank households, respondents predominantly report using the household head's occupation (49%) and whether households reported facing financial problems to them (49%). Finally, most households in our sample declare visiting the listed neighbors regularly (56%) and about half of them report receiving health/money advice from them.

5 Poverty rankings

5.1 Sample statistics

By design, in each of 34 EA we aim to rank up to 14 households for whom we have the name of the household head or his/her spouse – a total of 476 target households. This represents 91 household pairs per EA, or 3094 pairs in total. We surveyed 14.9 observers in each EA – 507 in total. Each of these observers is allowed rank up to 14 households.[27] If all 507 observers rank 14 households in their EAs, this would yield 46,137 pairwise rankings in total, with up to 15 distinct reports on each pairwise comparison. This undoubtedly would allow us to construct complete rankings of the 14 target households, either by averaging ranks, as was done by Alatas et al. (2012), or by our network matrix approach. We do not, however, expect all observers to rank 14 households because, unlike in rural areas, urban residents rarely know all their neighbors: 17% of the respondents in our urban sample arrived in their neighborhood within the last year, and only 6% were born in the area. To compensate for this, the number of observers is double the average number of 7.5 observers per target household in the Alatas et al. (2012) study.

Overall, we only collected 1820 distinct pairwise rankings – 3.9% of the maximum achievable figure of 46,137. This is well below what we were anticipating – and below Alatas et al. (2012) for whom this proportion is 100%. These 1820 pairwise rankings involve 837 distinct household pairs, which represent only 27% of the 3094 possible pairs of targeted households in our sample. These 837 distinct pairs involve 364 target households in total, which means that 24% of targeted households were not identified or ranked by any of our observers. For 442 of these pairs (52.8%), we have a single ranking – which rules out relying on averaging to minimize observation error. The average number of pair rankings per observer is 2.17 and the maximum is 9. This is much lower than in the Alatas et al. (2012) data, where the corresponding average number of rankings per pair is around 6 and 96% of households are ranked by five observers or more. It should be emphasized that, with such low level of overlap across observer rankings, the Alatas et al. (2012) method of averaging ranks across

[27] 15 for the half of the respondents who were asked to rank themselves; 14 for the others.

households would be meaningless. We also note that, when a pair of target households is ranked by multiple observers, there is considerable disagreement among them: only 38% of multi-ranked pairs have full agreement among observers, and 22% are equally split.

All this confirms the original motivation for our effort, namely, that urban households are less able than rural households to rank their neighbors by poverty level. What remains to be seen is whether our network matrix method can recover useful ranking information from such sparse and noisy data.

5.2 Reconstructing ranks using the network matrix method

We now combine the pairwise rankings that we collected to construct poverty ranking estimates using the methodology described in Section 2. To recall, for the purpose of constructing the adjacency matrix A, if there exist multiple reports on a particular pairwise ranking r_{ij}, the median rank $\overline{r}_{ij} = 1$ if more than half of the reports are positive, and 0 otherwise.[28]

Ranking graphs for the 34 locations in which relative rankings information was elicited are presented in the appendix (see Figures A1 and A2 in urban slums and rural villages, respectively). Each node represents a household, identified to respondents by the name of the head of household, spouse, age, and residence location. We also measured the PPI Index for each household whenever possible, divided it into four equal categories across the entire sample and added it directly on the directed graphs. Households with ids in the A's are households sampled for the Individual Survey.[29] Households with an id number in the B's are neighbors added as respondents for the ranking exercise. In contrast, households with an id number in the C's are "key informants" identified in the neighborhood – typically traders. We did not seek to explicitly elicit rankings on households with B's and C's IDs, but, as they are neighbors of A's, they are sometimes ranked, and we manually matched based on similar names, ages, and household sizes. We keep them in the graphs because omitting them sometimes breaks the graph into multiple components.

We immediately note that some locations provided much more information than others. Locations 13, 15, 22, and 28 only contain information on two or three pairs of households. Thirteen locations are broken into two or more components that cannot be ranked relative to each other (i.e., 1, 3, 5, 10, 11, 16, 17, 18, 23, 24, 25, 29, and 33). This pattern leaves fourteen locations with a single component containing at least five households (i.e., 2, 4, 6, 7, 8, 9, 12, 14, 19, 20, 21, 30, 31, 32). Of these 13 locations, some (i.e., 2, 4, 6, 7, 20, 30) contain

[28]When exactly half of the reports are 1 and the other 0, we set $\overline{r}_{ij} = 0 = \overline{r}_{ji}$ instead of setting both to 1. This avoids creating unnecessary cycles.

[29]These households are included in the graphs even if they could not be found when the listing survey was done.

at least one (directed) cycle involving a subset of nodes; while the others are transitive.

In Table A4, we report r_i^{up} and r_i^{down} statistics for the 14 locations with a single component and at least five ranked households. We immediately notice that r_i^{up}, the number of ranked households who are richer than a given household, is not the same as $n - r_i^{up} - 1$. For instance, in location EA 2, there are 16 ranked households. Household 203 has no household ranked richer but thirteen households ranked poorer, while household 301 has no household ranked richer, but ten households ranked poorer. If we look at the directed graph of relative rankings for location 2 (Figure A1), we note that household 203 is at the top of a long sequence of ranked households. In contrast, household 301 sits on a side branch above household 209, but is un-ranked relative to households 203 to 208. This characteristic means that household 301 could be as rich or even richer than 203, but in all likelihood, it is poorer. We cannot, however, clearly rank 301 relative to households 208, 205, 201, 202, and 203. We also do not know how 902 ranks relative to household 208: it could be poorer or richer. This example illustrates the partial nature of the information we can recover from the reported rankings.

We also observe situations in which multiple households share the same number of poorer and richer households. This arises when the constructed rankings are non-transitive, i.e., when they include a cycle. Households located at either end of the ranking chain stand on their own, but for instance, households 203, 206, 210, all have the same (large) number of households ranked above and below them in EA 6. This is because there is a directed cycle between them, meaning that, based on our definitions, they are both richer and poorer than each other – i.e., they are ranked, but not in a meaningful way. This is clear in the directed graph of relative rankings for location 6 (Figure A1). One extreme case of this is location EA 30, in which all households but two are located on a set of large cycles: pairwise rankings exist, but they do not induce a meaningful aggregate ranking.

While these findings do *not* constitute an indictment of the methodology, they reduce the usefulness of its results when, as in locations 2, 4, 6, 7, 20, 30 with one or more cycles, the rankings data is contradictory. Constructed *pairwise* rankings nonetheless remain potentially informative: as explained in Section 2, a cycle can be caused by a single misreported link by a single respondent. All the other directed links (i.e., inequality relationships) in this cycle may still be correct. Given this, in the subsequent statistically analysis we consider both the aggregate constructed rankings and relative position P_i, as well as the constructed and reported pairwise rankings $\hat{r}_{ij}$ and r_{ij}, respectively.

5.3 Confidence intervals

As indicated in Section 2, we use a wild-bootstrap to estimate the precision of our rank estimates. We find that index P_i for each household i varies a lot across simulated samples of observers. This is illustrated in Figure 7 which shows, for each estimated index value P_i, the distribution of P_i estimates obtained from 100 bootstrapped samples of observers, together with the corresponding 90% confidence interval. It is immediately clear that the standard error of each estimated P_i is large, reflecting the lack of agreement in rankings among observers.

Figure 8 does the type of analysis as Figure 7 for the intra-EA ranks implied by the estimated P_i. Again we see that estimated ranks are imprecise, with the possible exception of highly ranked households, for whom the confidence intervals are tighter. These households, however, only account for a small fraction of the ranked households. From this, we conclude that there is too much disagreement among observers' reported rankings to allow us to reconstruct accurate rank estimates.

5.4 Comparison with the Newman estimator

We started this section by noting that, in the data we collected in Cote d'Ivoire, averaging ranks to construct an aggregate poverty ranking would not work because of insufficient overlap across observers. Next we noted that, while our method does recover some relevant information about aggregate rankings, this information is partial and imprecise. Can the Zermelo algorithm do better?

We applied the Newman version of the Zermelo algorithm to our data, both with ties and without ties. Since the results are quite similar, we focus here on the simpler version without ties. We first note Newman estimates of ranks are correlated with the ranks based on our summary statistic P_i: the correlation coefficient is 0.786. However, the Newman method only ranks 207 households in total, compared with 364 ranked households with our method – i.e., it only ranks 57% of our ranked households.

Is this reduction in coverage compensated by an increase in precision? To investigate this possibility, we estimate Newman estimates for each of the 100 bootstrapped samples used in the previous sub-section. Results are shown in Figure 9. We do not observe any shrinkage in the confidence intervals: if anything, they appear larger than in Figure 8. Based on this, we conclude that, for our data, our estimator does better than the well-established Bradley and Terry 1952 estimator.

6 How informative are the rankings?

We now examine whether reported rankings are informative about differences in consumption levels across households. To this effect, we start by regressing differences in our various poverty indices between households i and j (in the dyadic dataset) on the reported and constructed pairwise ranks between i and j, as described in Section 2. Here, the "reported rank" variable is the share of reported ranks showing j richer than i. The "constructed rank" variable is a dummy equals to 1 if j is ranked richer than i, computed for part of the possible pairs. In the presence of a cycle, it is possible that j is ranked richer than i *and* i is ranked richer than j.

We also add an alternative way to aggregate rankings: the HodgeRank score, as developed by Bloch and Olckers 2021b and described in sub-Section 2. The set of HodgeRanks of all individual i is the set of scores s_i that minimizes the squared difference between scores and aggregated rankings. The difference in scores between j and i is used as an independent variable: the higher the difference, the richer j is ranked compared to i. A negative difference implies that j is ranked poorer than i. We note that the observed correlation between the constructed rank and the difference in HodgeRank scores is high ($\rho = 0.84$)[30].

All differences in outcomes are taken as j's value minus i's value. Results are presented in Table 4. We show regressions on four main outcomes, (1) food consumption per capita; (2) number of months during which the household suffered from food shortages; (3) PMT score; and (4) PPI score.[31] Except for food shortages, these measures are constructed such that a greater value means *less* poverty. Thus, if ranks are informative, we expect the coefficient of the different rank measures to be positive in columns 1, 3 and 4: when i is ranked poorer than j, j's consumption, PMT and PPI scores should be higher than i's. The reverse is expected for column 2.

We observe limited evidence that pairwise rank measures are informative about consumption differences per capita, PPI, or PMT. Most of the estimates are non-significant and are sometimes of the wrong sign (e.g., for food expenditure per capita). The coefficients reported in column 2 are negative as predicted, but only significant for constructed and hodgerank score's difference. The estimated R^2 is quite low throughout. From this, we conclude that, in general, rankings contain relatively limited information about consumption differences across ranked households.

Next we move the analysis to the level of the individual. Here the dependent variable

[30]Since HodgeRank scores are assigned to each household, it is possible to compute the differences in HodgeRank scores for all possible pairs $i - j$, thus increasing the number of observations.

[31]In Table A5, we run the same regression for an extensive set of outcomes, e.g., different consumption variables, shortfall in consumption, an index of improvement in consumption relative to last year.

is the level of the consumption measure. We estimate two sets of regressions, depending on which measure of aggregate rank for household i we use: its relative position P_i; and its estimated HodgeRank score $\hat{s}_i$. Results, presented in Table 5, show that estimated coefficients are not statistically significant. These results are perhaps not surprising, given the findings from pairwise regressions – and the fact that the information content of P_i is more affected by the presence of cycles than reported ranks.

To understand why individuals do not seem to make accurate rankings, we examine which characteristics of households appear to be predictive of reported ranks. The results are shown in Table 6. Regarding the asset/wealth data, as aggregated by the PPI index, the patterns are mostly consistent with expectations: if i has a lower PPI index than j, k is more likely to report that i is poorer than j. For consumption variables, j is ranked richer than i (positive coefficient) if j reports higher food consumption or higher spending on durables. The variables "months of food shortages", "expressed food worries in the last 12 months", or "received gifted food in the past week" also consistently predict rankings households with food deprivation poorer. Interestingly, conspicuous consumption expenditures do not predict reported rankings.

6.1 Ranking Accuracy

To examine the accuracy of reported ranks, we compare them with the rankings obtained from the PMT and PPI indices constructed survey data – which, for the purpose of this exercise, we regard as the true rankings. Overall, ranking accuracy is pretty low: reported rankings are right 52.5% of the time when we take PMT rankings as comparison, and 55.8% when we use PPI rankings.[32] Ranking accuracy is even lower if we use food expenditure per capita as comparison. Such average levels of accuracy are only barely above what could be achieved by random guessing. This is not because the indices themselves are random noise, though. We indeed find that PMT and PPI predict consumption per capita rankings well: the pairwise $i - j$ difference in PMT and PPI has the same sign as the difference in consumption per capita in 71% and 69% of cases, respectively. We nonetheless find that k respondents are more able to correctly rank ij pairs in terms of PPI or food consumption if the difference between i and j is large enough. When the PPI indices or food consumption levels of i and j are similar, k's ranking ability is not different from flipping a 50/50 coin. But when the difference is large (e.g., around the 80th or 90th percentile), k is 1.4 to 1.5 times more likely to provide a correct ranking (i.e., the probability of correct ranking is

[32]Because PPI is an integer index, some i, j pairs have the same PPI value and thus cannot be ranked. They are omitted from this analysis.

around 0.58 to 0.60%).[33] It follows that, with a sufficiently large number of reports on such ij pairs, our methodology yields a correct ranking with a reasonably high probability. To illustrate, when the probability that each k respondent ranks correctly an ij pair is 60%, the probability of mis-ranking falls from 40% for a single report to 35.2%, 31.7%, and 24.7% for 3, 5 and 11 reports, respectively.[34] This only applies to ij pairs that are quite different, however.

To get a better sense of how informative the reported ranks are overall, we simulate a ranking model calibrated on the data to assess how large the standard deviation of observation error would have to be in order to produce the ranking accuracy reported above.[35] To conduct this counterfactual experiment, we use as 'truth' the two welfare indices constructed from the data, PPI and PMT. All indices are standardized to have mean 0 and variance 1. We assume that each observer k sees a signal $y_i^k = y_i + e_i^k$ where y_i is either the PMT or PPI of i and where e_i^k is, as before, an i.i.d. observation error with mean 0 and variance σ_e^2 – and similarly for y_j^k. We then construct a simulated reported rank $r_{ij}^k = 1$ if $y_i^k > y_j^k$ and 0 otherwise. We do this for various values of σ_e until we find a value that gives the same ranking accuracy as above.[36]

Unsurprisingly, given the poor ranking accuracy of the actual reported ranks, we must posit quite a large σ_e in order to reproduce their ranking accuracy: across simulated vectors of observation errors, σ_e has to be at least 7.5 in order to reproduce the 52.4% PMT targeting accuracy of reported ranks; and the corresponding values for PPI is 3.5. In most simulations, σ_e has to be larger than 10 to match the accuracy of reported ranks. In other words, the standard deviation of the observation error e_i^k has to be a large multiple of the standard deviation of the truth y_i in order to account for the low ranking accuracy of reported ranks. This exercise is purely indicative, since we do not observe the 'true' welfare of individuals i and j. But it gives an idea of the magnitude of the observation error that characterizes our empirical setting.

Poor ranking accuracy may be due to a poor selection of observers k. If so, the usefulness of our method may be improved by selecting respondents with observable characteristics that

[33]We also find that k respondents are more sensitive to differences in PPI or food consumption in ij pairs that are on average richer. While this finding is not statistically significant, it nonetheless suggests that respondents are better able to distinguish differences in socio-economic status among the rich than among the poor.

[34]Based on binomial distribution calculations. Results for an even number of reports are slightly less precise because we classify a mean of 0.5 as missing, but they show the same improvement in accuracy with the number of reports.

[35]The simulated dataset includes all the i, j, k triads for which a rank is reported by k and PMT or PPI values exist for both i and j. Pairs i, j that have identical PPI are dropped from the simulation. The sample size is 446 distinct triads for PMT and 283 for PPI.

[36]By construction, the ranking accuracy of the simulated reported ranks is 100% when $\sigma_e = 0$.

predict ranking accuracy. To investigate this possibility, we regress each observer's ranking accuracy on a vector of respondent characteristics. To avoid oversampling observers who provide more rankings, we define, for each individual ranker k, the *Ranking Accuracy* of that observer as the *share* of pairs $i - j$ for which k accurately ranked i poorer or richer than j according to the $i - j$ difference in PPI, PMT, of household food expenditure per capita. We then regress this variable on observer characteristics. The results are shown in Table 7.[37] We do not find any convincing evidence that observer heterogeneity predicts ranking accuracy, implying little scope for improving on our method by over-sampling observers with certain characteristics.

6.2 Poverty Targeting

From a policy standpoint, accurate rankings between any given pair may not be needed. Instead, the policymaker may simply want to identify who is, say, below the median of the distribution. To test whether aggregate peer rankings can be used to do such identification, we create a dummy equal to 1 if a household's aggregate ranking puts it below the median of its EA. It is zero if the household is ranked at or above the median and missing if the household is not ranked. In Table 8, we use this dummy as a regressor, testing whether it correlates with whether the household is below the median based on the survey measures. Column 1 compares the categorization obtained thanks to the peer rankings exercise to that obtained from the PMT measure, column 2 to the categorization obtained from the PPI measure, and column 3 to the categorization based on food expenditure per capita. Quite strikingly, being categorized below the median does not significantly increase the likelihood that one is below the median based on any of the three survey measures, suggesting that even coarse categorizations are difficult to obtain from peer rankings. The below median classification based on the Hodgerank score does slightly better in predicting the PMT-based poverty classification, but not the others.

To test whether these mostly-zero results are driven by the fact that the probability of being ranked could itself be affected by one's position, we create a dummy equal to 1 if a household could not be given an aggregate ranking (this happened when none of the respondents surveyed listed that household as a known neighbor). Around 23% of the sample is "unranked". To test whether those unranked are disproportionately poor or disproportionately rich, the bottom panel of Table 8 shows regressions with this "unranked" dummy as the regressor. There is no statistically significant correlation, suggesting that categorizing those "unranked" as poor would not help improve targeting based on peer rankings.

[37] We can only estimate accuracy for respondents who ranked at least one pair of neighbors for which we have completed surveys. This represents 58% of respondents.

7 The self-ranking treatment

In this section we test whether including self-ranks improves accuracy by exploiting the fact that we randomized whether observers were asked to include themselves into their rankings. Since observers presumably have better information about themselves, they should be better able to rank themselves relative to others. Self-ranking may also be less accurate if observers do not rank themselves truthfully for instrumental reasons, e.g., they may understate their rank if they expect their report to be used in an anti-poverty program targeting (e.g.,Bloch and Olckers 2021a). To eliminate this concern, we ensure respondents understand that their rankings will not be used for any targeting purpose. Misreporting may also arise out of self-image or social-image considerations, e.g., to look less poor than they are (e.g.,Ghiglino and Goyal 2010).

