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

Pascaline Dupas† Marcel Fafchamps‡ Deivy Houeix§

August 3, 2024

Abstract

We introduce a novel approach for eliciting relative poverty rankings that aggregates partial orderings reported independently by multiple neighbors. We first identify the conditions under which the method recovers more accurate rankings than the commonly used Borda count method. We then apply the method to secondary data from rural Indonesia and to original data from urban Côte d'Ivoire. We find that the aggregation method works as well as Borda count in the rural setting but, in the urban setting, reconstructed rankings from both the pairwise and Borda count methods are often incomplete and sometimes contain ties. 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 in which individuals with incomplete information can only produce a partial ranking of alternatives.

*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, Doulo Sow, 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 Abidjan 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).

†Princeton University, NBER and CEPR. Address: Department of Economics, Washington Road, Princeton NJ 08540. Email: pdupas@princeton.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 identify the poor (Ravallion, 2000, 2009, 2015). In some instances, such as when poverty is highly concentrated, community-based or spatial targeting may be sufficient (Burke et al., 2021; Elbers et al., 2007). However, identifying the poor usually requires within-community targeting mechanisms.[1]

Various strategies have been developed to identify the poorest members of communities (Grosh et al., 2022). 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 (Skoufias et al., 1999). In practice, approaches vary in the type of information that is collected: detailed surveys on consumption and income (Deaton, 2019; Grosh and Glewwe, 2000) or light surveys on poverty indicators—e.g., assets (Elbers et al., 2003, 2007) or answers to subjective well-being questions (Ravallion, 2000, 2014; Ravallion and Lokshin, 2001; Ravallion et al., 2016). These methods all have shortcomings: detailed surveys are expensive and time-consuming; short surveys are thought to be limited due to their simplicity and model fit, and may be more easily manipulable by respondents (Banerjee et al. 2020); and subjective well-being is often not well correlated with material well-being, either over time or across countries (Blanchflower and Oswald, 2004; Fafchamps and Shilpi, 2008; Layard, 2009). Furthermore, the rankings are affected by measurement error and possible response bias or manipulation, leading to mis-assignment (Cruces et al., 2013).[2]

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 (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. (2019). 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

[1] Such as the Core Welfare Indicators Questionnaire (CWIQ) or the Simple Poverty Scorecards (SPS) pioneered by the World Bank and used in many surveys across the world (Diamond et al., 2016; Premand and Barry, 2022).

[2] Measurement error arises when respondents have only imperfect knowledge of the answer—e.g., because they do not recall or do not have full information about others. This noise leads to errors of assignment—known as type I and type II errors (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 (Ravallion, 2008).

neighborhood (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 (Trachtman et al., 2022). It has also been shown to yield valuable information in domains other than poverty rankings, notably entrepreneurial potential (Hussam et al., 2022) and long-term poverty (Trachtman et al., 2022). It is, however, vulnerable to local prejudices and views about who are the deserving poor (Alatas et al., 2019; Galasso and Ravallion, 2005; Ravallion, 2008). It also assumes that a small number of community members have the necessary information to provide all the requested rankings (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.[3]

This paper proposes a novel methodology for aggregating partial rankings and implements 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.[4] In Abidjan, we ask respondents—called observers—in 34 different neighborhoods to rank up to 14 target households in that neighborhood, which may not necessarily be the same 14 households. Limited overlap between the sets of households ranked by each observer can lead to bias when relying on rankings averaged across observers, as was done by Alatas et al. (2012). We instead develop a methodology to aggregate all the available—but partial—information provided by the respondents.

[3] 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.

[4] In particular, in 2019, Côte d'Ivoire started to roll out its universal health care coverage (CMU—Couverture Maladie 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 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.

We then compare the reported and aggregated rankings to poverty rankings based on household survey data in Indonesia and Côte d'Ivoire. In both countries, target households answered a survey covering income, 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 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;[5] 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. Building on the work of Tangian (2000), we propose a method for aggregating reported rankings from multiple observers. This method constitutes an alternative to the commonly used Borda count method that averages these reported ranks. In contrast, our method first averages *pairwise* rankings and then 'stitches' them together to produce an aggregate ordering. We show that this method outperforms the Borda method when observers only rank some of the available alternatives (e.g., some of their neighbors) and when they rank alternatives whose true ranks are proximate (e.g., the ranked neighbors are all poor or are all rich). We also develop a simple way of obtaining robust 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 Borda rank-averaging method used by the authors. In the urban context of Abidjan, however, we find that the estimated aggregate rankings fall short of expectations, due to two critical shortcomings: (1) they are incomplete in all cases—sometimes severely so; and (2) they often contain ties. In spite of this, the estimated orderings obtained from our method tend to outperform those obtained by the Borda method or by an algorithm commonly used in sports rankings (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—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.

[5] We apply the weights used by the Government of Côte d'Ivoire to construct the PMT index.

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 estimated 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 as well. But these poverty assessments 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.

We also investigate whether including self-ranks improves accuracy: since observers presumably have better information about themselves, they should be able to rank themselves relative to others. 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 the way observers rank themselves compared to others near them is unhelpful for the purpose of identifying the relatively poor: dropping self-ranks improves targeting accuracy somewhat. This appears driven by poor households over-stating their own material welfare relative others (as in Cruces et al. (2013)), suggesting that there may be a psychological cost to admitting one's own poverty (e.g., Bramoullé and Ghiglino, 2022; Ghiglino and Goyal, 2010).

The particularly low correlation between reported rankings and rankings based on survey 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. Alternatively, the challenge may come from social considerations. On the one hand, social arrangements forcing households to share resources with those around them may incentivize relatively well-off individuals to hide their income (Baland et al., 2011)—to appear 'average' to avoid attracting requests for assistance aimed at those who appear too rich. Consistent with this, we find that self-ranking respondents tend to rank themselves richer than how others rank them. On the other hand, as we show in a companion paper using data from the same Abidjan setting, relatively poor households may be keen to manipulate their consumption/behavior to appear 'average' and avoid being stigmatized as too poor (Dupas et al., 2024). The combination of these two forces creates a 'race to (appear in) the middle' that is a possible explanation for why it is difficult for observers in Abidjan to infer the incomes of their neighbors.

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 alternatives. 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. As we show in Section 2, averaging ranks can 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 some of these problems.[6]

The remainder of the paper is organized as follows. Section 2 presents the main methodological contribution of this paper. Section 3 tests the performance of our proposed method in the context of rural Indonesia. Section 4 explains the experimental design and data collection used in Côte d'Ivoire. Section 5 describes the empirical rankings obtained in Côte d'Ivoire using various methods. Section 6 investigates whether rankings are informative in the Côte d'Ivoire data while Section 7 examines the self-rank randomized treatment. Section 8 looks at characteristics that predict the propensity to rank others. Section 9 concludes.

[6] Truthful reporting is necessary for our method to yield correct rankings, but this is also true of other approaches.

2 Methodology

Our task is to aggregate orderings of alternatives that are provided by a multiplicity of respondents, each of which only observes a subset of alternatives. Our objective is to obtain an aggregate ordering that is as close as possible to the average ordering of alternatives in the population. The aggregation of orderings is a fundamental element of the field of social choice theory, which emerged from seminal contributions made by Condorcet, Borda, and Laplace in the 18th century (List, 2022). We focus our attention on a pairwise method for aggregating individual preference orderings. In this method, individual observers are asked to rank a series of alternatives. From these reported orderings, the researcher extracts comparisons between pairs of alternatives, which are then averaged across observers—in a way similar to majority voting. These averaged pairwise comparisons are then 'stitched together', in a way we describe below, to produce an aggregate ordering of preferences.

We start by noting that, if observers all rank the same alternatives, the asymptotic properties of the pairwise method are identical to the two most commonly used methods for aggregating preferences over alternatives: the Borda count method, which averages the ordinal rankings of alternatives reported by individual observers; and the scoring method, which averages the scores given to alternatives by individual observers. Since the scoring method requires more precise reporting by observers and is thus less practical in most contexts, we focus our attention to the pairwise method and Borda count methods, which both start from the same ordinal reports.

Next, we show that the pairwise method we develop performs better than the Borda count under some realistic conditions. These conditions - that individual observers often rank only partly overlapping subsets of alternatives - are directly relevant to poverty rankings. For instance, some observers only rank low alternatives while others only rank high alternatives. In many such situations, the pairwise method produces the true aggregate ordering while the Borda count has a large mean-square-error. We illustrate this feature with examples and simulations. We end the section with a Monte Carlo analysis of the performance of our rank estimator in the presence of mismeasurement.

2.1 Intuition

The task of aggregating income orderings of neighbors reported by different observers is closely related to a large literature in social choice theory that studies the aggregation of preference ordering across individual agents (e.g., Gehrlein, 1983). Ever since the foundational work of Condorcet, Laplace and Borda, this literature has followed three main tracks: a majority voting approach, which was the focus of Condorcet's attention; a social welfare scoring approach pioneered by Laplace; and a ranking methodology proposed by Borda—see, for instance, List (2022) for a recent survey of this literature.

In the majority voting approach, individuals vote on pairs of alternatives and these votes are then combined either to determine an aggregate ordering or, more simply, to designate a winner among all the alternatives. As already noted by Condorcet himself, this approach need not result in a transitive ordering: cycles may emerge that involve all or a fraction of the alternatives (Gehrlein, 1983). The approach favored by Borda instead turns individual preference orderings into normalized ranks that are then summed across individuals to construct an aggregate ranking of all the alternatives. While this rank averaging approach eliminates the possibility of ties, it does not rule out ties between alternatives.[7] The usefulness of these two approaches extends well beyond politics to include, among others, the ranking of applicants to a job and that of students in an exam.[8] The approach proposed by Laplace has led to welfare economics, whereby alternatives are chosen depending on their social welfare value.

A lesser-known yet insightful article by Tangian (2000) offers a useful bridge between the three approaches.[9] This author notes that, under fairly general assumptions that exclude coordinated action between agents, the Condorcet majority voting approach converges to the same aggregate ordering as the Laplace social welfare approach if the number of voters is large enough; and that, under similar conditions, the social welfare approach converges to the same ordering as the Borda rank averaging approach. Since the rank averaging approach always yields a transitive ordering (with possible ties), an immediate corollary is that, with a large enough number of voters, Condorcet cycles disappear. Tangian shows that, under some fairly realistic conditions, pairwise voting results in a transitive ordering with a high probability even with a relatively small number of voters—such as those involved in election polls. The Tangian results may be of little use in parliamentary politics because coordinated action by a small number of parties effectively rules it out. But it is useful in aggregating income rankings reported by many independent observers.

In the rest of this Section, we extend the Tangian results to settings in which individual agents are only able to rank some of the alternatives relative to each other. This could arise,

[7]As already noted by Laplace, the averaging approach also allows the use of cardinal information available to individual agents (i.e., information about the distance between two alternatives), in which case the averaging is done on scores, not ranks. This possibility receives little attention here because, in our empirical setting, cardinal information was not collected from ranking agents.

[8]For instance, each member of a hiring committee may start by ranking or scoring a subset of applicants. The committee then takes the average of these ranks or scores to construct an aggregate ranking, and may resort to voting to select a candidate or to break ties. Similar procedures are customarily followed when marking exams, assigning prizes or grants, and selecting papers for a conference.

[9]We are grateful to an anonymous referee for pointing this work to us.

for instance, because agents do not have sufficient information on all alternatives, or because ranking all alternatives is too costly for them. We call the alternatives they are willing or able to rank their *rankable set*. The majority voting approach still applies in this setting, but agents vote only on pairs of alternatives in their rankable set. Tangian (2000) shows that, as long as *rankable sets* are distributed randomly across agents, aggregating orderings through majority voting still converges to a transitive ordering—provided that some matrix rank conditions are satisfied (see below).