To test whether including self-ranks improves the accuracy of aggregate rankings, we re-estimate Table 8 excluding self-ranks from the data. Results, shown in Table 9, show that dropping self-ranks improves targeting accuracy somewhat. Those who rankbelow the median of their EA when self-ranks are excluded are 12.7 percentage points more likely to be below the median of the PMT index (Panel A, column 1). Similarly, those who are below the median HodgeRank score are 16.9 percentage points more likely to be below the median PMT (Panel A, column 2). These differences are significant at the 10% and 5% level, respectively, and they are larger than those reported in Table 8 when self-ranks are included. This suggests that the way observers rank themselves is unhelpful for the purpose of identifying the relatively poor.

To investigate why that is the case, we check whether respondents rank themselves differently from what other observers report about them. To do this, we estimate a regression of the form:

$$y_{ij}^k = \alpha S_{ij}^k + \theta_{ij} + u_{ij}^k$$

where: $y_{ij}^k = 1$ if observer k ranks i poorer than j, 0 if k ranks i richer than j, and missing otherwise; $S_{ij}^k = 1$ if $k = i$, -1 if $k = j$ and 0 otherwise; and θ_{ij} is a pairwise fixed effect. If $\alpha < 0$, this implies that respondents give themselves a higher rank than the rank others give them – possibly reflecting self-image or social-image considerations. In contrast, if α is larger than 0, it means that observers rank themselves lower than the rank others give them, i.e., it is more often the case that $y_{ij}^k = 1$ when $i = k$ than when $i \neq k$ and that $y_{ij}^k = 0$ when $j = k$ than when $j \neq k$. This may be due to a social norm of humility or as a learned heuristic to avoid requests for financial assistance from neighbors.[38] The experiment is not

[38] Alternatively, $\alpha > 0$ could arise because observers rank others based on their conspicuous consumption but rank themselves based on their own full consumption. This would imply some myopia: people systemat-

designed to identify which is the most likely explanation.

Results, presented in Table 10, show that α is significantly smaller than 0 and that the magnitude of the effect is large. By construction, actual ranks are equal to 1 half of the time. A coefficient of -0.28 means that respondents rank themselves poorer than others 22 percent of the time relative to the median of 50 percent. This indicates that a large fraction of respondents rank themselves as richer than others even though they are judged to be poorer by other observers. Further confirmation comes from observing that 62% of respondents rank themselves among the richest of their neighbors while only 22% rank themselves among the poorest. This rules out strategic under-reporting, but over-reporting is substantial. In addition, we find that self-ranking observers who are in the bottom half of their EA in terms of PMT or reported consumption tend to overestimate their ranking more than those in the upper half (see Columns 3-7 of Table 10). This suggests that there may be a psychological cost to admitting one's own poverty (e.g.,Ghiglino and Goyal 2010,Bramoullé and Ghiglino 2022).

8 Propensity to rank and to be ranked

A main finding from our application is that rankings are far from complete. This is because many respondents did not know some of their neighbors enough to list and rank them. To further investigate the correlates of the propensity to rank and the propensity to be ranked, we construct a *dyadic dataset* indexed by the respondent k and a ranked household i in the same EA. We create a dependent variable $m_{ki} = 1$ if respondent k ranks household i relative to any other household, and 0 otherwise. We regress this dummy on characteristics of both i and k in Table 11.

We find that the geographic distance between the respondent k and household i has a significant (negative) effect on reporting. The absolute magnitude of the coefficient is small, but this is primarily because average reporting is low to start with.[39]

In column (2), we add information about consumption. We see that, some variables indicating that household i is poor tend to be negatively correlated with being ranked by k. For instance, households who experience food shortages over more extended periods are less likely to be reported on by k. We find limited evidence, however, that detailed consumption expenditures as reported by household i consistently helps predict reporting by k. If anything, the higher the food consumption, the less likely a household would be ranked

ically misjudge the true poverty of others even though they realize that their own conspicuous consumption gives an inflated image of their true prosperity.

[39]Experimentation with alternative functional forms indicate that the log form chosen here fits the data well.

by others. The category of expenditures classified as 'conspicuous', e.g., beauty products, eating out, and charitable contributions, positively predict being ranked by others.

Overall, these findings confirm that k's propensity to rank i can be partly accounted for by observable characteristics of i and how they compare to k's. This is reassuring because it indicates that k takes relevant characteristics of i into consideration when choosing to rank i relative to other households. The findings also suggest that richer households, at least in terms of assets, are in general more likely to be ranked. A plausible explanation is that their wealth is easily observable. With this ranking methodology, the rich are more likely to be ranked and thus the poor are less likely to appear in constructed rankings. This finding is problematic if the purpose of eliciting income and wealth rankings is, as is often the case, to identify the poor.

9 Conclusion

This paper introduced a new method for eliciting relative poverty rankings that aggregates partial poverty rankings obtained from multiple individuals. We demonstrated that the method works in principle and performs well in the rural setting of Alatas et al. 2012. In the urban setting of Abidjan, however, the constructed rankings are incomplete in all studied neighborhoods. Furthermore, they are not always transitive and many contain cycles. A more accurate picture would require increasing the density of reporting.

We also find that, in both settings, pairwise rankings reported by respondents are only mildly correlated with various poverty measures collected on households targeted by the ranking exercise. The same holds for constructed aggregate rankings. These results confirm for Côte d'Ivoire the conclusion of Alatas et al. 2012 for rural Indonesia, namely, that reported ranks do capture relevant information about relative welfare but this information is noisy. We also find that reported rankings seem to reflect a few observable expenditures only. In addition, we investigate whether reported rankings correlate better with the conspicuous consumption expenditures of the target households. We find that they do not. We also note that respondents asked to include themselves in their ranking tend to overstate their relative income position.

From this experiment, we conclude that, in an urban setting where people know little about their neighbors, rankings constructed based on peer rankings are probably insufficient to achieve poverty targeting at a cost lower than surveying households directly. For the same reason, our method seems to work better in a rural setting – and does not require the extensive overlap in comparison sets across observers that averaging ranks requires.

In this paper, we have demonstrated the potential usefulness of an original methodology that equals – and often surpasses – existing methods for aggregating partial rankings: it requires fewer assumptions than the commonly used Zermelo-Bradley-Terry algorithm and is able to rank more pairs; and it avoids the potential biases of averaging ranks or scores across partially overlapping observers. We also provided a way of producing confidence intervals for estimated ranks – which, in our urban data, confirmed that ranks were only estimated noisily. This method can be extended to many situations in which individuals face options over which they have identical true preferences, but only have partial information and can only rank some options relative to each other. Our method is not applicable if respondents are known to have different preferences and the purpose of aggregation is to achieve a compromise between them. More research is needed on aggregation approaches suitable for such situations.

References

Alatas, Vivi, Abhijit Banerjee, Arun G. Chandrasekhar, Rema Hanna, and Benjamin A. Olken, "Network Structure and the Aggregation of Information: Theory and Evidence from Indonesia.," *American Economic Review*, 2016, *106* (7), 1663–1704.

_ , _ , Rema Hanna, Benjamin A. Olken, and Julia Tobias, "Targeting the Poor: Evidence from a Field Experiment in Indonesia.," *American Economic Review*, 2012, *102* (4), 1206–4–.

_ , _ , _ , Benjamin A Olken, Ririn Purnamasari, and Matthew Wai-Poi, "Does elite capture matter? Local elites and targeted welfare programs in Indonesia," Technical Report, National Bureau of Economic Research 2013.

_ , _ , _ , Benjamin Olken, Ririn Purnamasari, and Matthew WaiPoi, "Does Elite Capture Matter? Local Elites and Targeted Welfare Programs in Indonesia," *AEA Papers Proceedings*, 2019, p. 109: 334–339.

Banerjee, Abhijit, Rema Hanna, Benjamin Olken, and Sudarno Sumarto, "The (lack of) Distortionary Effects of Proxy-Means Tests: Results from a Nationwide Experiment in Indonesia," *Journal of Public Economics Plus 1*, 2020.

Basurto, Maria Pia, Pascaline Dupas, and Jonathan Robinson, "Decentralization and efficiency of subsidy targeting: Evidence from chiefs in rural Malawi," *Journal of public economics*, 2020, *185*, 104047.

Beaman, Lori, Niall Keleher, Jeremy Magruder, and Carly Trachtman, "Urban Networks and Targeting: Evidence from Liberia," *AEA Papers and Proceedings*, 2021, *111*, 572–576.

Blanchflower, David G. and Andrew J. Oswald, "Well-being over time in Britain and the USA," *Journal of Public Economics*, 2004, pp. 88(7–8), 1359–1386.

Bloch, Francis and Matthew Olckers, "Friend-Based Ranking," *American Economic Journal: Microeconomics*, 2021.

_ **and** _ , "Friend-Based Ranking in Practice," *AEA Papers and Proceedings*, 2021, pp. 111: 567–71.

Bradley, R. A. and M. E. Terry, "Rank analysis of incomplete block designs: I. The method of paired comparisons," *Biometrika*, 1952, *39*, 324–345.

Bramoullé, Y. and C. Ghiglino, "Loss Aversion and Conspicuous Consumption in Networks," *CEPR Discussion Paper*, 2022, *17181.*

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller, "Bootstrap-Based Improvements for Inference with Clustered Errors," *The Review of Economics and Statistics*, 08 2008, *90* (3), 414–427.

Chris, Fujii Tomoki Lanjouw Peter Ozler Berk Elbers and Wesley Yin, "Poverty alleviation through geographic targeting: How much does disaggregation help?," *Journal of Development Economics*, 2007, pp. 83(1): 198–213.

Chris, Jean O. Lanjouw Elbers and Peter Lanjouw, "Micro–Level Estimation of Poverty and Inequality," *Econometrica*, 2003, pp. 71(1): 355–364.

Deaton, Angus, "The Analysis of Household Surveys : A Microeconometric Approach to Development Policy.," *Washington, DC: World Bank. © World Bank.*, 2019.

Dupas, Pascaline, Marcel Fafchamps, and Eva Lestant, "Report on the AUDRI 2019-2020 Cote d'Ivoire Surveys.," *King Center, Stanford University*, 2021.

Elbers, Chris, Tomoki Fujii, Peter Lanjouw, Berk Özler, and Wesley Yin, "Poverty alleviation through geographic targeting: How much does disaggregation help?," *Journal of Development Economics*, 2007, *83* (1), 198–213.

Fafchamps, Marcel and Forhad Shilpi, "Subjective Welfare, Isolation, and Relative Consumption," *Journal of Development Economics*, 2008, pp. 86(1): 43–60.

Galasso, Emanuela and Martin Ravallion, "Decentralized targeting of an antipoverty program," *Journal of Public economics*, 2005, pp. 89 (4): 705–727.

Ghiglino, C. and S. Goyal, "Keeping Up with the Neighbors: Social Interaction in a Market Economy," *Journal of the European Economic Association*, 2010, *8* (1), 90–119.

Grosh, Margaret and Paul Glewwe, "Designing Household Survey Questionnaires for Developing Countries.," *World Bank Publications, The World Bank*, 2000, p. 25338.

Hussam, Reshmaan, Natalia Rigol, and Benjamin N Roth, "Targeting high ability entrepreneurs using community information: Mechanism design in the field," *American Economic Review*, 2022, *112* (3), 861–98.

Hvidberg, Kristoffer B, Claus Kreiner, and Stefanie Stantcheva, "Social Positions and Fairness Views on Inequality," Technical Report, National Bureau of Economic Research 2020.

Jackson, Matthew, "Social and Economic Networks," *Princeton University Press*, 2010, *Princeton.*

Jiang, Xiaoye, Lek-Heng Lim, Yuan Yao, and Yinyu Ye, "Statistical Rank- ing and Combinatorial Hodge Theory," *Mathematical Programming*, 2011, *127* (1), 203–244.

Krackhardt, D., "Predicting with Networks: Nonparametric Multiple Regression Analysis of Dyadic Data," *Social Networks*, 2022, *10*, 359–81.

Layard, Richard, "Well-Being Measurement and Public Policy," *NBER Chapters, in: Measuring the Subjective Well-Being of Nations: National Accounts of Time Use and Well-Being*, 2009, pp. 145–154.

Martin, K Himelein Ravallion and K Beegle, "Can subjective questions on economic welfare be trusted?," *Economic Development and Cultural Change*, 2016, pp. 64 (4): 697–726.

Newman, M.E.J., "Fast computation of rankings from pairwise comparisons," *arXiv*, 2022, *2207*, 00076.

Ravallion, Martin, "Monitoring targeting performance when decentralized allocations to the poor are unobserved," *The World Bank Economic Review*, 2000, pp. 14 (2): 331–345.

_ , "Miss-targeted or Miss-measured?," *Economics Letters*, 2008, pp. 100 (1): 9–12.

_ , "How relevant is targeting to the success of an antipoverty program?," *World Bank Research Observer*, 2009, pp. 24 (2): 205–231.

_ , "The economics of poverty: History, measurement, and policy," *Oxford University Press*, 2015.

_ **and M Lokshin**, "Identifying welfare effects from subjective questions," *Economica*, 2001, pp. 68 (271): 335–357.

Skoufias, Emmanuel, Benjamin Davis, and Jere R Behrman, "An evaluation of the selection of beneficiary households in the education, health, and nutrition program (PROGRESA) of Mexico," *International Food Policy Research Institute, Washington, DC*, 1999.

Trachtman, Carly, Yudistira Hendra Permana, and Gumilang Aryo Sahadewo, "How much do our neighbors really know? The limits of community-based targeting.," *University of California Berkeley*, 2022.

UN-Habitat, "State of the World's Cities 2006 and 2007.," *Nairobi: United Nations Human Settlements Programme.*, 2006.

Zermelo, E., "Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung," *Mathematische Zeitschrift*, 1929, *29*, 436–460.

Appendices

Appendix A. Estimating the precision of spatial targeting

In this Appendix, we examine the extent to which material well-being in our study area can be predicted on the basis of spatial information. To investigate, we rely on data from an individual survey collected by AUDRI in December 2019 to March 2020. The population covered by the AUDRI survey is comparable to this paper in terms of spatial coverage and sampling. It includes respondents sampled from 622 urban enumeration areas (EA) and 84 peri-urban villages located in and around Abidjan.[40] The median number of respondents is 4-5 per EA (80% of the sample) and 8-9 per rural areas (village), with sizeable variation across sampling units. In the analysis, we combine both sampling units and refer to them as sampling unit (SU).

Our measure of household material well-being is the proxy-means test (PMT) index based on the formula and weights used by the government of Côte d'Ivoire to measure poverty. This index is chosen for two main reasons: it is available for all but one of the 2940 individuals covered in the AUDRI survey; and it is arguably the most reliable measure we can build and the most acceptable to policy-makers in the country. For the purpose of our calculation, we take the median PMT of the sample as poverty cutoff.[41] We estimate the precision of spatial targeting by calculating the proportion of individuals correctly identified as poor or non-poor based purely on the predicted poverty level in their location –i.e., either their sampling unit or their GPS location. For simplicity, we ignore measurement error in the PMT index itself; our focus is on the extent to which a policymaker can rely on spatial targeting to approach the level of poverty targeting comparable to a PMT survey of all households in the Greater Abidjan region.

We start by regressing the PMT index on SU fixed effects and we check which proportion of the respondents in the individual survey would be corrected assigned to a poor or non-poor status if targeting is based on interviewing a small sample of individuals in each SU and relying on the SU fixed effects to identify the poor. Based on this calculation, 69.7% of the poor and 70.0% of the non-poor are correctly assigned to their respective category, leaving 30.3% of poor identified as non-poor and 29.0% of non-poor identified as poor. This simple calculation seems to suggest that a relatively high level of accuracy (70%) could be achieved by surveying a small sample of respondents in each SU and targeting all individuals in the SU's that have a sample average of the PMT index below the cutoff.

This calculation, however, is misleading because the sample mean in each SU suffers from

[40] The list of EA's and the distinction between EA's and villages mirrors the methodology used for the population census by the the National Statistical Institute of Cote d'Ivoire.

[41] This is estimated as follows. Respondents are selected among 15,075 adults living 5,127 households with a total population of 26,101 individuals, children included. From the individual survey, we know that the median income per adult is $54 (Dupas et al. 2021). With approximately three adults per household, this implies a median income per person of $31.5 per month or $1.04 per day, which we use as a conservative value for the poverty cutoff. From this calculation, it follows that the cutoff corresponds roughly to the material welfare of the median PMT index household in our sample.

sampling error, i.e., the SU fixed effects over-fit the small sample in each SU. To investigate this issue, we begin to testing whether the variation in SU fixed effects could entirely be driven by sampling error across SU's. This is achieved by bootstrapping the distribution of estimated SU fixed effects that would be obtained under the null hypothesis of no SU fixed effects. To this effect, we scramble the observed values of the PMT but keep the division of the sample into sampling units of the original size. After scrambling, each SU contains the same number of observations as in the original sample, but these observations have been permutated and come, at random, from other SU's. We then regress the PMT index on SU fixed effects and calculate the standard deviation of the 635 fixed effects.[42] By repeating this process multiple times, we are able to simulate the standard deviation of SU fixed effects under the null of zero systematic variation in the PMT index across SU's. The distribution of the simulated standard deviations is presented in Figure 4, and is to be compared to the observed standard deviation of the SU fixed effects in the original sample, which is 0.3. We see from Figure 4 that the distribution of the simulated standard deviation remains well below 0.3, which indicates that there is systematic variation in PMT index across SU's. In other words, targeting on the basis of SU fixed effects is *a priori* informative. We just do not know how much.

We then simulate the impact of sampling error on targeting by mimicking predicting out of sample. To this effect, we first recover the prediction errors $\hat{e}_i$ from the regression of the PMT index on the SU effects. These errors capture the variation of individual indices among households within each SU. We then scramble the $\hat{e}_i$'s across observation and add these permutated errors to the estimated SU fixed effects to construct an artificial sample of PMT values. This artificial sample can be thought of as representative of other households in the same SU since it suffers from the same amount of prediction error as the sampled individuals. We then calculate the accuracy of SU fixed-effect targeting on this artificial sample. We find that targeting accuracy drops dramatically for out-of-sample households:[43] only 58.6% of the poor are properly assigned to the poor category, while 39.0% of the non-poor are assigned to the poverty status. This demonstrates that interviewing a small number of individuals in each SU's so as to select SU's to receive a poverty intervention would only achieve around 60% targeting accuracy – which non-negligible but only 10% better than randomly targeting half of the SU's.