A similar extension can be made for rank averaging provided that ranks are normalized before averaging, to account for the varying size of each observer's rankable set. More precisely, let the rankable set of observer k have size m_k and let the ranks reported by observer k for alternative i be denoted as $r_i^k \in \{1, 2, ..., m_k\}$. The normalized rank given by observer k to alternative i is then defined as $z_i^k \equiv \frac{r_i^k}{m_k+1}$. With this normalization, observers with a rankable set of size 1 report a rank of 0.5 (no information); those with two observations report normalized ranks $\{1/3, 2/3\}$; those with three have normalized ranks $\{1/4, 2/4, 3/4\}$; etc.

Importantly, the requirements for the random assignment of observers are stricter than for ranking through majority voting: *individual alternatives* have to be assigned randomly to agents, not just rankable sets. The reason for this stricter requirement can be illustrated with a simple example. Consider two sets of observers who all share the same latent preferences over five alternatives $y_1 < y_2 < y_3 < y_4 < y_5$. One set of observers is assigned rankable set $\{y_1, y_2, y_3\}$ and the other set of observers is assigned rankable set $\{y_3, y_4, y_5\}$. If we ask all observers to vote on each possible pair of alternatives in their rankable set, we can recover the true ordering by 'stitching together' their votes on the two sets, given that they have alternative y_3 in common. Averaging ranks, however, does not produce the correct ordering. With m_1 and m_2 observers in each set, respectively, the average ranks for the five alternatives are $\{1/4, 2/4, \frac{0.75m_1+0.25m_2}{m_1+m_2}, 2/4, 3/4\}$, which shows that not only are options y_2 and y_4 tied, option y_3 is ranked above y_4 or below y_2 depending on the relative number of observers of each set—or tied with them if $m_1 = m_2$. The correct ordering can, however, be recovered if observers are all assigned a *randomly selected* set of three alternatives, in which case the averaged normalized rankings converge to $\{1/4, 1.5/4, 2/4, 2.5/4, 3/4\}$ as the number of observers increases—which is the same ranking as that obtained by 'stitching together' two orderings sharing a common alternative.

These observations are the main motivation behind our effort to develop a methodology that relies on a pairwise majority voting approach rather than rank averaging when rankable sets vary across observers and the researcher has reasons to expect true ranks to be correlated within rankable sets—which arises, for instance, because some observers see mostly low-ranked alternatives (e.g., they have poor neighbors) while others see mostly high-ranked alternatives (they have rich neighbors).

2.2 Theory

2.2.1 Extending Tangian's results to pairwise rankings

To formalize our intuition, we borrow heavily from Tangian (2000) whose results we start by summarizing here. Consider n individuals each with a latent ordering on two alternatives y_A and y_B which we denote by a utility function $u_k(y_i)$ for agent k and alternative y_i. We say that y_A is socially preferred to y_B if:

$$\sum_{k=1}^{n} u_k(y_A) > \sum_{k=1}^{n} u_k(y_B)$$

In the empirical part of the paper, we consider the special case in which $u_k(y_i)$ becomes observer k's belief $E_k[y_i]$ about household i's income y_i and we wish to rank y_A higher than y_B is the observers' average belief is higher for y_A than y_B.

Following Tangian (2000), we decompose the cardinal utility $u_k(y_i)$ into an ordinal part η^k (i.e., the ordering) and a cardinal residual ξ^k (i.e., the difference in value between $u_k(y_A)$ and $u_k(y_B)$). Formally: let $\eta^k = 1$ if $u_k(y_A) \geq u_k(y_B)$, i.e., y_A is ranked higher than y_B, and $\eta^k = 0$ otherwise; and define:

$$\xi^k = |u_k(y_A) - u_k(y_B)|$$

With this notation, a majority of observers ranks y_A above y_B if $\sum_{k=1}^{n} \eta^k > n/2$ or, equivalently, if:

$$\sum_{k=1}^{n} (2\eta^k - 1) > 0 \tag{1}$$

Similarly, y_A is socially preferred to y_B if:

$$\sum_{k=1}^{n} (2\eta^k - 1)\xi^k > 0 \tag{2}$$

Using this approach, Tangian (2000) shows that the probability that an ordering obtained by majority voting (1) differs from an ordering based on reported utilities (2) converges to 0 as $n \to \infty$ if η^k and ξ^k are independent random variables.[10] Tangian (2000) also offers an approximation formula for this difference in finite samples.

[10] There is also a special 'knife-edge' case where $\sum_{k=1}^{n} \eta_k = n/2$, in which case this probability converges to $1/2$, i.e., the two alternatives are equally ranked and the outcome of the vote is random.

This result generalizes to multiple alternatives $m > 2$. In this case, we have $m(m-1)/2$ unordered pairs of alternatives on which we have elicited rankings from observers, thereby mimicking majority voting on all pairs of alternatives. This leads to Tangian (2000)'s main result [Theorem 1] that the probability that the ordering obtained from pairwise majority voting (equation 1) is equal to the ordering based on summing utilities (equation 2) tends to 1 as $n \to \infty$, again assuming that η_i and ξ_i are independent random variables across observers and alternatives—and ruling out ties. An immediate consequence of this theorem is that, under the same conditions, the fraction of random samples in which the pairwise majority voting ordering is not transitive vanishes as $n \to \infty$ [Theorem 2]. This follows from the property that, in the absence of ties, an ordering obtained by averaging utilities is always transitive by construction. Tangian (2000) [Theorem 3] also notes that these properties extend to Borda counts, that is, to orderings obtained by averaging normalized ranks instead of utilities, where, as before, a normalized rank is defined as:

$$z_i^k \equiv \frac{r_i^k}{m+1} \tag{3}$$

where the $r_i^k \in \{1, 2, ..., m\}$ are the ordinal ranks reported by agent k and m is the number of alternatives ranked by that agent.

The asymptotic equality between orderings obtained from pairwise majority voting and Borda counts remains if each agent k observes only a randomly selected subset m_k of the m alternatives. This derives directly from the fact that Theorem 1 only depends on the two pairwise inequalities (1) and (2), and these can be modified to allow for a varying number of n_{ij} observations per $\{i, j\}$ pairs of alternatives. Thus, as long as each $n_{ij} \to \infty$ for each $\{i, j\}$ pair, Theorem 1 holds and, by consequence, also Theorems 2 and 3 as summarized here.[11]

Things are different if the m_k subsets of alternatives are selected such that the true ranks r_i (or true values $U(y_i)$) are correlated within subsets. This arises when certain observers only see low-ranked alternatives while others only see high ranked ones. In this situation, orderings based on Borda counts no longer coincide with orderings based on reported utilities or values, and Theorem 3 breaks down. The example presented earlier provides a proof *a contrario* that Borda counts need not converge to true orderings in the presence of rank correlation in m_i subsets. It follows that when observers self-select the alternatives that they rank, orderings based on Borda counts may be unreliable due to the possible (unobserved) correlation in true ranks that this self-selection may generate. In such situations, we may still ask observers to report their value or score $u_k(y_i)$ for each observation in their subset.

[11] This does not, however, apply to the approximation formula for the difference between the two methods, which would have to be adapted to reflect the varying number of observations per pair of alternatives. We do not use the approximation formula in this paper.

But this raises the level of reporting difficulty for respondents—which may lead to excess observation noise, response bias, or sample attrition. It also requires that each observer use the same scoring scale, a requirement that may not hold in practice, e.g., because observers have different scoring rules. Furthermore, this system is manipulable: observers can increase the weight of their preferences in the final average by artificially inflating the range of their scores.

In contrast, pairwise comparisons (1) remain relatively unaffected: their asymptotic properties are unchanged as long as, for each specific $\{i, j\}$ pair of alternatives, the observed "votes" come from a sample of observers that is representative of the distribution of utilities or beliefs in the study population. This requirement can be satisfied if the assignment of rankable subsets to observers is orthogonal to their (unobserved) preference ranking over all latent alternatives. This holds even if true ranks are correlated within rankable subsets., i.e., some rankable subsets are top-heavy while others are bottom-heavy. It can also be satisfied even if each observer self-selects his or her rankable set m_k, as long as the preferences of that observer over all latent alternatives is independent of their selected m_k set. Put differently, it does not matter for equation (1) that certain observers self-select to vote on pairs of alternatives that they rank lowly, while others do the opposite—as long as the distribution of the values $u_k(y_i)$ or beliefs $E_k[y_i]$ (or ranking scale) of these observers is representative of the population of observers. This condition is satisfied when the preferences of observers over all latent alternatives are independent of their observed rankable set, in which case each sample of pairwise comparisons comes from a representative sample.

This leads to the following proposition:

Proposition 1: Assume that all observers share the same social welfare function $u(y_i)$ but have different rankable sets M_k. Let $M_s \subset M$ denote the set of all the alternatives that belong at least one M_k. Let C be the set of all consecutive pairs in the full ordering implied by $u(y_i)$. Then, irrespective of how alternatives are assigned to rankable sets:[12]

1. The scoring method recovers a correct full ordering of all the alternatives in M_s
2. The voting method recovers a correct full ordering of all the alternatives in M_s provided that all the pairs of alternatives in C appear in at least one M_k.
3. The voting method recovers a correct partial ordering of all the alternatives in M_s provided that some the pairs of alternatives in C appear in at least one M_k.

[12]Part 2 of the Proposition guarantees a unique ordering of all ij pairs that appear in at least one M_k. Other pairs may also be rankable by combining information from different rankable sets M_k, a point we revisit in detail in the next subsection.

Proof: Part 1 follows directly from the assumptions of common preferences and no reporting error. For Part 2, the proof is in two parts. (1) The assumption of common preferences and no reporting error implies that all ranked pairs are correctly ranked. (2) If all the pairs in C are correctly ranked, the full ordering of the pairs in M_s can be recoved by daisy-chaining these pairs together (see the next subsection for an algorithm that achieves this). Once such a chain has been found, all the remaining pairwise ranks are subsumed in the full ordering. Part 3 follows from Part 2 and the fact that if a pair in C is missing from the pairs in M_k, than the two alternatives can not be ranked relative to each other.

Proposition 1 does not, however, apply to Borda counts. This is because, as examplified in the previous sub-section, the normalized rankings based on Borda counts from each M_k are biased estimates of the true normalized ranks z_i in M. Indeed, they ignore the fact that the lowest rank alternative in a particular M_k may be ranked higher in M and, similarly, the highest rank alternative in a particular M_k may be ranked lower in M. Furthermore, the spacing between alternatives appearing in M_k is also likely to be irregular, with large gaps in z_i between some pairs and small gaps between others. It follows that all the normalized rankings in each M_k set are biased representations of the true normalized ranks. Averaging these biased estimates over observers does not eliminate the bias. [13] We expect the bias to be particularly large when alternatives in rankable sets M_k are correlated in the sense that alternatives i and j are more likely to be in a set M_k if $|z_i - z_j|$ is small than if it is large. The reason is that, when rankable sets are correlated, the lowest ranked alternative in some of them are highly ranked in M and vice versa.