We also examine whether better targeting can be achieved by using the detailed GPS coordinates of respondents instead of relying on their sampling unit. To this effect, we fit a two-way kernel regression in (decimal) latitude and longitude on the PMT index.[44] The idea behind the approach is that average poverty varies relatively smoothly across space in and around the city. It is intended to capture the way a knowledgeable city planner would form a mental representation of the spatial distribution of affluent and poor neighborhoods – and may target anti-poverty interventions on that basis.

The resulting fit is illustrated in Figure 5, where each color represents a prediction quartile. We see that, as could be anticipated, predicted index values are higher in the vicinity of the city center and lower in the periphery, with a few exceptions corresponding to secondary

[42]71 SU's with a single observation are omitted from this procedure.

[43]The predicted fit varies somewhat from one simulation to another. We present here the result from one representative simulation.

[44]We use the *npregress* command in Stata with an Epanechnikov kernel and an optimal bandwidth.

towns or affluent sub-urban neighborhoods. We expect this map to provide a reasonable match of where an informed Ivorian planner would expect the poor to live. We then compare these spatial predictions to actual PMT index values in our sample, using the same cutoff value as before. We find that this method misses most of the poor: it assigns 21.1% of the poor (and 12.6% of the non-poor) to the poverty category. Using this method would lead to massive under-targeting of anti-poverty interventions. Consistent with this, we observe large wealth differences (as measured in the PPI indexes) across households within EA in the directed graphs, Figures A1 and A2.

From this analysis, we conclude that there is considerable variation in poverty levels within small geographical units, making it difficult for policymakers to effectively reach the poor by targeting on the basis of local averages in our urban setting – irrespective of how these averages are obtained. This means that much room is left to improve targeting by refining poverty information within small sampling units.

Appendix B. Simulation of the performance of reconstructed ranks

To illustrate the effectiveness of the method, we conduct a simulation analysis that loosely mimics our empirical setting. The policy maker wishes to rank a sample of N individuals residing in the same locality. To this effect, B local observers are asked to rank the N individuals by income – which, as discussed above, is equivalent to ranking each pair of individuals in the locality.[45] We set $N = 30$ and $B = 9$. The number of distinct ij pairs is thus $N(N-1)/2 = 435$. True ranks r_{ij} are set by drawing, for each individual i, a log income y_i from a standard normal distribution with unit variance.[46] We generate 100 different localities for each simulation; they can be understood as replications of the data generating process (DGP).

Each local observer only knows a random subset S of the N individuals, with S taking values $\{0.7, 0.5, 0.3, 0.1\}$. For instance, $S = 0.5$ means that an observer asked to rank individuals i and j knows each of them with probability 0.5 – and is thus capable of ranking the ij pair with probability 0.25. Each of the nine observers thus ranks on average a quarter of the individuals pairs in their locality. Since there are $B = 9$ observers, this generates overlap in rankings across ij pairs, making it unlikely that no ranking is reported on any particular pair. In contrast, when $S = 0.1$, each observer ranks any ij pair with a 1 percent probability – i.e., provides on average 4.35 pairwise rankings. The 9 observers thus collectively provide at most 39 distinct rankings on average or 9% if the total number of 435 ij pairs. It follows that, when $S = 0.1$, directly elicited pairwise rankings do not rank all individuals in the locality – and our reconstruction method comes to the fore. We then combine these partial rankings r^k_{ij} to calculate reconstructed ranks P_i as explained above. Un-ranked individuals end up in the middle of the reconstructed distribution with $P_i = n/2$.

Simulation results are presented in Tables A6 and A7. The first row, first column of the

[45]For ease of interpretation, we set the B observers to be distinct from the N target individuals.

[46]The log-normal assumption guarantees that income is positive and mimics the actual distribution of income in many populations.

Table focus on the individual reports r_{ij}^k without observation error. As anticipated, when $S = 70\%$ the proportion of pairs on which we have a report is $49.06\% \approx 0.7^2$. This proportion falls to 24.95% when $S = 50\%$, 8.88% when $S = 30\%$, and 1.02% when $S = 10\%$. There is considerable variation around this average across the 100 simulated localities. In Table S1 the variance of observation errors is 0, which implies that ranked pairs are correctly ranked.

The second parts of Tables A6 and A7 summarize our results regarding our main object of interest, namely, the reconstructed ranks $\hat{r}_i = P_i$. These ranks are constructed by sorting all individuals according to their relative position index P_i: individuals i is estimated to be ranked higher than j if $P_i > P_j$. When reconstructed ranks are not complete, $P_i = P_j$ for some ij and individuals i and j cannot be ranked relative to each other. Having sorted individuals according to their P_i value, we examine what proportion of the *consecutive* pairs are unranked. This measures the completeness of the reconstructed ranks. Since each locality has 30 individuals, there are 29 consecutive sorted pairs $\{\hat{r}_i, \hat{r}_{i+1}\}$. From Table A6 we see that, when $S = 70\%$, the proportion of missing consecutive pairs is small: 0.41%. This proportion increases rapidly as S falls: it rises to 3.93% when $S = 50\%$, 16% when $S = 30\%$ and 60% when $S = 10\%$ (Table A7). We also note the presence of considerable variation across replications, especially at lower values of S. This is because, when S is small, the number of reports is very small – vanishingly so in some replications.

Conditional on a consecutive pair $\{i, i+1\}$ being ranked relative to each other (i.e., $P_i \neq P_{i+1}$), the reconstructed ranking often agrees with the true ranking, in the sense that $r_i > r_{i+1}$ if and only if $P_i > P_{i+1}$. The proportion of correct reconstructed ranks, however, falls rapidly with S. Furthermore, since randomly assigning ranks between two individual pairs results in a correct ranking half of the time, we see that the predictive power of the method falls rapidly with S.

Combining the loss of information due to erroneous and missing rankings, we find that when $S = 70\%$, our method produces a correct ranking for 99.55% of the consecutive pairs. This confirms that the method proposed here can work in the sense of producing a close approximation of the true ranking of target individuals by income level in each given location. The performance of the method, however, deteriorates rapidly when S falls, that is, when observers know a smaller proportion of the target individuals. In particular, when each observer only knows 10% of the target individuals, i.e., 3 individuals on average, the reconstructed rankings only match 25.93% of the true rankings. We also note that there is considerable variation in the performance of the method across replications, suggesting that, in some cases, it can still yield valuable results even when S is low.

Next we investigate the role played by averaging by introducing a observation error term e_i^k to the income of individual i that is observed by k. This error enters the model as $log(y_i) + log(e_i^k)$ where $log(e_i^k)$ is normally distributed with mean 0 and a variance σ_e^2. The variance of $log(y_i) = 1$. In Tables A6 and A7, we then report simulation results for values of σ_e^2 between 0.1 and 0.9 in columns 2-5. The proportion of correctly ranked pairs is close to the theoretical mean of 0.49, with some variation around that mean. The next part of the Table shows the proportion of observer reports that correctly rank ij pairs. As expected, this proportion falls as σ_e^2 increases, with a little variation across locations. But the majority of pairs remain correctly ranked.

The rest of the table shows the effect of observation error on reconstructed ranks P_i. We first look at the proportion of missing reconstructed rankings, that is, the proportion of

ij pairs such that $P_i = P_j$. Equal reconstructed ranks arise for two reasons: (1) because the income of individual i, say, was not observed by any of the B observers and hence no report was given that involves that individual who therefore get $P_i = n/2$; or (2) because the reconstructed graph $\hat{M}$ is either incomplete (e.g., with multiple branches) or contains one or more directed cycles. Simulation results (not shown here) indicate that the first channel has a severe effect only when $S = 10\%$: in this case, around 38-39% of individuals are not ranked at all. But observation error does not affect this proportion since it does not increase the frequency of missing pairwise *reports* r_{ij}^k.[47] Hence the effect of observation error on missing ranks comes entirely from the way it degrades the reconstructed graph. Turning to this channel, we know that when $\hat{M}$ contains two or more branches, it is possible for two individuals to share the same difference $r_i^{down} - r_i^{up} = r_j^{down} - r_j^{up}$ even if $r_i^{down} \neq r_j^{down}$. In this case i and j have the same reconstructed rank P_i even though they sit on separate branches of $\hat{M}$. In addition, when the reconstructed ranking graph contains a directed cycle involving L individuals, all these L individuals share the same values of r^{down} and r^{up} – and thus the same value of the reconstructed index P. Since observation error makes directed cycles more common and can introduce multiple branches in $\hat{M}$, it increases the proportion of missing reconstructed rankings. This effect, however, can be partly compensated by the averaging effect of multiple reports on the same ij pair: even though individual reports may be distorted by error, their average may still be correct since the mean observation error $E[e_i - e_j]$ on $y_i - y_j$ tends to 0 as $kS \to \infty$. Based on this, we expect the mitigating effect to be stronger when S is larger. But it is unclear how much of a mitigating effect averaging has.

We show that the degrading effect of observation error on missing ranks is dramatic: even when $S = 70\%$, the proportion of missing reconstructed ranks rises to 83.48% when $\sigma_e^2 = 0.9$. This arises even though the large value of S implies a relative large number of reports r_{ij}^k on the same ij pair. This means that, within the parameters of our simulation exercise (and the context of our data collection), averaging is unlikely to help much. We also see that there is a lot of variation in the proportion of missing ranks across localities/replications.

Next, the Tables A6 and A7 shows the proportion of correctly reconstructed consecutive ranks, as a fraction of non-missing consecutive ranks. Here the picture is more encouraging: non-missing ranks do, in their majority (76 to 87%), agree with the true ranks. Again this varies a lot across replications. The bottom of the Table combines the two to document the effect of observation error on the proportion of correctly ranked consecutive pairs. The Table shows that, even with $S = 70\%$, this proportion falls rapidly from 74.83 to 12.97% is σ_e^2 increases.

We repeat the same exercise for $S = 50, 30$ and 10%. The same pattern is by and large reproduced: a very gradual decline in the proportion of correct reports; a rapid increase in the proportion of missing reconstructed ranks; and a relative stability of the proportion of correct reconstructed ranks with respect to observation error. All in all, we note a very rapid deterioration in the performance of our rank reconstruction method as observation error increases.

[47]For other values of S in our simulations, the proportion of totally unranked individuals is never more than 3%., irrespective of observation error. The fact that the size of the giant component in the reconstructed graph abruptly falls when S falls below a threshold is a well-known phenomenon in network analysis (e.g., Jackson 2010).

Things are different regarding the proportion of missing ranks: it was already 60% when $\sigma_e^2 = 0$; but it does not increase further with observation error. We also find that the proportion of correct ranks is quite low (its expected value under pure random guess is 50%) – but it is relatively constant. As a result, the proportion of correctly reconstructed ranks under $S = 10\%$ is fairly constant and insensitive to observation error. This arises for the same reason that the reconstructed graph $\hat{M}$ contains a lot of unconnected nodes when $S = 10\%$: being very sparse, the graph $\hat{M}$ contains few branches and nearly no cycles (e.g., Jackson 2010), thereby ruling out the two main sources of missing ranks in the better connected $\hat{M}$ generated when $S \geq 30\%$.

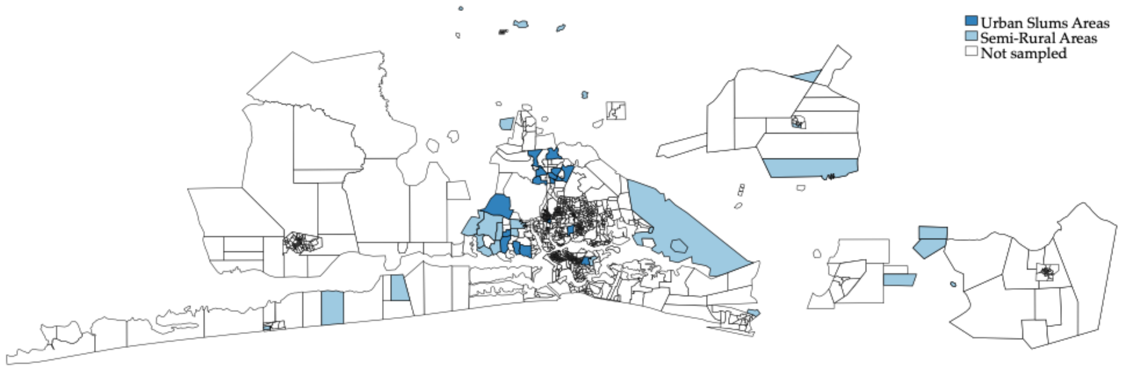
The Tables A6 and A7 have focused on a specific section of the reconstructed ranks, namely, the consecutive pairs. We now broaden our focus to include all ij pairs, whether or not they are consecutive in the reconstructed ranks. We expect reconstructed ranks to be more accurate for ij pairs that are far apart in the true ranks, since the larger difference between their incomes is more likely to survive observation error. Results are presented in Figures 6. Each Sub-Figure shows,for a particular value of S, the proportion of correctly ranked ij pairs as a function of the distance in their true ranks $r_j - r_i$. Each line corresponds to a different value of σ_e^2.

Starting with the top-left sub-Figure, we immediately see that, in the absence of observation error, the constructed P_i index correctly rank ij pairs irrespective of the distance in their true ranks. When observation error is added, we find, as expected, that the P_i index is more likely to correctly rank distant ij pairs than pairs with a more similar income. This contrast remains as σ_e^2 is increased from 0.1 to 0.9, at which point, as we have seen in Table A6, most pairwise ranks are not correctly reconstructed – primarily because of missing ranks. As we move to lower values of S in Figure 6, the proportion of correctly ranked pairs initially falls with S – but then it rises somewhat at high values of σ_e^2, as already observed in the simulation Tables. This may be due to the fact that, with large observation errors, the abundance of reports when $S = 50\%$ increases the likelihood of cycles. With $S = 30\%$, the reconstructed graphs are more sparse and, as a result, are less likely to contain cycles. The bottom right sub-Figure confirms the very different pattern observed in the Tables with $S = 10\%$. Here the likelihood of cycles is low and index P_i is more able to identify correct ranks at high levels of observation error – as already noted.

To summarize, these simulations have demonstrated that index P_i is capable of identifying true ranks with high accuracy in situations where a sufficiently large number of reports are provided by observers, and when these reports are not affected too much by observation error. The performance of the index nonetheless deteriorates when S falls, and especially when σ_e^2 increases. The deterioration in performance due to a lower S arises because of an increase in the proportion of unranked pairs; the deterioration associated with observation error arises because of the creation of branches and, especially, cycles in the reconstructed graphs. It may be possible to improve the performance of the method by penalizing cycle formation, but this is left for future research.

Figures

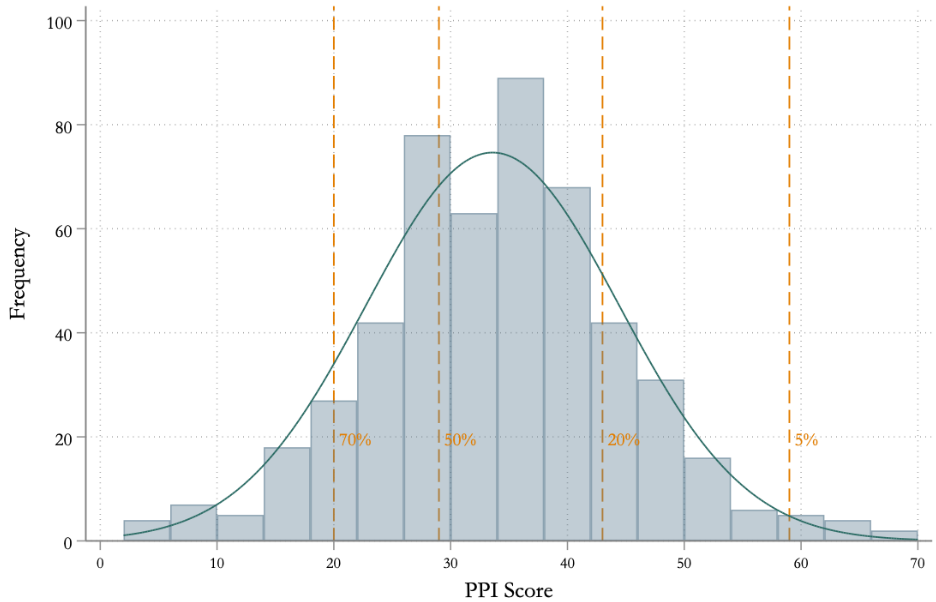
Figure 1: Sampling Areas



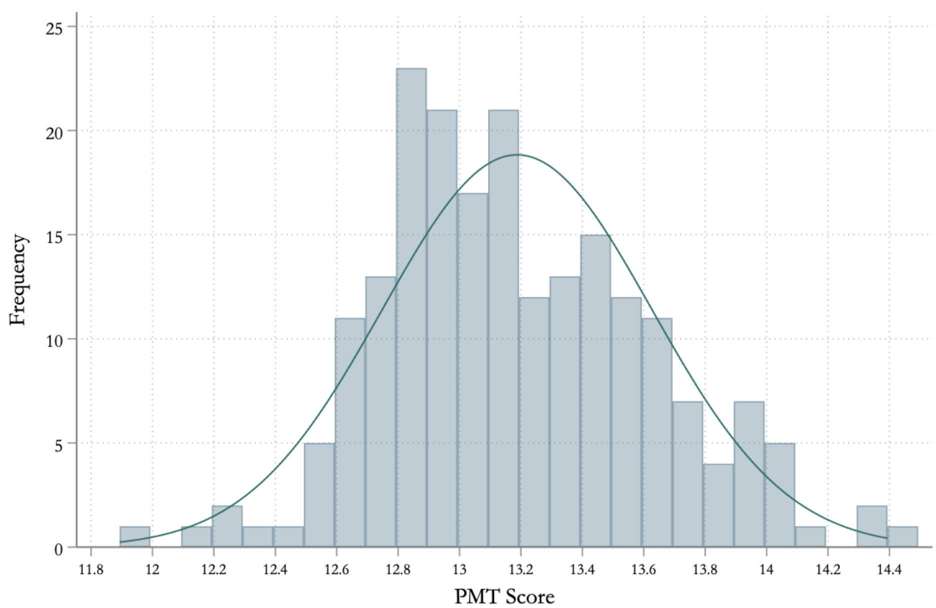
Notes: Enumeration areas selected for the ranking study are indicated in blue.