Definition: Rankable sets are said to be *correlated* when

$$E\left[\sum{}_k \sum_{i\in M_k} \sum_{j\in M_k} (z_i - z_j)^2\right] < E\left[\sum{}_k \sum_{i\in M_k} \sum_{j\notin M_k} (z_i - z_j)^2\right]$$

From this analysis, it appears that, when observers only rank M_k subsets, the scoring method a priori dominates the voting (pairwise comparison) method and, certainly, the Borda count method. The problem is that it requires each observer to score using the *exact* same $u(y_i)$. Since any monotonic transformation of the $u(.)$ function yields the same normalized ranks z_i, there is no a priori reason to expect observers to score using the same $u(.)$ function even when they have the same ordering over alternatives. When observers

[13] We conjecture that the bias may disappear if each alternative in an M_k is randomly assigned to the observer (i.e., there is no self-selection in which alternative he/she observes) and there is a large enough number of observers to eliminate the small sample bias that arises even in that case. Proving this formally is left for future work.

only rank M_k subsets, averaging $u(y_i)$ scores subject to different monotonic transformations is not guaranteed to yield an average score from which the true normalized ranks can be recovered[14]. It follows that the scoring method only fits situations in which (1) scores have an intuitive cardinal meaning to observers and/or (2) observers are given precise instructions on how to score (e.g., range, average, etc). In situations where these conditions are unlikely to hold, aggregating pairwise comparisons dominates because it does not require making cardinal assumptions about $u(y)$.

The above results can be generalized to situations in which observers have different preferences $u(y_i)$. In this case, the researcher is interested in aggregating preferences to get the population ordering of alternatives. This requires selecting a random set of observers from the population of interest. As demonstrated by Tangian (2000), the ordering of alternatives recovered from a sample can be regarded as a close approximation of the population preferences *if* that sample is large enough. This also extends to the case where observers want to report a common true ranking of alternatives, but they either observe or report these true rankings with error. In such cases, averaging scores or pairwise ranks over observers should closely approximate the true ranks if the number of observers is large enough.

2.2.2 Simulated Performance of Ranking Methods

To illustrate this point, we simulate the performance of the Borda count and pairwise methods of preference aggregation when observers' rankable sets only partially overlap with each other. To isolate the role of rankable sets, we assume that all observers have the same preference ordering over all the alternatives in set M and that they report these orderings without error.[15] This means that the only source of statistical variation is the composition of the rankable subsets M_k.

Results are illustrated in Figure A1 for 14 observers ranking a set of 30 households, which aligns with our Côte d'Ivoire study's parameters. Each panel presents the true ordering in blue. The orderings derived from the Borda and pairwise methods are depicted in red and black, respectively. In the top left graph, observers can observe only $S = 5\%$ of the households on average, and the mistake value is $V = 0.05$.[16] Given the minimal overlap between the observers, both methods perform poorly at recovering the correct ranking. However, the pairwise method performs better in predicting the correct order by combining

[14] Precision falls when the M_k's are small relative to M, and a larger sample is needed to achieve the same level of precision.

[15] Given Tangian (2000)'s results, these assumptions are equivalent to assuming that we have a large enough sample for the aggregation of reported orderings to match the true ordering *within rankable sets*.

[16] A higher mistake number indicates greater errors in ranking income. This number represents the standard deviation of a log-normal distribution of the mistake, ultimately multiplied by the true income, which is assumed to follow a normal distribution with a standard deviation of 1.

the orderings from different sets. In Figure A1 (c), we also observe that the ranks obtained using the pairwise method appear less disperse across simulations. Overall, the performance of the Borda method improves as the average proportion of the sample observed by each observer increases, as seen in Figure A1 (d). From this, we conclude that the pairwise method offers a significant improvement over the commonly used Borda method when observers rank a limited subset of alternatives, a common scenario in many settings.

2.3 The pairwise method

2.3.1 Challenges

We have shown that it is a priori appealing to resort to pairwise majority voting as a method for eliciting preferences over alternatives: it remains accurate in many situations in which Borda counts are not; and it imposes less of a burden on respondents than eliciting $u_k(y_i)$ directly. The method, however, raises a number of specific difficulties that we now address.

The first challenge is how to *elicit* pairwise votes on a set of alternatives. Since the number of pairs in a rankable set m_k is $(m_k - 1)m_k/2$, it increases rapidly with m_k. The solution, however, is simple: ask respondents to rank all m_k alternatives and extract pairwise 'votes' from that report. For instance, if an observer reports $U(y_1) < u(y_2) < u(y_4)$, we can recover $3 * 2/2 = 3$ pairwise votes on $\{y_1, y_2\}$,$\{y_1, y_4\}$, and $\{y_2, y_4\}$ that can be used to estimate (1) using all available information on those pairs from a multiplicity of observers with different rankable sets.

The second challenge is how to *aggregate* these votes across all pairs to produce an estimated ordering. The solution we propose is to resort to a specific set of matrix manipulations. For the reader familiar with network matrices, reported ranks are first organized in the form of a directed adjacency matrix. By analogy with the concept of distance in directed matrices, all the alternatives that are above a specific alternative i can be identified by taking forward powers of the adjacency matrix, while all the alternatives below i can be identified by taking backward powers of the same matrix.

The third challenge is to *summarize* partial information. When the sample of observers is small, our proposed method is not guaranteed to produce a full or even transitive ordering of all the alternatives. To this effect we propose a $[0, 1]$ summary index similar in spirit to the normalization of Borda counts: no information means an index value of 0.5, while values below 0.5 implies that the alternative is ranked lower than other alternatives, and vice versa if the index value is above 0.5. With this approach, all alternatives in a Condorcet cycle end up with the same index value, i.e., they are tied.

The fourth challenge is to estimate the *accuracy* of the method in a particular sample. To address this challenge, we rely on a randomization inference approach to mimic the random assignment of rankable sets to observers. Counter-factual samples are generated by sampling observers with replacement from the surveyed sample, and confidence intervals are obtained by reporting the distribution of ranks across simulated samples.

2.3.2 From partial to complete ordering

Our objective is to use partial orderings reported by individual observers to estimate the ordering that would be obtained by averaging normalized ranks (or scores) collected from the entire population. We assume that partial rankable sets are randomized across observers, but we allow alternatives to be correlated within each rankable set.

We start by illustrating how orderings can be represented in matrix form. Consider a set S of m alternatives ranked in the *order* of their values $u(y_i)$:

$$u(y_1) < u(y_2) < ... < u(y_m)$$

with the resulting ordering taking rank values $r_i \in \{1, 2, ..., m\}$. We take this ordering to represent the aggregation of individual true orderings that we would obtain through a Borda count if we had information on all individual orderings. Consequently, we assume this '*true*' ordering to be transitive, possibly with ties.

The true ordering has an $m \times m$ matrix representation R in which each matrix element $\eta_{ij} = 1$ if $u(y_i) < u(y_j)$ and $\eta_{ij} = 0$ if $u(y_i) > u(y_j)$. For instance, ordering $u(y_1) < u(y_2) < u(y_3) < u(y_4)$ is represented as:

$$R \equiv [\eta_{ij}] = \begin{bmatrix} . & 1 & 1 & 1 \\ 0 & . & 1 & 1 \\ 0 & 0 & . & 1 \\ 0 & 0 & 0 & . \end{bmatrix} \tag{4}$$

Diagonal elements η_{ii} are missing since alternative i is not ranked relative to itself. Ties are represented as $\eta_{ij} = \eta_{ji} = 0$.

The sum $\sum_j \eta_{ij}$ of the elements in row i is the number of alternatives ranked higher than i, and the sum $\sum_j \eta_{ji}$ of the elements in column i is the number of alternatives ranked lower than i. It follows that the rank r_i of alternative i is:

$$r_i = m - \sum_j \eta_{ij} = 1 + \sum_j \eta_{ji} \tag{5}$$

where sums are taken over non-missing values. Normalized ranks z_i can also be computed as:

$$z_i \equiv \frac{r_i}{m+1} = \frac{1}{2}\frac{P+m+1}{m+1} \tag{6}$$

where:

$$P \equiv \sum_j \eta_{ji} - \sum_j \eta_{ij}$$

is the difference between the number of alternatives above i minus the number of alternatives below i. We will use this property below.

We observe a series of partial orderings reported by independent observers. We assume these partial orderings to be unbiased but they possibly contain observational or reporting error. Partial orderings can also be represented in matrix form. For instance, if observer a does not rank y_3, the reported ordering can be written as:

$$R_a = \begin{bmatrix} . & 1 & . & 1 \\ 0 & . & . & 1 \\ . & . & . & . \\ 0 & 0 & . & . \end{bmatrix} \tag{7}$$

where the row and columns for alternative 3 are both missing. The matrix representation can also accommodate disconnected orderings – e.g., respondent b reporting $y_1 < y_2$ and $y_3 < y_4$ only—as well as other partial orderings.

The next step is to use a 'majority voting' approach to obtain relative rank estimates for as many pairs as possible and to fill a matrix with them. Formally, let vector $r^k = \{r_1^k, r_2^k, ...r_m^k\}$ be the ordering reported by observer k for alternatives from 1 to m, with some alternatives possibly not ranked by k, i.e., $m_k \leq m$. From this ordering, we construct $m_k(m_k-1)$ pairwise comparisons η_{ij}^k for each observer k as explained earlier. We then apply the 'majority voting' rule (1) to aggregate across observers each pairwise comparison into an estimate of an element of matrix R:

$$\hat{\eta}_{ij} = 1 \text{ if } \sum_{k=1}^{n}(2\eta_{ij}^k - 1) > 0 \text{ and } 0 \text{ otherwise} \tag{8}$$

with $\hat{R} = [\hat{\eta}_{ij}]$. There is a tie whenever $\sum_{k=1}^n \eta_{ij}^k = n/2$, in which case, by (8), both $\hat{\eta}_{ij} = 0$ and $\hat{\eta}_{ji} = 0$. Majority voting on pairs can produce many ties, especially when the number of observations on a pair is small (and even).

To reduce the proportion of ties, we also implement an alternative voting rule that takes rank differences into account. Formally, let r_i^k be the integer rank of alternative y_i reported by observer k, and similarly for r_j^k. Denote the rank difference as $d_{ij}^k = r_j^k - r_i^k$. This difference can be positive or negative, depending on whether $r_j^k >$ or $< r_i^k$. We then estimate the elements of matrix R using:

$$\ddot{\eta}_{ij} = 1 \text{ if } \sum_{k=1}^{n} d_{ij}^{k} > 0 \text{ and } 0 \text{ otherwise} \tag{9}$$

with $\ddot{R} = [\ddot{\eta}_{ij}]$. In a pairwise comparison, formula (9) gives more weight to an alternative i that, on average, is several ranks higher than j by observers. The rationale behind this approach is that, on average, large rank differences are correlated with large differences in values $u_k(y_i)$ or beliefs $E_k[y_i]$. This method brings the estimated preference ordering closer to a social welfare ordering and, by Tangian (2000)'s Theorems 1 and 3, it also brings it closer to an ordering obtained by averaging complete normalized rankings, which is our objective.

Matrices $\hat{R}$ and $\ddot{R}$ summarize the available information about all the pairwise comparisons implicit in the data—but they contain no information on pairs that are not compared by any observer. We may nonetheless be able to fill some of these missing pairs by combining the available information. To illustrate, imagine that one observer reports partial ordering $y_1 < y_2 < y_3$ and a second observer reports $y_3 < y_4$. Given our assumptions of transitivity in true rankings and unbiased reporting, a best guess about the true ordering is $y_1 < y_2 < y_3 < y_4 < y_5$, where we have 'stitched' together the two partial rankings using the common alternative y_3. This methodology can be summarized as follows. Any ordering can be represented as a ranking matrix (e.g., matrix R_a in equation 7) and each ranking matrix can be seen as the adjacency matrix of a directed network in which an arrow from y_i to y_j implies that $y_i < y_j$. It follows that alternatives above y_i can be identified by following all network paths leading from y_j. All these paths can, in turn, be identified by taking powers of the adjacency matrix of the directed network which, in our case, is (a transformation of) $\hat{R}$ or $\ddot{R}$. A similar process can be used to identify all the alternatives below y_i. The result of both these calculations is summarized in a new rank matrix $\tilde{R}$. This matrix identifies all the directed chains of alternatives that can be recovered from the data.