Figure 2: Poverty Measures Distribution

Poverty Probability Index PPI Distribution in Our Sample

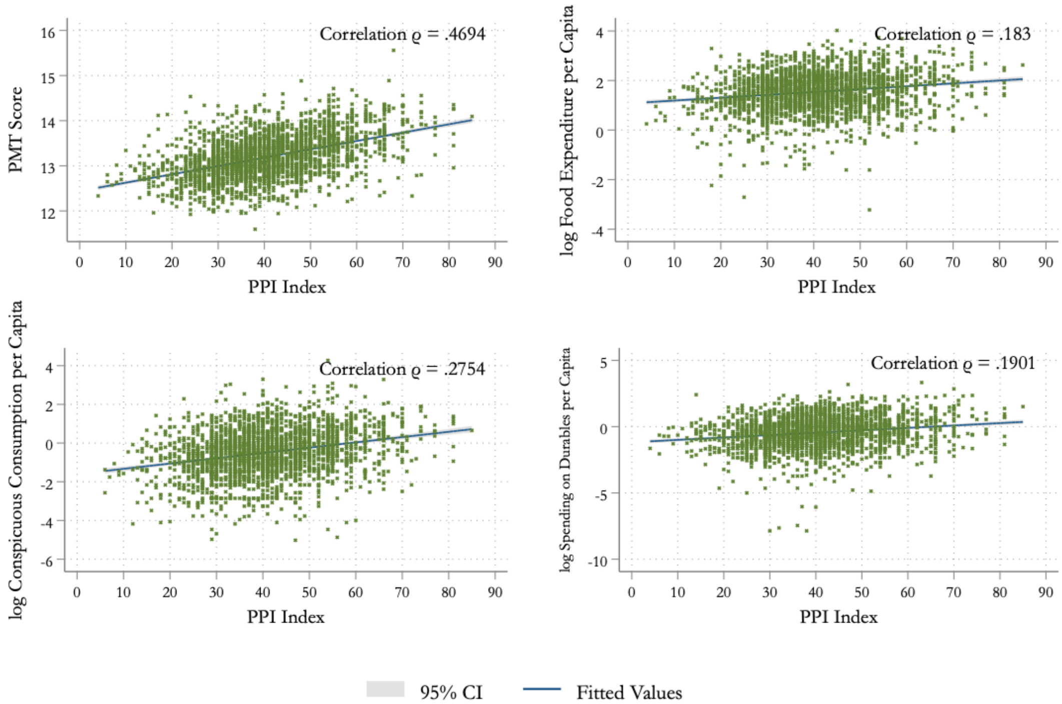


PMT Score in Our Sample



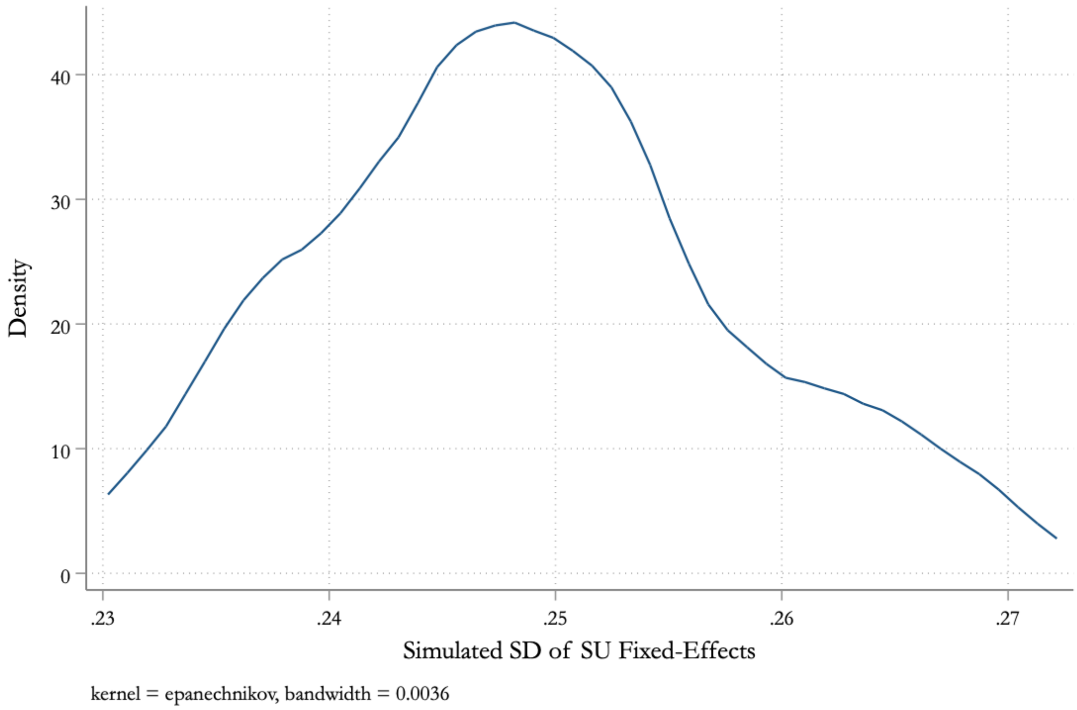
Notes: We plot the distribution of the PPI and PMT indexes. In the top figure, we compute PPI in our sample, as built by Innovations for Poverty Action (IPA) in April 2018 using Côte d'Ivoire's 2015 Enquête sur le Niveau de Vie des Ménages Survey. The "Poverty Likelihood", i.e., the probability to be below the National Poverty Line is indicated in orange. The bottom figure indicates the PMT score developed by the Ivorian Government. The two indexes are described in more details in Section 4.3.

Figure 3: Correlation across Poverty Measures



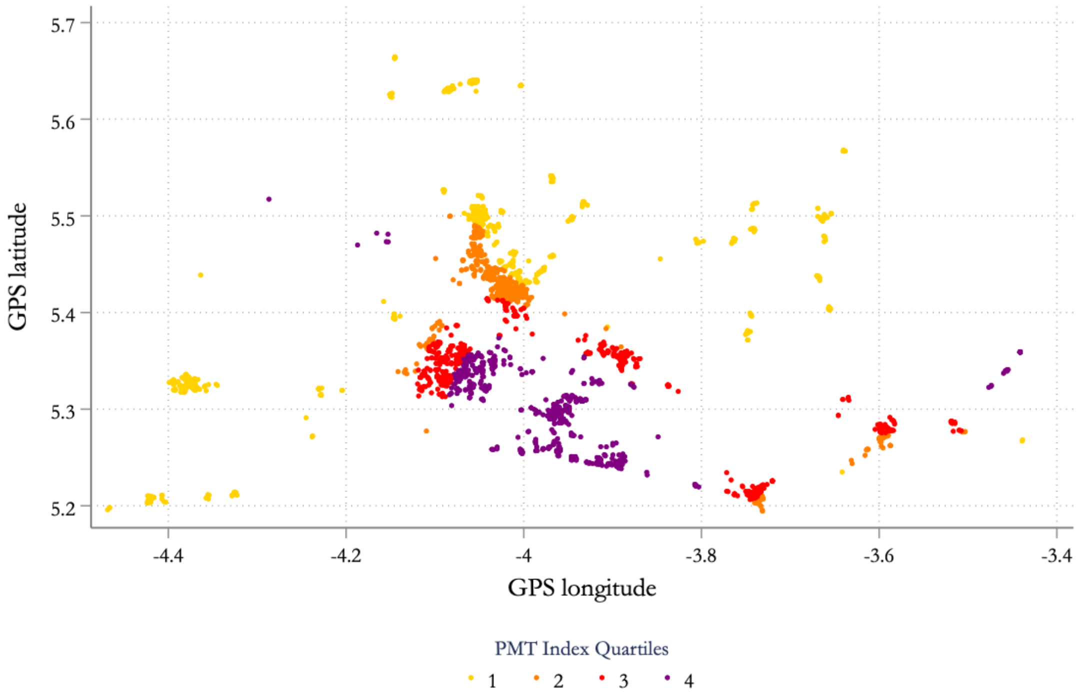
Notes: Correlation across the three measures of poverty as described in Section 4.3. We used the full AUDRI study sample, no matter whether they participated in the ranking exercise or not, in order to maximize the number of observations.

Figure 4: Distribution of the Standard Deviations of SU-Fixed Effects



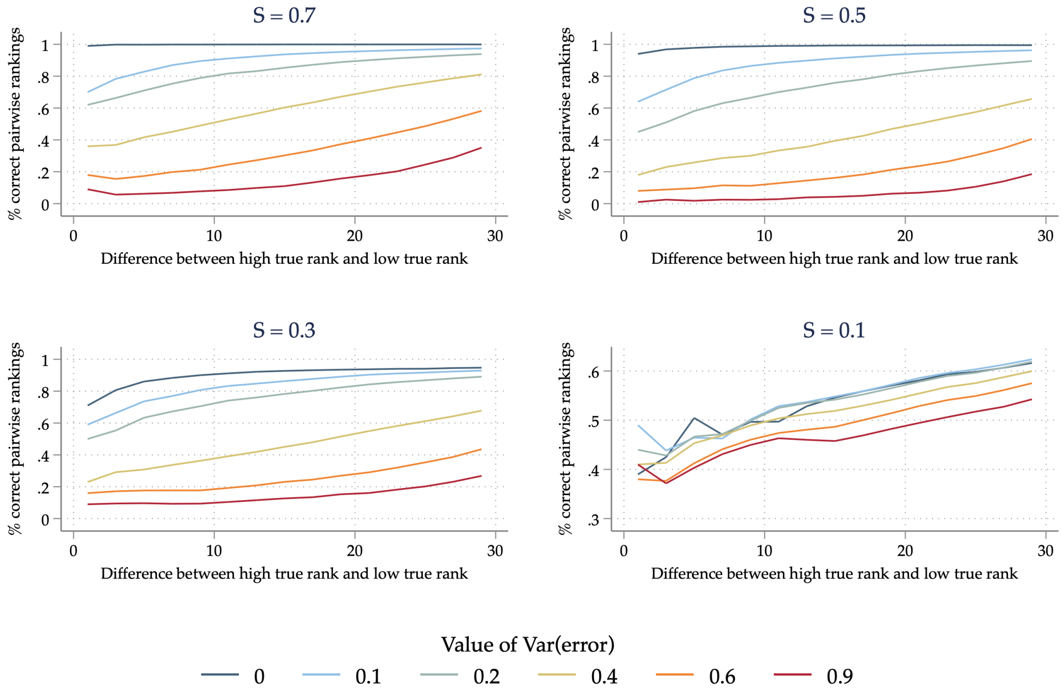
Notes: To obtain the distribution of SD of SU fixed effects, we bootstrap the distribution of estimated SU fixed effects that would be obtained under the null hypothesis of no SU fixed effects. For that, we scramble the observed values of the PMT but keep the division of the sample into sampling units of the original size. We then regress the PMT index on SU fixed effects and calculate the standard deviation of the 635 fixed effects. We are able to simulate the standard deviation of SU fixed effects under the null of zero systematic variation in the PMT index across SU's by repeating this process multiple times.

Figure 5: Spatial Predictions of the PMT Index in and around the City of Abidjan



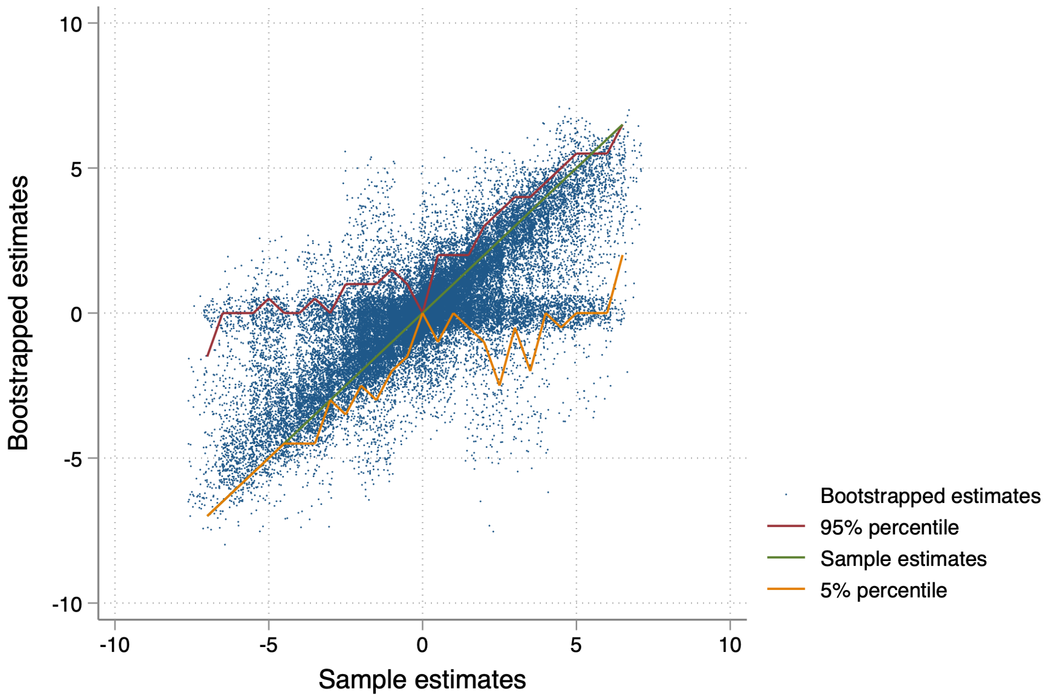
Notes: We fit a two-way kernel regression in (decimal) latitude and longitude on the PMT index. To do so, we use the *npregress* command in Stata with an Epanechnikov kernel and an optimal bandwidth. The idea behind the approach is that average poverty varies relatively smoothly across space in and around the city. It is intended to capture the way a knowledgeable city planner would form a mental representation of the spatial distribution of affluent and poor neighborhoods – and may target anti-poverty interventions on that basis. Each color represents a prediction quartile.

Figure 6: Index Accuracy by Difference in True Ranks



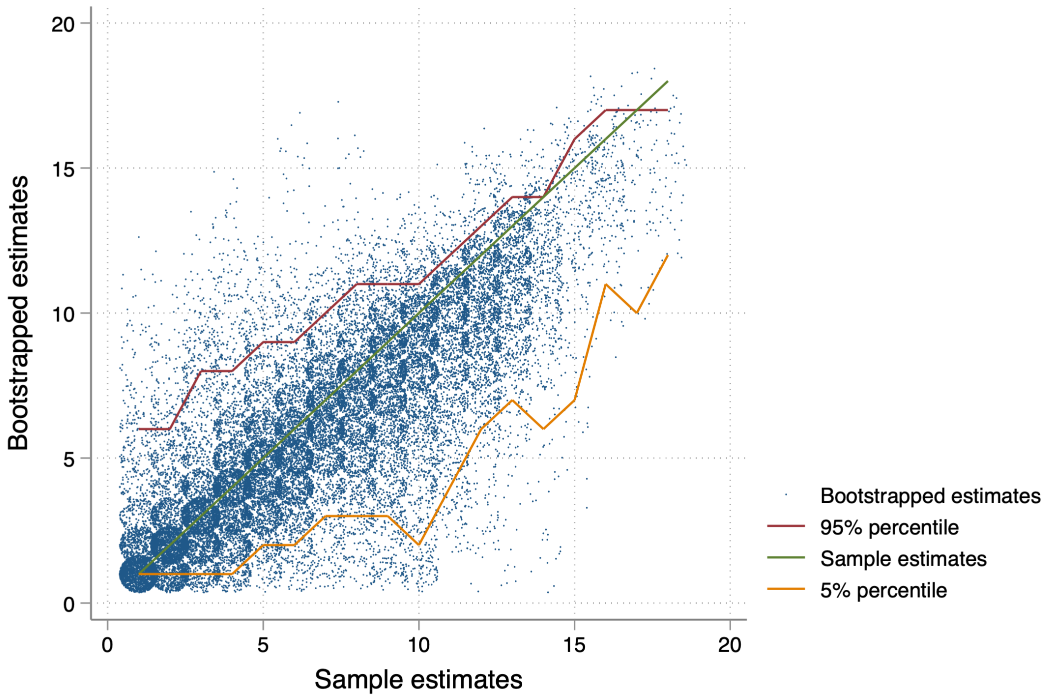
Notes: Each line is a fitted fractional polynomial of % of correctly predicted pairwise comparisons. We show four graphs with different values of S, the random subset the individual knows.

Figure 7: 90% confidence intervals for the estimated index



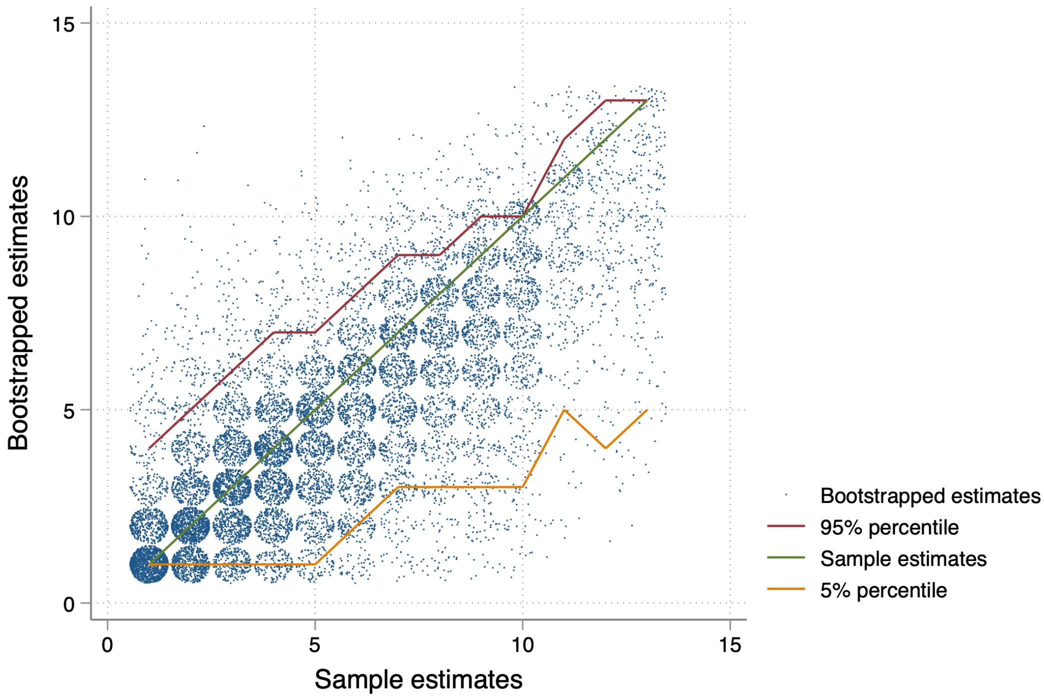
Notes: Bootstrapped estimates have been 'jittered' to show their frequency distribution over 100 replications. Each dot corresponds to an index estimate from a bootstrap sample, conditional on its corresponding sample estimate..

Figure 8: 90% confidence intervals for ranks estimated by the matrix method



Notes: Bootstrapped estimates have been 'jittered' to show their frequency distribution over 100 replications. Each dot corresponds to a rank estimate from a bootstrap sample, conditional on its corresponding sample estimate..

Figure 9: 90% confidence intervals for ranks estimated by the Newman method



Notes: Bootstrapped estimates have been 'jittered' to show their frequency distribution over 100 replications. Each dot corresponds to a rank estimate from a bootstrap sample, conditional on its corresponding sample estimate..

Tables

Table 1: Respondents' Types

Ranking respondent	**#**	**%**	**% of Household Head**	**% of Women**
Households from the listing only - A's	88	17.36 %	38.64%	69.32%
Resp. from the indiv. survey - A's	119	23.47%	47.06%	56.30%
Resp. selected on the spot - B's	230	45.36%	46.52%	46.96%
Key informants - C's	70	13.81%	25.71%	71.43%
Total	**507**	**100%**	**42.41%**	**56.41%**

Table 2: Summary Statistics - Poverty Measures

	(1) Urban	(2) Rural
Consumption Expenditures		
Value of food expenditure in the last week	15.43	13.17
	(8.12)	(7.20)
Value of conspicuous expenditures in the last month	2.47	2.59
	(3.29)	(2.74)
- Communication expenditures	1.09	1.00
	(1.25)	(1.39)
- Entertainment (concert, bar, cinema, games) expenditures	0.22	0.29
	(1.55)	(0.91)
- Beauty products/hairdresser expenditures	0.48	0.75
	(0.79)	(1.40)
- Charitable expenditures	0.67	0.55
	(2.38)	(1.18)
Spending on durables in the last 12 months	2.34	2.10
	(3.00)	(2.38)
- Clothes/shoes HH expenditures	1.10	1.34
	(1.09)	(1.48)
- Furniture HH expenditures	0.39	0.27
	(1.16)	(0.84)
- School fees HH expenditures	0.89	0.52
	(2.80)	(1.35)
Value of food expenditure in the last week per capita	3.76	3.74
	(2.89)	(3.39)
Spending on durables in the last 12 months per capita	0.53	0.61
	(0.67)	(1.46)
Indexes		
PPI Index	37.13	28.70
	(9.55)	(10.64)
Score PMT	13.23	13.13
	(0.43)	(0.44)
Other variables		
HH's head unemployed or inactive	0.20	0.15
	(0.40)	(0.36)
# of mobile phones per capita	0.81	0.82
	(0.47)	(0.48)
Observations	294	213

Notes: Consumption expenditures are *all* normalized per week and in 1,000 FCFA

Table 3: How do Respondents Think About Poverty? Summary Statistics from Survey Data

	(1) Share of respondents
Uncertain about their ranking	0.09
	(0.29)
Own poverty's perceptions	
Consider their household to be poor	0.29
	(0.45)
Consider their household to be poorer than neighbors	0.21
	(0.41)
Think that other households consider their household to be poor	0.21
	(0.41)
Criteria used to classify	
Household expressed their financial problems	0.49
	(0.50)
Household members' health	0.14
	(0.35)
Household head's occupation	0.49
	(0.50)
Households' daily number of meals	0.19
	(0.39)
Household children's school enrollment	0.07
	(0.26)
Respondents' own definition of poverty	
Food deprivations	0.80
	(0.40)
No decent housing	0.31
	(0.46)
Unresolved health problems	0.43
	(0.49)
No proper toilet/bathroom	0.16
	(0.37)
Knowledge about neighbors	
# of neighbors listed in total	5.75
	(1.47)
% of neighbors they regularly visit	0.56
	(0.38)
% of neighbors receive health/money advice from	0.44
	(0.39)
% of neighbors they'd ask money from	0.38
	(0.38)
Observations	507

Notes: Survey data collected early March 2020. Definition of poverty manually entered by the enumerators and re-classified by the research team.

Table 4: Relative Rank vs. Relative Survey-Measured Poverty: Pair-Level

Difference j - i in:	(1) Food exp per ca	(2) Months food short	(3) PMT Score	(4) PPI Index
Reported rank	0.219	-0.485	0.078	1.165
	(0.355)	(0.347)	(0.050)	(0.996)
log(distance)	0.070	0.051	0.014	0.017
	(0.072)	(0.078)	(0.012)	(0.244)
Constant	-0.077	0.119	-0.125*	-1.223
	(0.481)	(0.463)	(0.067)	(1.349)
R2	0.003	0.004	0.009	0.005
Observations	691	645	416	751
Constructed rank	-0.243	-0.843***	0.145***	0.794
	(0.283)	(0.315)	(0.046)	(0.819)
log(distance)	0.009	-0.000	0.007	0.068
	(0.063)	(0.065)	(0.011)	(0.209)
Constant	0.394	0.753*	-0.142**	-1.494
	(0.416)	(0.424)	(0.064)	(1.171)
R2	0.001	0.009	0.019	0.006
Observations	1014	951	570	1099
Diff in HodgeRank	0.101	-1.190***	0.135*	2.642**
between j and i	(0.421)	(0.407)	(0.071)	(1.174)
log(distance)	0.014	-0.001	-0.001	-0.135*
	(0.024)	(0.021)	(0.006)	(0.073)
Constant	0.173	0.419***	-0.024	-1.056**
	(0.163)	(0.148)	(0.035)	(0.470)
R2	0.000	0.003	0.003	0.002
Observations	4579	4436	1625	5590

Notes: We report three separate regressions for each constructed variable of aggregated rankings, as described in Section 6. The reported rank variable is the share of reported ranks showing j richer than i. The constructed rank variable is a dummy equals to 1 if j is ranked richer than i (including the case where it is both). The difference in HodgeRank scores between j and i is used as an independent variable: the higher the difference, the richer j is ranked compared to i, following the HodgeRank algorithm described in part 2. A negative difference would indicate that j is ranked poorer than i. All dependent variables are differences in consumption variables. They are calculated as the value for household j minus the value for household i. The complete list of dependent variables is given below. For consumption variables, a positive difference means that i is poorer than j. We also display the PMT and the PPI indexes. Robust standard errors are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All dependent variables are the difference in values between households i and j and described the following:
Food expenditures: Total of consumption expenditures collected with a one week recall period: staples, meat, vegetables, fruits, drinks, alcohol. *Months food short:* Number of months the household experienced a food shortage over the last twelve months. PPI and PMT Scores are wealth measures, computed following the methodology described in Section 4.3.

Table 5: Aggregate Rank vs. Survey-Measured Poverty Level : Individual Level

Levels of:	(1) Food exp per ca	(2) Months food short	(3) PMT Score	(4) PPI Index
Relative Position	0.005	-0.024	0.010	0.017
	(0.037)	(0.039)	(0.007)	(0.127)
log(distance)	-0.023	-0.050	0.011	-0.210
	(0.075)	(0.080)	(0.015)	(0.295)
Constant	4.095***	1.835***	13.097***	35.134***
	(0.482)	(0.471)	(0.084)	(1.686)
R2	0.001	0.003	0.049	0.003
Observations	291	282	167	291

HodgeRank score	0.367	-0.959	0.246	1.867
	(1.535)	(1.202)	(0.182)	(4.948)
log(distance)	-0.065	-0.078	0.013	-0.389*
	(0.054)	(0.056)	(0.013)	(0.210)
Constant	4.084***	1.949***	13.071***	35.664***
	(0.355)	(0.357)	(0.077)	(1.285)
R2	0.004	0.005	0.049	0.008
Observations	474	464	207	474

Notes: We report two separate regressions for each constructed variable of aggregated rankings, as described in Section 6. Only the surveyed respondents for which we recovered a rank are included in this regression. The constructed rank difference is the difference between the number of nodes that, in the constructed ranking graph, i looks down towards, and the number of nodes that i looks up towards. The difference is 0 when everyone is in a circle, meaning no one is ranked above anyone else. The HodgeRank scores follow the HodgeRank algorithm described in part 2. A high score means the individual is ranked richer.
All dependent variables are levels of consumption variables. The complete list of dependent variables is given below. Robust standard errors are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Outcomes: *Food expenditures:* Total of consumption expenditures collected with a one week recall period: staples, meat, vegetables, fruits, drinks, alcohol. *Months food short:* Number of months the household experienced a food shortage over the last twelve months. PPI and PMT Scores are wealth measures, computed following the methodology described in Section 4.3.

Table 6: Predictors of rankings: Pairwise comparisons

	(1)	(2)	(3)
	Dep. var. =1 if resp. k reported that i is poorer than j		
	OLS	OLS	OLS
Indep. Variables: Differences between j and i in:			
PPI Index	0.002	0.001	0.003**
	(0.001)	(0.001)	(0.001)
HH's head unemployed or inactive	-0.035	-0.038	-0.033
	(0.029)	(0.029)	(0.031)
Household Size	0.005	0.013**	0.005
	(0.004)	(0.005)	(0.005)
Value of food expenditure in the last week per capita		0.010***	
		(0.004)	
Value of conspicuous consumption expenditures in the last month		-0.006	
		(0.004)	
Spending on durables in the last 12mo per capita		0.018	
		(0.019)	
Received gifted food last week (yes=1)			-0.118***
			(0.043)
Food worries during last 12mo (yes=1)			-0.044
			(0.034)
Months with food shortages last 12mo			-0.004
			(0.006)
Days with skipped meals last 3mo			0.000
			(0.001)
Improvement in food situation last year (1 to 5)			-0.029*
			(0.017)
log(distance from k to i)	0.011	0.012	0.004
	(0.011)	(0.011)	(0.011)
Semi-Rural EA	-0.006	0.010	-0.021
	(0.040)	(0.041)	(0.044)
Constant	0.515***	0.490***	0.537***
	(0.047)	(0.047)	(0.048)
R2	0.009	0.020	0.030
Observations	887	887	813

Notes: The unit of observation at the triad level. The outcome variable is a dummy equal to 1 if k ranked j poorer than i, 0 otherwise. It is missing if no ranked were assigned. Pairs i-j involving the respondent k are dropped. The three columnns contain different types of predictors, e.g., assets and expenditures, and assets and experienced poverty. The predictors are all differences between j and i. Missing distance is replaced by the average distance in the EA and we control for such a case in the regressions. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Determinants of Ranking Accuracy

	(1)	(2)	(3)
	Ranking Accuracy compared to survey measure:		
	PMT	PPI	Food expenditure per capita
Respondent: Woman	0.044	-0.079	0.091*
	(0.056)	(0.054)	(0.052)
Respondent: Migrant	0.022	0.003	-0.068
	(0.059)	(0.057)	(0.055)
Respondent: Non-Ivorian	0.022	-0.027	-0.012
	(0.060)	(0.059)	(0.054)
Respondent: Key Informant	0.063	0.080	-0.062
	(0.085)	(0.095)	(0.065)
Respondent: Household Head	0.036	-0.025	0.092
	(0.057)	(0.056)	(0.056)
Semi-Rural EA	-0.035	0.041	0.037
	(0.055)	(0.051)	(0.052)
PPI Index	-0.000	0.004*	0.002
	(0.002)	(0.002)	(0.002)
Value of food expenditure in the last week per capita	0.014*	0.002	0.014**
	(0.008)	(0.008)	(0.007)
Asked to self rank	0.037	0.016	0.047
	(0.045)	(0.044)	(0.041)
# of neighbors listed in total	-0.008	0.013	0.002
	(0.012)	(0.011)	(0.010)
Household Size	0.014*	0.001	0.026***
	(0.008)	(0.008)	(0.007)
Belong any community group	0.003	-0.039	0.009
	(0.045)	(0.046)	(0.043)
Constant	0.402***	0.330**	0.139
	(0.145)	(0.137)	(0.123)
R2	0.026	0.037	0.070
Observations	285	278	285
Sample Mean of Ranking Accuracy	0.535	0.521	0.511

Notes: Regression of the propensity to rank accurately two respondents according to their respective indexes. The number of observations is the number of respondents for whom we could obtain the accuracy measures (i.e., they ranked enough neighbors within our sample). The number of observations for the PPI is lower since accuracy is missing when two ranked neighbors had the exact same index (which happened more often for PPI than for the PMT/food expenditure that are more precise indexes). Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. M

Table 8: Poverty Targeting: Can peer rankings identify those below the median?

	Dep. Var: Ranked below the median according to:		
	PMT	PPI	Food Expenditure per capita
Panel A. Indep. Var: Ranked below the median (aggregate ranking)			
Below median	0.105	0.019	-0.006
	(0.081)	(0.058)	(0.058)
Constant	0.395***	0.411***	0.453***
	(0.046)	(0.036)	(0.036)
Observations	172	311	311
Panel B. Indep. Var: Ranked below the median (HodgeRank score)			
Below median	0.194***	0.022	-0.012
	(0.073)	(0.051)	(0.051)
Constant	0.372***	0.427***	0.484***
	(0.046)	(0.033)	(0.033)
Observations	189	390	390
Panel C. Indep. Var: Unranked			
Unranked	0.167	-0.023	0.035
	(0.121)	(0.052)	(0.053)
Constant	0.444***	0.459***	0.469***
	(0.036)	(0.025)	(0.025)
Observations	207	507	507

Notes: In the first two panels, we create a dummy equal to 1 if the aggregate ranking puts the household below the median of its EA. It is zero if the household is ranked at or above the median and missing if the household is not ranked. We use two separate constructed variable of aggregated rankings: (1) the relative position of individual i in the constructed network, i.e., how many people can be ranked as poorer than i and subtracts how many can be ranked richer. (2) the HodgeRank algorithm assigns a score to each households, described in part 2. Only the individuals ranked by at least one other respondent are considered.
We run OLS regressions of the dummy on a dummy for whether the household is below the median based on the survey measure (PMT in column 1, PPI in column 2, and the food expenditure per capita in column 3). The table reads as follows: individuals ranked below the median in the aggregate peer ranking are 10.1 percentage points more likely to be indeed below the median of the PMT score distribution (Column 1). In the bottom panel, the dependent variable is a dummy equal to 1 if the household was ranked by no one.
Robust standard errors are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: Poverty Targeting: Predict being ranked below median- No Self-Rank

	Dep. Var: Ranked below the median according to:		
	PMT	PPI	Food Expenditure per capita
Panel A. Indep. Var: Ranked below the median (aggregate ranking)			
Below median	0.127	0.101	0.020
	(0.081)	(0.067)	(0.067)
Constant	0.406***	0.410***	0.458***
	(0.049)	(0.041)	(0.042)
Observations	161	236	236
Panel B. Indep. Var: Ranked below the median (HodgeRank score)			
Below median	0.169**	0.058	-0.007
	(0.079)	(0.065)	(0.066)
Constant	0.383***	0.423***	0.469***
	(0.050)	(0.044)	(0.044)
Observations	161	236	236
Panel C. Indep. Var: Unranked			
Unranked	0.081	0.048	0.061
	(0.084)	(0.044)	(0.044)
Constant	0.441***	0.428***	0.445***
	(0.039)	(0.032)	(0.032)
Observations	207	507	507

Notes: See Table 8 notes.
Robust standard errors are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 10: Testing for Self-Ranking Bias

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Bottom Cons	Bottom PMT	Bottom PPI	Men	Women
S_{ij}^{k}	-0.281***	-0.283***	-0.262***	-0.594***	-0.245***	-0.202***	-0.306***
	(0.031)	(0.031)	(0.050)	(0.081)	(0.048)	(0.064)	(0.042)
Constant	0.518***	0.515***	0.535***	0.531***	0.531***	0.500***	0.532***
	(0.004)	(0.004)	(0.005)	(0.002)	(0.007)	(0.008)	(0.006)
R2	0.136	0.139	0.112	0.450	0.112	0.078	0.159
Observations	1298	1337	680	313	732	598	739
Number of ij pairs	704	732	443	244	479	421	459

Notes: The dependent variable is 1 if the respondent k reports that i is poorer than j, 0 if i is richer, and missing if k does not rank i and j. Variable S_{ij}^{k} is 1 if k=i and -1 if k=j, and 0 if k is not i or j. Column 1 only uses the 30 EAs without sampling issues / Column 2 uses all 34 EAs. The other columns are restricted to respondent k in top and bottom 50% of different wealth measures. Observations from the no-self-ranking treatment are omitted since they contain no useful information. Including them anyway produces identical results. Fixed effects are include for each (i,j) pair. Robust standard errors are provided in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 11: Predictors of propensity to rank others

	(1) OLS	(2) OLS
log(distance from k to i)	-0.017***	-0.017***
	(0.001)	(0.001)
Semi-Rural EA	-0.036***	-0.044***
	(0.008)	(0.008)
Respondent k: Key Informant	-0.048***	-0.048***
	(0.010)	(0.010)
Value for i of the following var:		
PPI Index	-0.002***	-0.001**
	(0.000)	(0.001)
Household Size	-0.003*	-0.003
	(0.002)	(0.002)
HH's head unemployed or inactive	0.013	0.023*
	(0.012)	(0.012)
Respondent: Non-Ivorian	0.010	0.026*
	(0.013)	(0.014)
Respondent: Migrant	-0.020	-0.022*
	(0.013)	(0.013)
Respondent: Woman	-0.014	-0.011
	(0.010)	(0.010)
Received gifted food last week (yes=1)		0.007
		(0.016)
Food worries during last 12mo (yes=1)		-0.018*
		(0.010)
Value of food expenditure in the last week per capita		-0.001
		(0.002)
Value of conspicuous expenditures in the last month		0.005***
		(0.002)
Spending on durables in the last 12 months per capita		0.002
		(0.005)
Value for i - Value for k of the following var:		
PPI Index	0.001***	0.002***
	(0.000)	(0.000)
Household Size	0.002*	0.001
	(0.001)	(0.001)
HH's head unemployed or inactive	0.004	-0.001
	(0.009)	(0.009)
Respondent: Non-Ivorian	0.006	-0.001
	(0.009)	(0.010)
Respondent: Migrant	0.005	0.008
	(0.009)	(0.009)
Respondent: Woman	0.009	0.004
	(0.007)	(0.007)
Received gifted food last week (yes=1)		0.014
		(0.012)
Food worries during last 12mo (yes=1)		0.017**
		(0.007)
Value of food expenditure in the last week per capita		-0.002
		(0.001)
Value of conspicuous consumption expenditures in the last month		-0.001
		(0.001)
Spending on durables in the last 12mo per capita		-0.001
		(0.004)
Constant	0.290***	0.269***
	(0.024)	(0.026)
R2	0.042	0.050
Observations	6835	6620

Notes: The unit of observation at the dyad level. The outcome variable is a dummy equal to 1 if k report a ranking for the individual i, 0 otherwise. Pairs i-k involving the respondent k are dropped. Missing distance is replaced by the average distance in the EA and we control for such a case in the regressions. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