Matrix $\tilde{R}$ is then used to compute an estimate $\widetilde{z}_i$ of the normalized ranks using formula (6). If matrix $\tilde{R}$ contains a complete transitive ordering of all the alternatives and reported ranks are sufficiently accurate, the $\widetilde{z}_i$ estimates coincide with the normalized ranks z_i in the true matrix R. Matrix $\tilde{R}$ is not complete, however, if an alternative m is not ranked by any observer. In this case, formula (6) yields a value of $\widetilde{z}_m = 0.5$ for that alternative, which correctly denotes an absence of information. The formula also generalizes the concept of normalized ranks to situations in which the directed network represented by matrix $\tilde{R}$ contain forks, splits, or cycles: forks and splits result in an incomplete ordering by $\widetilde{z}_i$, and cycles result in ties. Examples are provided in Appendix C. For these reasons, $\widetilde{z}_i$ is the main object of interest in our analysis.

To assess the accuracy of the $\widetilde{z}_i$ estimates, we construct confidence intervals using a cluster bootstrap. The approach is based on the two maintained assumptions behind our methodology, namely that: (1) rankable sets are assigned to observers independently from their preference ordering over all alternatives; and (2) observers have orderings over the full set of alternatives that are independently distributed from their rankable sets. Under these two assumptions, each reported partial ordering r^k can be seen as an i.i.d. realization of a data generating process that randomly samples from the true orderings. This data generating process can therefore be mimicked by drawing with replacement from the set of reported orderings, so as to produce counterfactual samples of orderings. By calculating $\widetilde{z}_i$ estimates for each counterfactual sample, we can approximate the distribution of $\widetilde{z}_i$ estimates produced by the data generating process. This produces confidence intervals for estimated $\widetilde{z}_i$.

2.4 Monte-Carlo simulations

From Tangian (2000)'s theoretical results, we know that we can reliably estimate a population-wide ordering by averaging 'votes' over pairwise alternatives obtained from a large enough sample of observers – provided that the rankable sets of each observer is distributed independently from their ordering over all the available alternatives. What is less clear is how large this sample has to be. Tangian (2000) reports minimal sample size formulas for various parameter values, but in a context in which (1) the number of alternatives is small and each observer ranks all alternatives and (2) observers have different orderings over these alternatives and the object of the estimation is to aggregate observers' orderings to predict the ordering of the entire population. The example discussed in Tangian's is one of a poll of voters used to predict the outcome of an election.

In our setting, observers are asked to rank local households by income level, which implies that all observers ought to have the *same true ordering*. This difference from Tangian's setting should work in our favor in the sense that we are not attempting to recover the average ordering of a potentially diverse voter population. In our setting, all observers are asked to report on the same objective ordering – and they deviate from this ordering only because of observation and reporting error. Consequently, the number of observations on a pairwise comparison that is needed to converge to their true rank should be smaller if observation and reporting error is small. However, our setting also differs from Tangian's in another important way: observers only rank a subset of the available alternatives, not all of them. Because of these two differences, it is unclear how many observers would be required to obtain accurate estimates of the true ordering since this number depends on (1) the proportion of alternatives each observers ranks and (2) the magnitude of the observation and reporting error.

To address this issue, a Monte Carlo simulation analysis is conducted to examine how precisely true normalized ranks z_i are approximated by orderings estimated using the scoring, Borda count, and pairwise method, respectively. We do this under different scenarios regarding the information available to observers, namely: the average number S of target households they can rank; and the variance V of the error they make about other households' incomes. To this effect, we start by generating artificial samples of $\log(y_i)$ realizations by drawing from a standard normal variable with mean 0 and unit variance—which means that the distribution of income y_i is log-normal. We organize these random draws into 100 sets of 30 income realizations, intended to represent households in a village or urban neighborhood. The true normalized ordering of households by income in each set is denoted by vector $z = z_1, ..., z_{30}$.

Each set of households is associated with 9 observers who know a fraction S of them. Each Observer k has an estimate y_i^k of the true income y_i of household i in their location. This estimate is given by:

$$log(y_i^k) = log(y_i) + u_i^k \tag{10}$$

where observation error u_i^k is an i.i.d. random draw from a mean-zero normal variable with standard deviation V.[17]

From the income realizations and observation errors, we construct two reports that each observer makes to the researcher: a vector of income estimates y_i^k for each of the households in their rankable set s_k; and a normalized vector of estimated rankings $z^k = \{z_1^k, z_2^k, ...z_{m_k}^k\}$ for the households in their rankable set, where ranks are based on their observed income y_i^k and z_i^k are computed using formula (3). From these reports we then construct three estimates: (1) a scoring estimate obtained by first averaging the reported income of household i across observers, i.e., $y_i^a = \sum_k y_i^k$, and then calculating i's normalized income rank ω_i^s from 0 to 1 based on that; (2) a Borda count rank obtained by first averaging the normalized rank of household i across observers, i.e., $z_i^a = \sum_k z_i^k$, and then calculating i's 0-to-1 normalized rank ω_i^b based on that;[18] and a pairwise method index z_i calculated according to equation

[17]The important assumption here is that observers k and observer m see the income of individual j with mean-zero errors that are independent (and identically distributed) between k and m. This assumption implies that averaging across observers reduces expected measurement error. Although equation (9) also implies that the errors that observer k makes for individual j and individual i are i.i.d. as well, observer fixed effects could be accommodated when they only report *relative* rankings, as in the Borda and pairwise methods.