ONLINE APPENDIX

Table A1: PPI – Scorecard Côte d'Ivoire 2015 National Poverty Line

Indicators	Responses	Points (National Poverty Line)
1. In which district does this household reside?	A. Abidjan	7
	B. Yamoussoukro	5
	C. Bas-Sassandra	9
	D. Comoé	4
	E. Denguélé	0
	F. Gôh-Djiboua	3
	G. Lacs	3
	H. Lagunes	2
	I. Montagnes	5
	J. Marahoué	0
	K. Savanes	2
	L. Vallée du Bandama	2
	M. Woroba	4
	N. Zanzan	4
2. How many members does the household have?	A. Three or less	17
	B. Four or more	0
3. What is the highest educational level that the household head has completed?	A. None	0
	B. Primary	4
	C. Secondary	5
	D. Higher	12
4. Did all children aged 6 to 16 attend school this school year?	A. There are no children aged 6 to 16	11
	B. All children aged 6 to 16 attended school this year	7
	C. At least one child aged 6 to 16 did not attend school this year	0
5. What is the mode of water supply?	A. Tap water in the dwelling	10
	B. Tap water in the yard	4
	C. Tap water outside of the property	4
	D. Well in the yard	1
	E. Public well	2
	F. Village pump	2
	G. Surface water (creek, river, etc.) or other	0
6. What type of toilet do you use?	A. W-C inside	7
	B. W-C outside	6
	C. Latrines in the yard	5
	D. Latrines out of the yard	5
	E. In nature (no toilet) or other	0
7. Where do you take your shower?	A. Outside	0
	B. Rudimentary shower	3
	C. Bathroom	9
	D. Other	1
8. Did the household own a moped, car or van in good working order in the last 3 months?	A. The household owns a car or van	15
	B. The household owns a moped and does not own a car or van	9
	C. None	0
9. Did the household own a fan in good working order in the last 3 months?	A. Yes	6
	B. No	0
10. Did the household own a bed in good working order in the last 3 months?	A. Yes	4
	B. No	0
	PPI Score	SUM

Table A2: Poverty Predictions Models: Proxy-Means Test (PMT) and Poverty Probability Index (PPI)

	PMT Score					PPI Index	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log(per capita conso)	Rural - Gov	Urban - Gov	Tot - Gov	Tot - Ranking	Tot - AUDRI	Tot - Ranking	Tot - AUDRI
R^2	0.497	0.612	0.568	0.563	0.491	0.484	0.446
Observations	7,076	5,748	12,773	193	2,871	493	2,666

The table reports the R^2 and the number of observations from the regressions run by the government of Côte d'Ivoire to build their PMT score (columns 1, 2, 3). The numbers were shared to us by the CNAM in Côte d'Ivoire. The government regressed log(food expenditure per capita) on the variables used to build the PMT score. In Column 4, we report the R^2 from the same regression run on the data for households involved in the ranking exercise. Column (5) reports the R^2 from the same regression run on the data for households involved in the full AUDRI sample. Column (6) reports the fit from the PPI regression, i.e., regressing log(food consumption per capita) on the variables used to build the PPI index. Column (7) reports the latter PPI regression on the full AUDRI sample. Note that the sample size is not exactly the same between columns (5) and (7) due to differential missing patterns between variables used in the PMT vs. the PPI score.

Table A3: Summary Statistics Across EAs

	(1)	
# of intended targets	14.00	
	(0.00)	
# of observers	8.65	
	(4.26)	
Coverage overlap between observers >= 2	0.31	
	(0.23)	
Coverage overlap between observers = 1	0.15	
	(0.16)	
Coverage overlap between observers = 0	0.54	
	(0.23)	
Full Agreement on the Pairs i-j	0.24	
	(0.43)	
Observations (EAs)	34	1

Notes: The extent of overlap in coverage between observers states that the share of the observers that were ranked by at least 2 observers, 1, or none. Full agreement across observers report, among all reported pairwise ranks by individual observers, the share where the proportion of responses that state that i¡j is equal to 0 or 1..

Table A4: Reconstructed aggregate rankings

Number of ranked households who are richer or poorer than the target household

EA 2	richer	poorer	EA 4	richer	poorer	EA 6	richer	poorer	EA 7	richer	poorer	EA 8	richer	poorer	EA 9	richer	poorer	EA 12	richer	poorer
201	2	11	201	0	9	201	10	10	201	10	2	201	0	5	205	4	4	203	4	1
202	1	12	202	5	3	202	13	0	202	10	11	203	1	5	208	0	6	204	1	2
203	0	13	203	3	5	203	10	10	203	10	11	207	1	5	209	5	0	205	5	0
204	10	10	204	2	7	206	10	10	204	10	11	209	2	5	210	3	5	305	0	3
205	3	10	205	5	0	207	12	2	205	11	1	211	12	5	211	2	2	307	0	2
206	10	10	206	6	0	208	11	3	206	10	11	212	11	2	212	1	7	308	0	2
208	10	10	207	1	8	209	13	0	207	10	11	213	10	3	213	8	1			
209	10	10	209	7	0	210	10	10	208	1	11	214	8	4	301	1	0			
210	12	1	210	6	1	211	2	11	212	12	0	301	0	6	303	0	1			
211	14	0	211	3	5	302	10	10	302	10	11	302	0	6	304	6	2			
212	10	5				303	0	10	303	10	0	304	13	0	306	1	3			
213	11	3				304	10	4	304	10	11	306	0	4	307	9	0			
214	11	3				308	1	12	305	0	12	307	0	6	901	0	9			
301	0	10				309	3	0	903	0	11	308	0	6						
302	0	10																		
902	13	0																		

EA 14	richer	poorer	EA 19	richer	poorer	EA 20	richer	poorer	EA 21	richer	poorer	EA 30	richer	poorer	EA 31	richer	poorer	EA 32	richer	poorer
202	5	3	201	0	8	201	6	1	202	7	2	201	0	10	201	1	9	201	12	0
203	5	7	205	1	3	203	7	0	203	1	3	203	13	10	202	12	0	202	9	1
207	9	0	206	2	2	204	6	2	204	0	4	204	13	10	203	6	5	203	0	7
209	7	1	208	4	0	205	6	12	205	0	6	205	13	10	204	11	4	205	7	7
210	6	2	211	1	2	206	0	12	207	3	9	206	0	11	205	7	1	206	7	7
211	5	7	212	3	0	207	6	12	208	10	1	207	13	10	206	0	8	207	7	1
214	0	1	213	5	1	208	0	12	210	4	5	208	13	10	208	4	6	209	7	7
303	0	7	214	0	3	209	6	12	211	5	2	209	13	10	211	3	7	214	1	1
304	5	7	302	3	1	210	6	12	301	3	9	210	13	10	212	11	4	302	0	7
305	0	7	303	6	0	214	7	0	302	5	2	211	1	10	214	11	4	303	1	2
304	2	3	304	2	3	301	6	1	303	3	9	212	13	10	302	0	4	304	7	7
902	0	4	902	0	4	302	6	0	304	11	0	213	13	10	305	2	6	306	0	7
						303	7	0				214	13	10	902	0	10	903	0	4
						902	7	0												

This Table reports the reconstructed rankings for the most informative enumeration areas (EAs). In each EA column appears the id code of the household in that enumeration area. Numbers from 201 to 214 represent individuals from the individual or listing surveys. Numbers above 301 were given to respondents added on the spot. Numbers from 901 and above are key informants who appear in this Table because they self-ranked.

Table A5: Relative Rank vs. Relative Survey-Measured Poverty: Pair-Level

Difference j - i in:	Social exp per ca	Durables per ca	Total cons per ca	Was given food (yes)	Food worries (yes)	Days skipped	Improvement in food
Reported rank	-0.345*	-0.149	-0.275	-0.064**	0.006	0.496	0.042
	(0.183)	(0.105)	(0.514)	(0.032)	(0.053)	(1.452)	(0.090)
log(distance)	0.021	0.002	0.094	-0.002	-0.003	-0.535*	-0.014
	(0.039)	(0.024)	(0.107)	(0.007)	(0.013)	(0.310)	(0.021)
Constant	0.222	0.224*	0.369	0.008	0.003	1.661	0.044
	(0.265)	(0.124)	(0.692)	(0.044)	(0.074)	(1.790)	(0.118)
R2	0.049	0.006	0.006	0.006	0.000	0.011	0.001
Observations	691	691	691	650	650	639	650
Constructed rank	-0.388**	-0.186**	-0.816*	-0.081***	-0.006	-1.200	0.154**
	(0.158)	(0.085)	(0.420)	(0.027)	(0.045)	(1.267)	(0.076)
log(distance)	-0.002	0.004	0.012	-0.003	-0.015	-1.181***	-0.014
	(0.034)	(0.021)	(0.095)	(0.006)	(0.011)	(0.404)	(0.018)
Constant	0.243	0.169	0.807	0.029	0.094	5.789***	-0.056
	(0.232)	(0.114)	(0.608)	(0.038)	(0.066)	(2.212)	(0.103)
R2	0.022	0.007	0.006	0.012	0.002	0.022	0.005
Observations	1014	1014	1014	960	960	945	960
Diff in HodgeRank between j and i	-0.827***	-0.290***	-1.016*	-0.173***	-0.051	-0.188	0.126
	(0.209)	(0.094)	(0.579)	(0.040)	(0.061)	(1.326)	(0.102)
log(distance)	-0.001	0.011	0.024	0.006**	-0.004	-0.409***	-0.000
	(0.012)	(0.009)	(0.037)	(0.003)	(0.004)	(0.117)	(0.006)
Constant	-0.025	-0.013	0.135	-0.040**	0.053**	1.826**	0.011
	(0.085)	(0.054)	(0.246)	(0.017)	(0.024)	(0.729)	(0.039)
R2	0.004	0.001	0.001	0.006	0.004	0.004	0.002
Observations	4579	4579	4579	4450	4450	4422	4450

Notes: We report three separate regressions for each constructed variable of aggregated rankings, as described in Section 6. The reported rank variable is the share of reported ranks showing j richer than i. The constructed rank variable is a dummy equals to 1 if j is ranked richer than i (including the case where it is both). The difference in HodgeRank scores between j and i is used as an independent variable: the higher the difference, the richer j is ranked compared to i, following the HodgeRank algorithm described in part 2. A negative difference would indicate that j is ranked poorer than i. All dependent variables are differences in consumption variables. They are calculated as the value for household j minus the value for household i. The complete list of dependent variables is given below. For consumption variables, a positive difference means that i is poorer than j. We also display the PMT and the PPI indexes. Robust standard errors are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All dependent variables are the difference in values between households i and j and described the following:

Social expenditures Total of consumption expenditures collected with a one month recall period: telecom, beauty products, entertainment, charitable contributions. *Annual expenditures:* Total of consumtion expenditures collected with a one year recall period: shoes and clothing, furniture, school fees. *Total consumption:* Weekly expenditures x 52 + monthly expenditures x 12 + annual expenditures. We then divide by the number of household members (adults and children). *Was given food (yes):* Dummy equal to 1 if members of the household have received free food from other households or organizations. *Food worries (yes):* Dummy equal to 1 if respondents answers yes to question . *Days skipped:* Number of days with skipped meals over the last three months. *Improvement in food:* Likert scale from 1 (much worse) to 5 (much better) on whether food situation of the respondentś household has improved relative to previous year.

Table A6: Performance of the method - Simulations (1)

$S = 70\%$	$Var(error)$				
	0	0.1	0.2	0.6	0.9
Pairwise rankings $r_i jk$					
Reported {i,j,k} pairs / total {i,j,k} pairs	49.06% [38.83 - 59.44]	48.21% [39.64 - 56.37]	48.21% [39.64 - 56.37]	48.21% [39.64 - 56.37]	48.21% [39.64 - 56.37]
Correctly ranked {i,j,k} pairs / ranked {i,j,k} pairs	100.00% [100.00 - 100.00]	96.70% [93.62 - 98.52]	93.59% [89.73 - 96.85]	82.44% [73.39 - 88.32]	76.34% [66.51 - 83.31]
Reconstructed consecutive ranks r_i and r_{i+1}					
Missing reconstructed consecutive ranks / 29	0.41% [0.00 - 3.45]	12.59% [0.00 - 34.48]	23.52% [3.45 - 58.62]	69.97% [17.24 - 100.00]	83.48% [51.72 - 100.00]
Correctly reconstructed consecutive ranks / reconstructed consecutive ranks	99.96% [96.43 - 100.00]	85.95% [62.96 - 100.00]	80.64% [56.25 - 100.00]	78.17% [46.15 - 100.00]	81.78% [33.33 - 100.00]
Correctly reconstructed consecutive ranks / 29	99.55% [93.10 - 100.00]	75.00% [55.17 - 86.21]	61.34% [31.03 - 86.21]	22.62% [0.00 - 58.62]	12.97% [0.00 - 34.48]

$S = 50\%$	$Var(error)$				
	0	0.1	0.2	0.6	0.9
Pairwise rankings $r_i jk$					
Reported {i,j,k} pairs / total {i,j,k} pairs	24.95% [18.47 - 33.79]	24.27% [18.01 - 31.09]	24.27% [18.01 - 31.09]	24.27% [18.01 - 31.09]	24.27% [18.01 - 31.09]
Correctly ranked {i,j,k} pairs / ranked {i,j,k} pairs	100.00% [100.00 - 100.00]	96.76% [92.84 - 98.55]	93.62% [89.07 - 97.15]	82.37% [72.88 - 88.93]	76.19% [64.63 - 83.37]
Reconstructed consecutive ranks r_i and r_{i+1}					
Missing reconstructed consecutive ranks / 29	3.93% [0.00 - 10.34]	16.83% [0.00 - 51.72]	35.38% [6.90 - 72.41]	81.14% [55.17 - 100.00]	91.31% [68.97 - 100.00]
Correctly reconstructed consecutive ranks / reconstructed consecutive ranks	98.23% [85.71 - 100.00]	81.58% [65.22 - 95.65]	78.12% [47.06 - 100.00]	80.50% [0.00 - 100.00]	87.52% [0.00 - 100.00]
Correctly reconstructed consecutive ranks / 29	94.38% [79.31 - 100.00]	67.72% [37.93 - 86.21]	49.83% [27.59 - 68.97]	14.83% [0.00 - 37.93]	7.14% [0.00 - 17.24]

Notes: Results using different values of S, the random subset each local observer knows, and $var(error)$. Mean across 100 localities. The bounds below the mean are for the minimum and maximum across localities in brackets.

The total number of {i,j,k} pairs in each locality is 9*(30*29/2)=3915. The total number of {i,j} pairs is 30*29/2=435. Ranks are reconstructed using the relative position index P_i. A reconstructed consecutive rank is missing when $P_i - P_j = 0$. All batches of simulations use the same randomization seed.

Table A7: Performance of the method - Simulations (2)

$S = 30\%$	$Var(error)$				
	0	0.1	0.2	0.6	0.9
Pairwise rankings r_ijk					
Reported {i,j,k} pairs / total {i,j,k} pairs	8.88% [4.90 - 12.67]	9.09% [5.26 - 13.69]	9.09% [5.26 - 13.69]	9.09% [5.26 - 13.69]	9.09% [5.26 - 13.69]
Correctly ranked {i,j,k} pairs / ranked {i,j,k} pairs	100.00% [100.00 - 100.00]	96.87% [92.72 - 99.18]	93.87% [86.59 - 97.50]	82.54% [72.89 - 89.75]	76.43% [64.84 - 84.68]
Reconstructed consecutive ranks r_i and r_{i+1}					
Missing reconstructed consecutive ranks / 29	15.97% [3.45 - 41.38]	20.45% [3.45 - 51.72]	28.83% [6.90 - 62.07]	75.86% [37.93 - 100.00]	85.38% [58.62 - 100.00]
Correctly reconstructed consecutive ranks / reconstructed consecutive ranks	83.67% [62.96 - 100.00]	73.15% [52.63 - 90.00]	68.48% [40.00 - 88.89]	71.76% [0.00 - 100.00]	68.46% [0.00 - 100.00]
Correctly reconstructed consecutive ranks / 29	70.34% [41.38 - 86.21]	58.10% [31.03 - 79.31]	48.62% [24.14 - 68.97]	16.52% [0.00 - 41.38]	9.59% [0.00 - 24.14]
$S = 10\%$	$Var(error)$				
	0	0.1	0.2	0.6	0.9
Pairwise rankings r_ijk					
Reported {i,j,k} pairs / total {i,j,k} pairs	1.02% [0.26 - 2.76]	1.08% [0.28 - 2.89]	1.08% [0.28 - 2.89]	1.08% [0.28 - 2.89]	1.08% [0.28 - 2.89]
Correctly ranked {i,j,k} pairs / ranked {i,j,k} pairs	100.00% [100.00 - 100.00]	96.45% [78.57 - 100.00]	93.82% [75.00 - 100.00]	83.15% [62.26 - 96.77]	76.37% [52.17 - 93.55]
Reconstructed consecutive ranks r_i and r_{i+1}					
Missing reconstructed consecutive ranks / 29	59.97% [37.93 - 86.21]	58.03% [31.03 - 82.76]	58.38% [34.48 - 82.76]	59.86% [34.48 - 79.31]	59.62% [27.59 - 79.31]
Correctly reconstructed consecutive ranks / reconstructed consecutive ranks	64.98% [30.00 - 100.00]	63.05% [36.36 - 100.00]	63.64% [28.57 - 100.00]	56.18% [14.29 - 83.33]	55.15% [30.77 - 88.89]
Correctly reconstructed consecutive ranks / 29	25.93% [6.90 - 44.83]	26.31% [10.34 - 44.83]	26.21% [6.90 - 48.28]	22.72% [3.45 - 44.83]	22.28% [6.90 - 37.93]

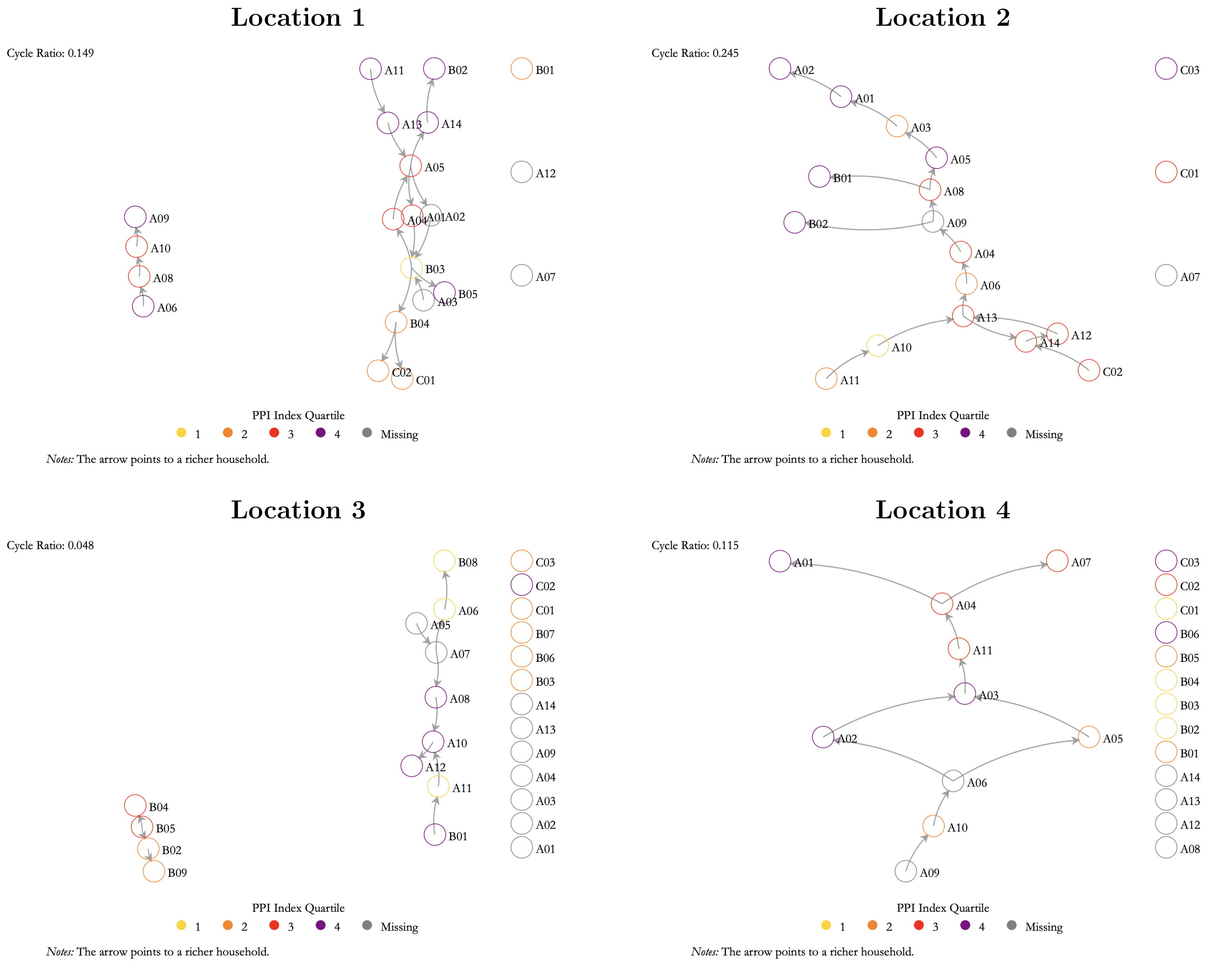
Notes: Results using different values of S, the random subset each local observer knows, and $var(error)$. Mean across 100 localities. The bounds below the mean are for the minimum and maximum across localities in brackets.

The total number of {i,j,k} pairs in each locality is 9*(30*29/2)=3915. The total number of {i,j} pairs is 30*29/2=435. Ranks are reconstructed using the relative position index P_i. A reconstructed consecutive rank is missing when $P_i - P_j = 0$. All batches of simulations use the same randomization seed.

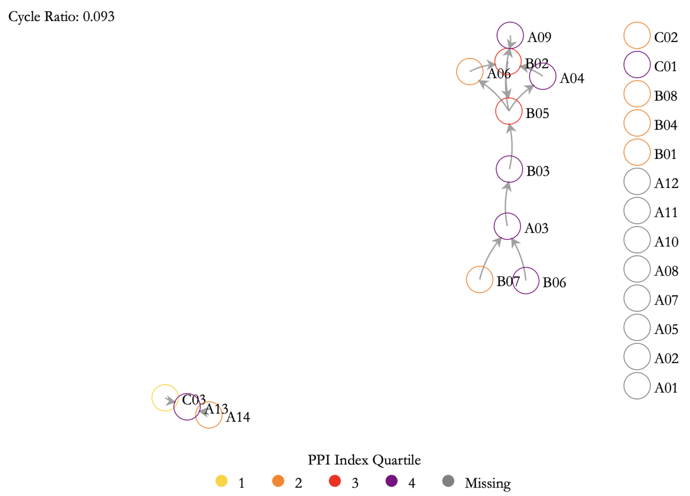
Network Figures

Targeted households (i.e., those on which we have listing or survey information) are identified in the graphs by the letter A. Households that were added as external observers for the sole purpose of the ranking exercise are identified in the graphs by the letter B. Key informants are identified by the letter C. No consumption information was collected on B and C observers.

Figure A1: Directed graph of relative rankings - Urban Slums

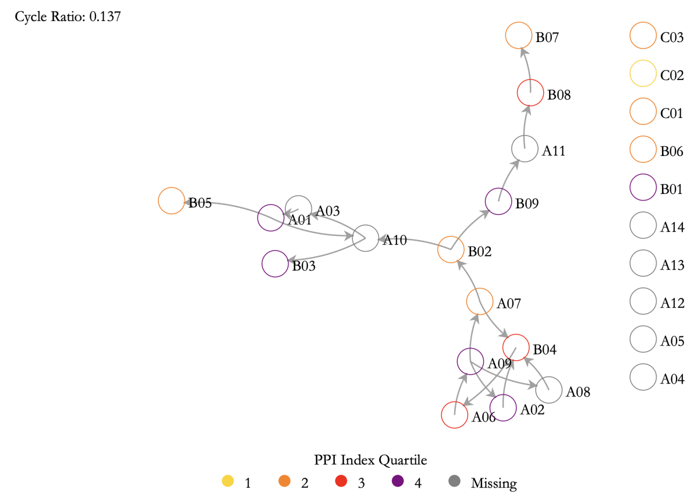


Location 5



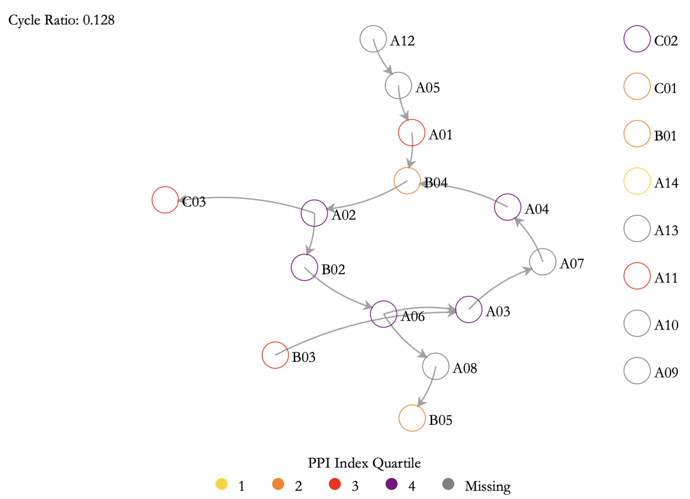
Notes: The arrow points to a richer household.

Location 6



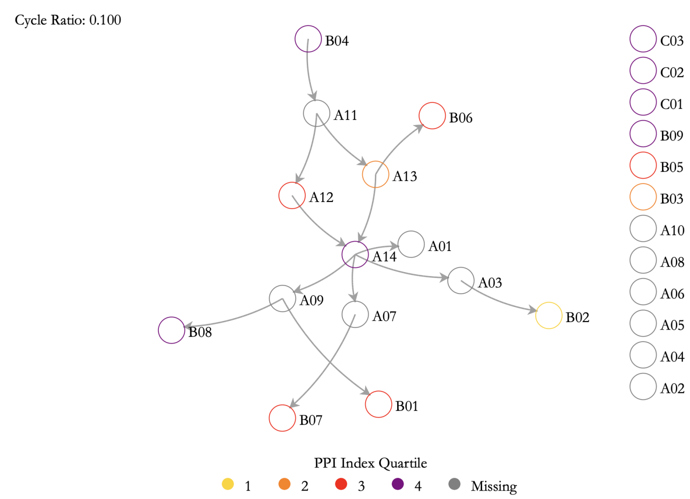
Notes: The arrow points to a richer household.

Location 7



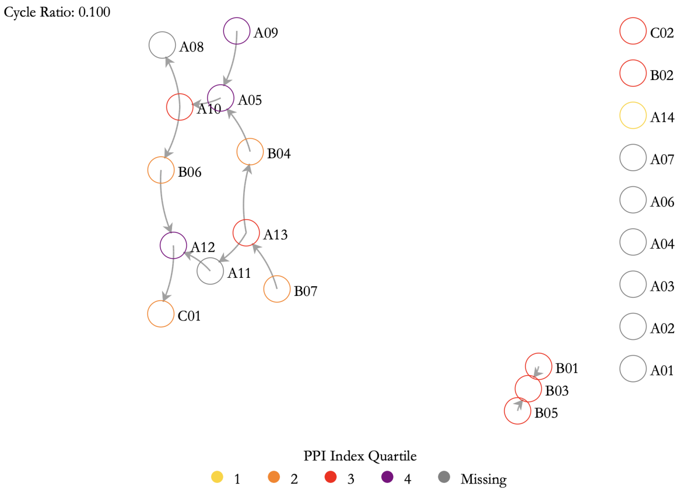
Notes: The arrow points to a richer household.

Location 8



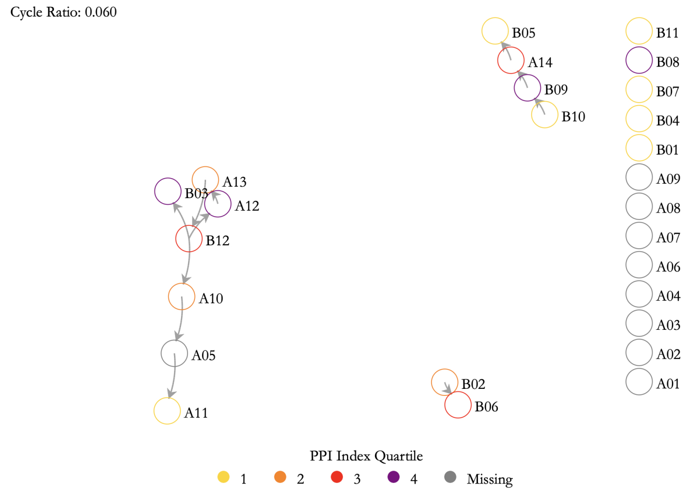
Notes: The arrow points to a richer household.

Location 9



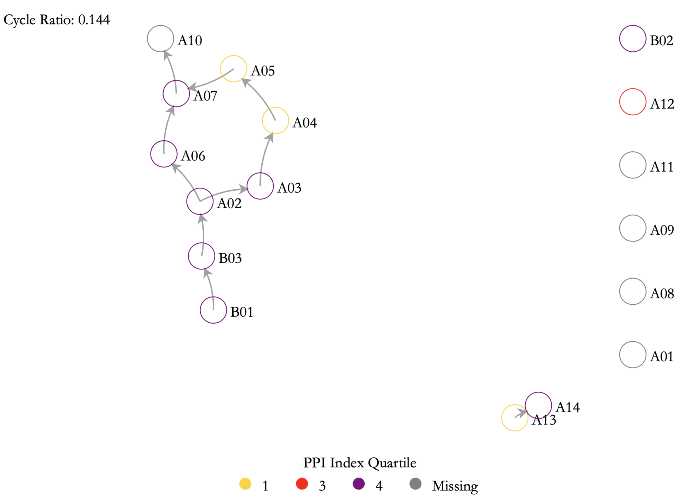
Notes: The arrow points to a richer household.

Location 10



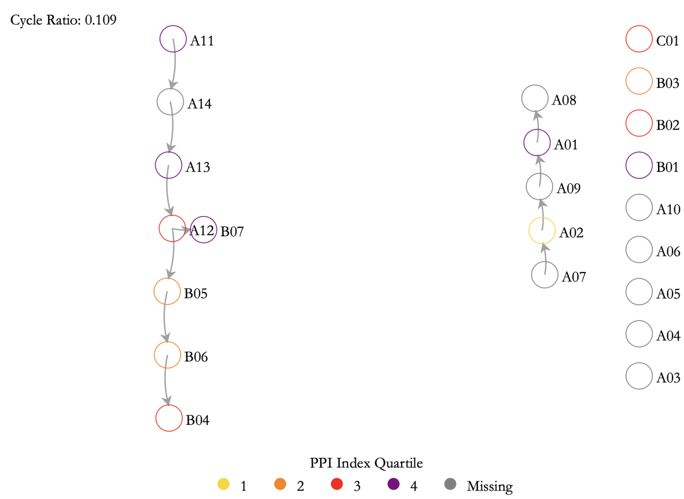
Notes: The arrow points to a richer household.

Location 17



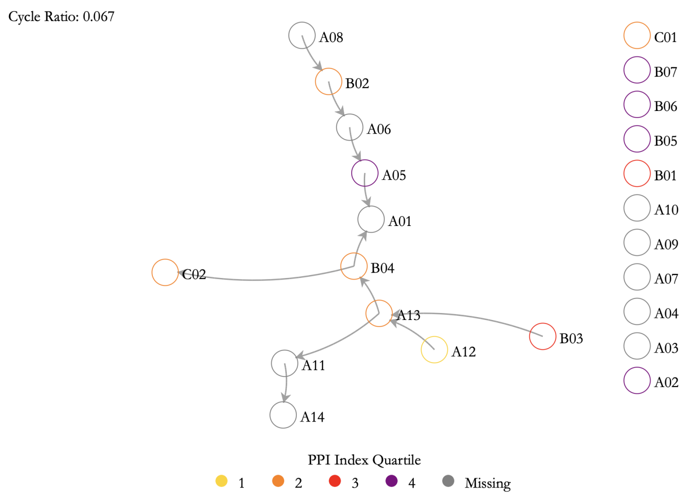
Notes: The arrow points to a richer household.

Location 18



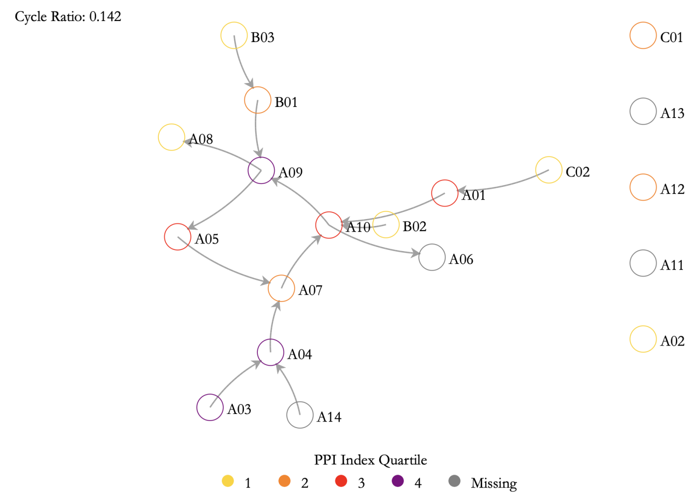
Notes: The arrow points to a richer household.

Location 19



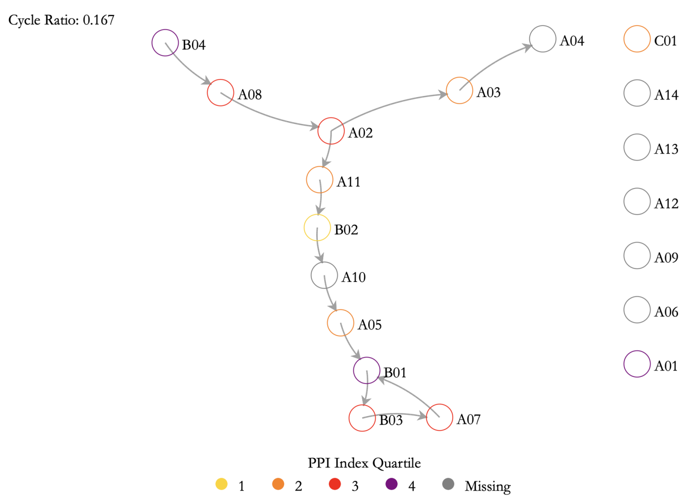
Notes: The arrow points to a richer household.

Location 20



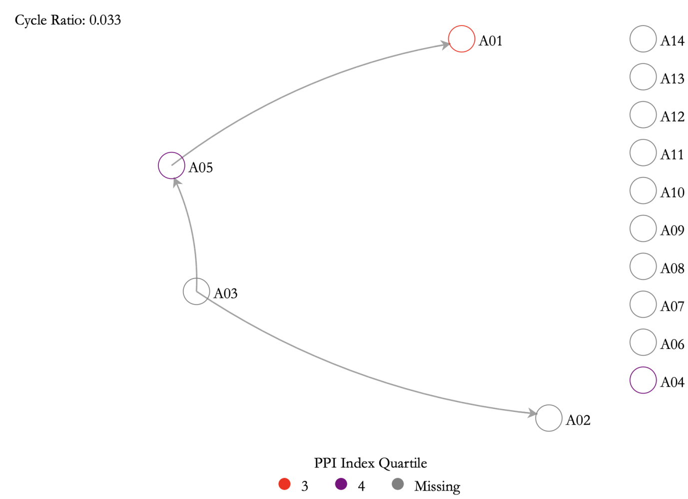
Notes: The arrow points to a richer household.

Location 21



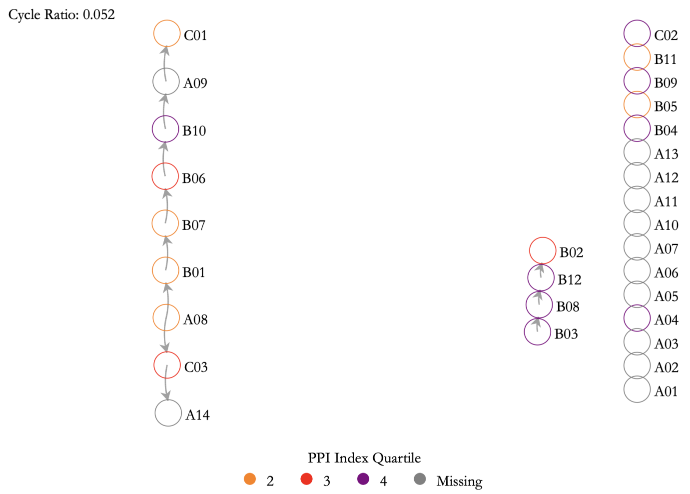
Notes: The arrow points to a richer household.

Location 22



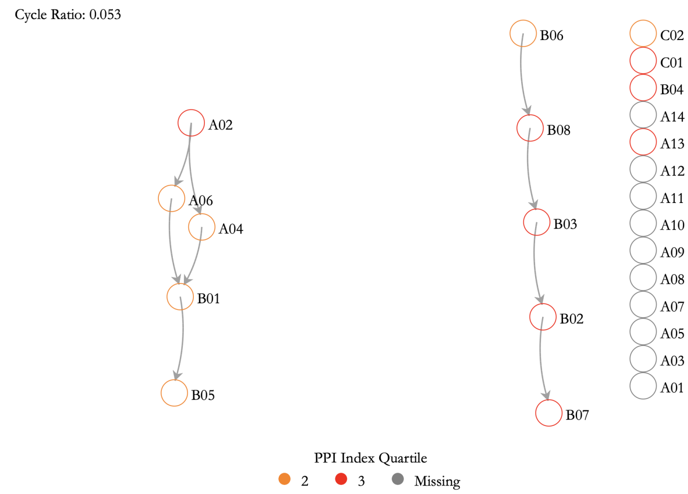
Notes: The arrow points to a richer household.

Location 23



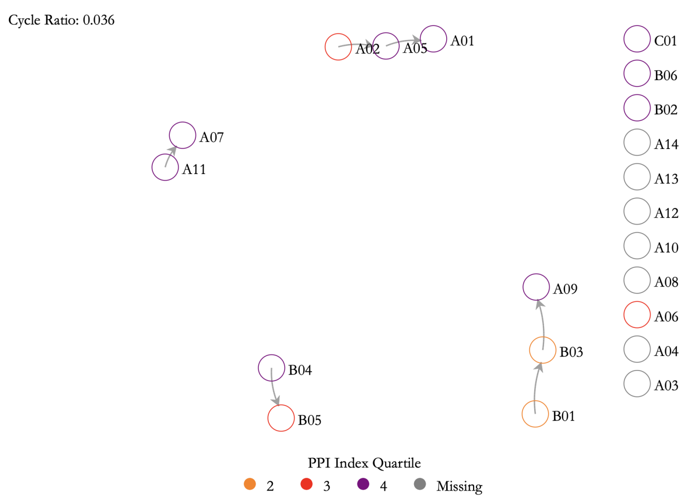
Notes: The arrow points to a richer household.

Location 24



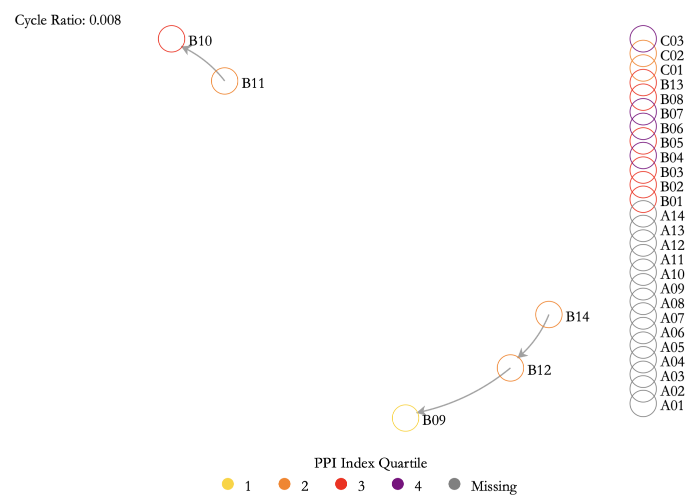
Notes: The arrow points to a richer household.

Location 25



Notes: The arrow points to a richer household.

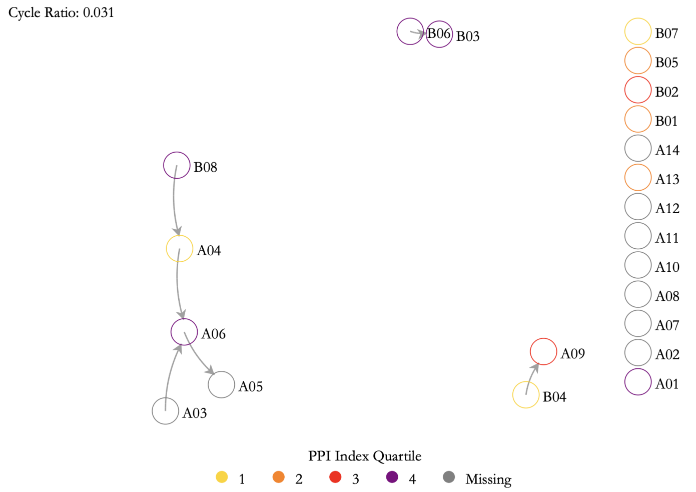
Location 26



Notes: The arrow points to a richer household.

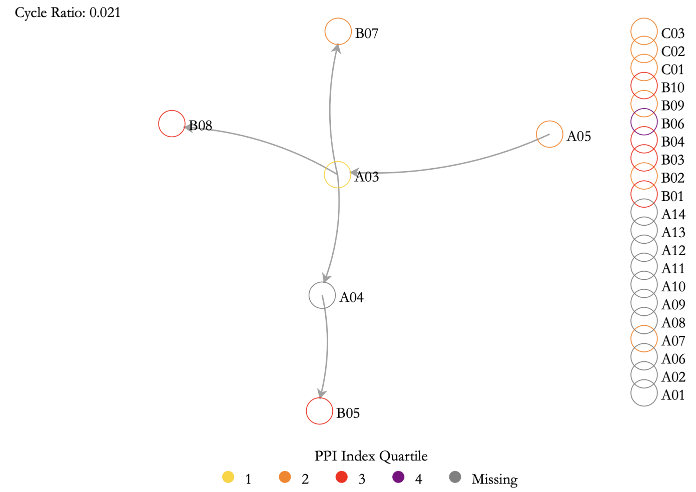
Figure A2: Directed graph of relative rankings - Rural Villages

Location 11



Notes: The arrow points to a richer household.

Location 12

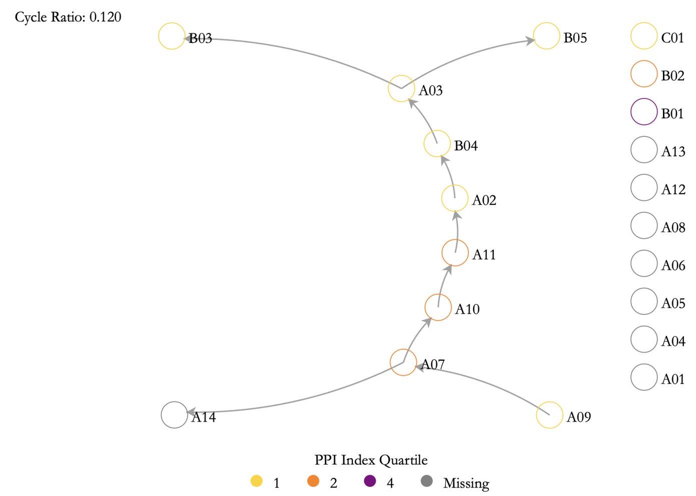


Notes: The arrow points to a richer household.

Location 13

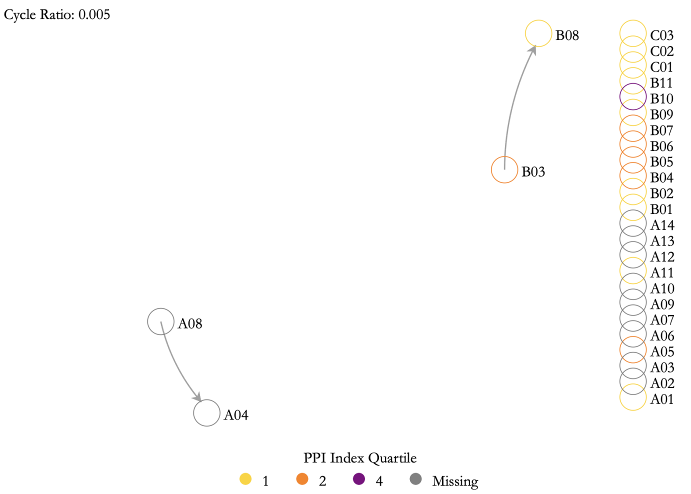
Notes: The arrow points to a richer household.

Location 14



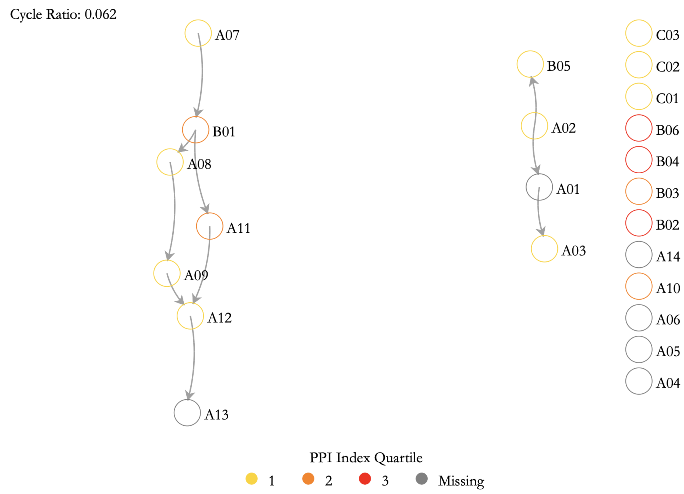
Notes: The arrow points to a richer household.

Location 15



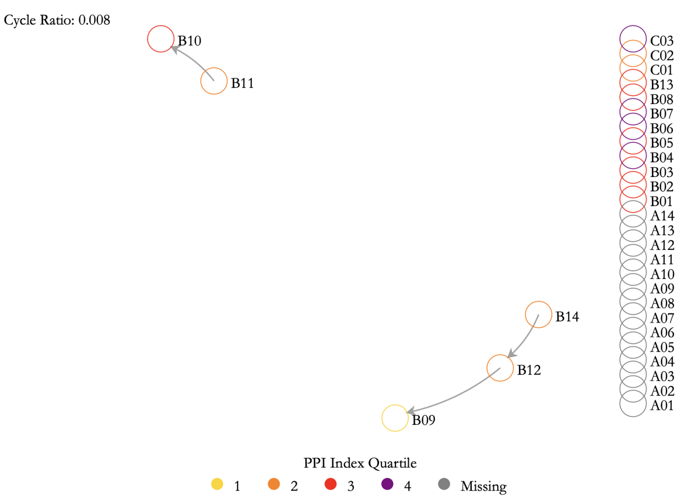
Notes: The arrow points to a richer household.

Location 16



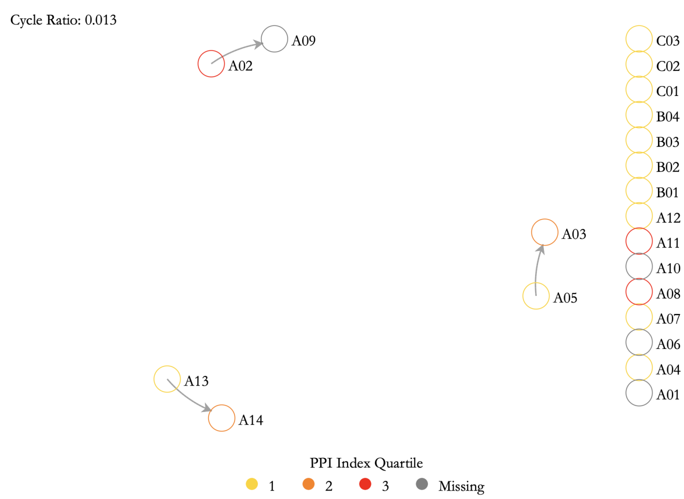
Notes: The arrow points to a richer household.

Location 27



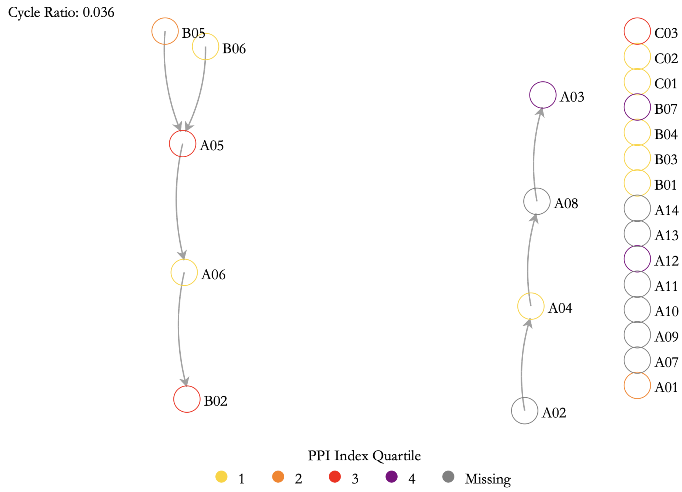
Notes: The arrow points to a richer household.

Location 28



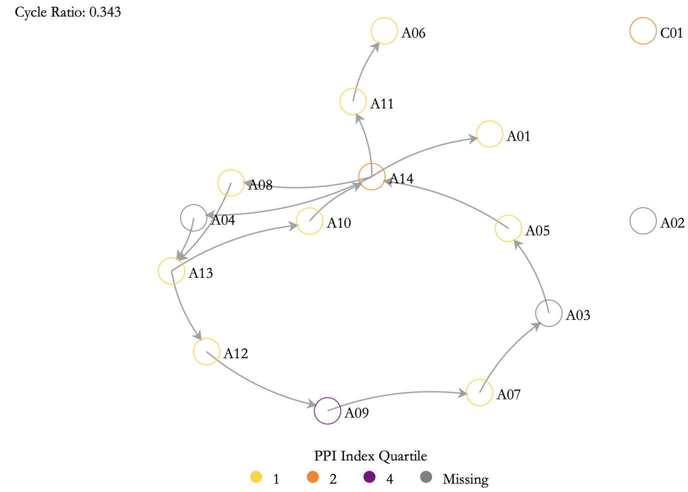
Notes: The arrow points to a richer household.

Location 29



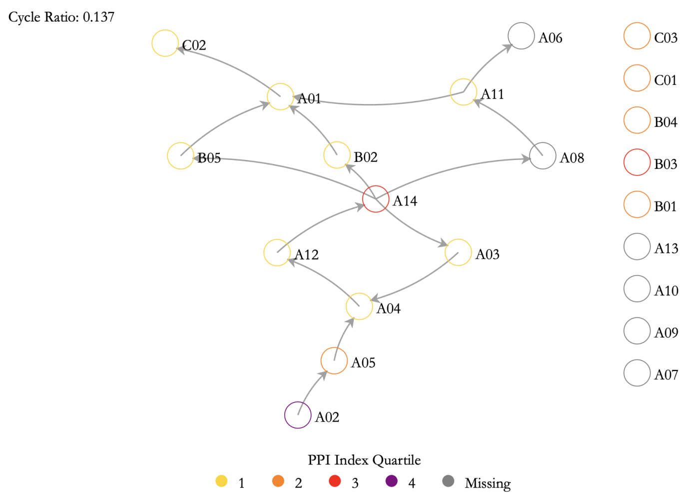
Notes: The arrow points to a richer household.

Location 30



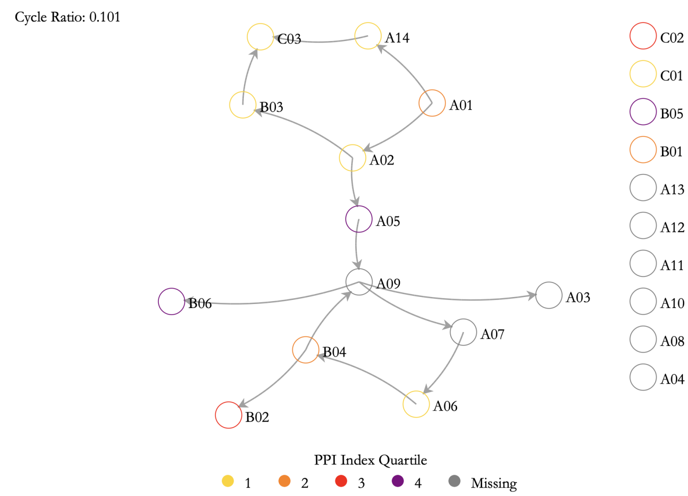
Notes: The arrow points to a richer household.

Location 31



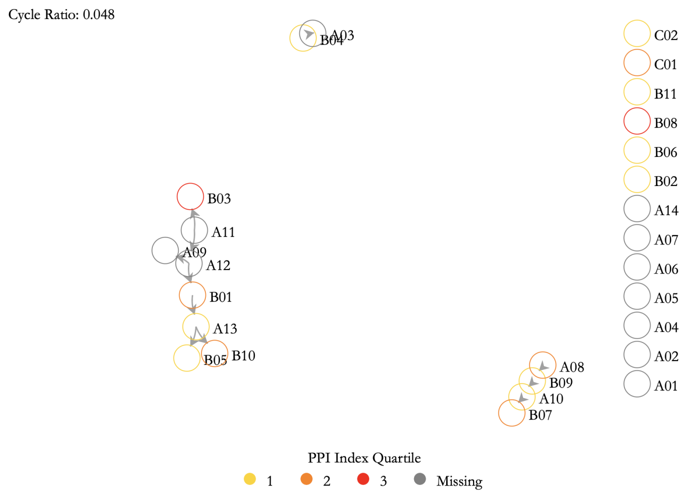
Notes: The arrow points to a richer household.

Location 32



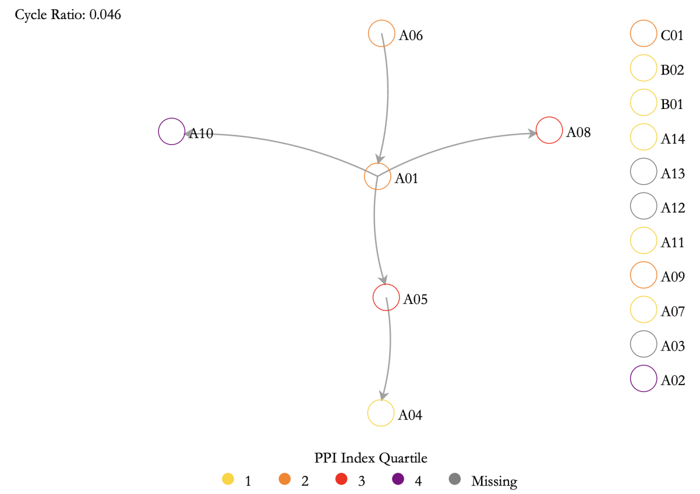
Notes: The arrow points to a richer household.

Location 33



Notes: The arrow points to a richer household.

Location 34



Notes: The arrow points to a richer household.