[18]By construction, the average of normalized ranks compresses the distribution of ranks towards the middle. In order to compare the Borda count method to the scoring ranks, it has to be resclaled to span the same range before renormalizing the result.

(6), from which we construct a rescaled normalized ordering ω_i^p from 0 to 1.[19] We then calculate the Mean Square Error (MSE) of each estimator $m \in$ {s=scoring, b=Borda count, p=pairwise} as:

$$MSE^m = \frac{1}{3000} \sum_{i=1}^{3000} [\omega_i^m - z_i]^2 \tag{11}$$

The lower bound of MSE^m is 0 when the estimated ordering perfectly matches the true ordering. In our simulation, the measure takes value 0.0832 when all households are tied at the mean rank (i.e., 15.5) and it is equal to 0.333 when the estimated ordering is the reverse of the true ordering. This means that any MSE above 0.0832 is worse than having no information. In all simulations we expect the MSE obtained by the scoring method to perform the best because it relies on richer information provided by observers. But it can diverge from the true ordering due to observation error. Hence the MSE value for the scoring method should be seen as the lowest achievable MSE of the pairwise and Borda count methods relying on the same observers.

We report in Tables 1 and 2 the results from two sets of Monte Carlo simulations. All simulations reported use the same vector of income realizations, so as to eliminate any random sampling noise. Each table reports the three MSE^m for 20 different simulations based on the amount of observation error V—from 0 to 0.9 (to recall, the standard deviation of log(income) is 1)—and the size of the rankable sets S—from 10% to 70% of the set of 30 households in the observer's location set. Each table has two panels. In the right-hand panel, we only report the MSE estimates for ranked alternatives—as well as the total number of ranked alternatives. Given that methods and simulations vary in the fraction of ranked alternatives, we also report in the left-hand panel MSE estimates calculated over the whole sample of 3000 households, so as to facilitate comparison. In this panel, unranked alternatives (ie., households) are assigned the median normalized rank of 0.5—meaning 'no information'.

In Table 1, rankable sets are uncorrelated by construction, which means that each observer has an independent S% probability of observing each of the 30 households in their set. We know that, in this case, the Borda count method is expected to do reasonably well, but we are not sure how the pairwise method will perform relative to it. Results show that, for all three methods, estimated orderings deteriorate with observation error: for $V = 0.9$ and $S = 0.1$, orderings estimated by the pairwise and Borda count methods are both worse than no information, i.e., their MSE exceeds 0.0832. We also note that all methods improve

[19] To ensure full comparability between the three methods, each of the three 1-to-30 orderings is obtained using the same Stata command 'egen ordering=rank(estimate), by(zd)' where 'zd' is the identifier of the set of 30 households. Each of these rescaled 1-to-30 orderings is then normalized using formula (3).

with observer coverage, as could be expected. For instance, with a coverage of 70% and low observation error ($V <= 0.1$), all methods do quite well. We also see that all three methods fail to rank a sizeable proportion of alternatives when coverage is low—e.g., $S = 0.1$ or 0.2. Finally, and more importantly, we note that, in half of the simulations with less observation error, the pairwise method does better than the Borda count method. This is particularly noticeable in the right-hand panel where we ignore unranked alternatives (which add noise to the MSE's).

Earlier in this Section, we have argued that the pairwise method is expected to do particularly well when rankable sets are correlated. To confirm this prediction, we repeat in Table 2 the same set of 20 simulations with maximally correlated rankable sets, in the sense that the set of households observed by each observer k is contiguous in income. In practice, this is achieved by randomly picking an integer U between 1 and $30(1 - S)$ and letting observer k see households U to $\text{int}(U + 30S)$ in the true income ordering. To illustrate, if $S = 0.1$ and $U = 5$ then the set of observed households is those with true ranks r_5, r_6, r_7, r_8. As predicted, we find that the pairwise method does much better than the Borda count method, outperforming it in 18 of the 20 simulations. In some cases, the difference is qualitatively quite large. For instance, when $S = 0.2$ and $V = 0$, the pairwise method yields an MSE of 0.061 (0.043 if only ranked alternatives) while the MSE of the Borda count method well exceeds 0.0832, meaning that it does worse than having no information. The superiority of the pairwise method is maintained even with a high level of observation error, although both methods perform poorly when coverage is very low ($S = 0.1$). In that case, only the scoring method manages to recover a meaningful ordering estimate of 0.074, which is less than 0.0832—and 0.063 when ignoring unranked alternatives.

From this exercise we conclude that the pairwise method has a useful role to play in the estimation of orderings when observers rank different subsets of the available alternatives. Furthermore, the pairwise method tends to outperform the commonly used Borda count method when the rankable subsets of observers are correlated—as is likely to be the case in many empirical applications.

3 First application: Rural Indonesia

We start by validating our method using data on rural Indonesia from Alatas et al. (2012). In this study, the authors collect poverty rankings, from richest to poorest, of, on average, 8.7 randomly selected rural households in each of 640 villages.[20] Each household is ranked

[20]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.

separately by 7.5 observers on average.[21] These observers are the other sample households living in the same village. Since observers are not free to choose which neighbors to rank, there is no self-selection of rankable sets and thus no possibility of correlated rankable sets across observers.

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 reports an income ordering in a given village. 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. To show this, we compute the standard deviation of the ranks reported by different observers around the average rank of a given household. We find it to be small (1.24) relative to the standard deviation of ranks across the entire sample, which is 2.25. Alatas et al. (2012) also 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 pairwise method to the same data in order to obtain estimates of $\widetilde{z}_i$ for each sampled household in each village. From this value, we construct an estimated rank $\hat{r}_i$ by sorting observed households according to $\widetilde{z}_i$ in each village—and taking proper care of ties. We then compare these estimates to estimated ranks similarly obtained by sorting households by their averaged normalized ranks reconstructed from the Alatas et al. (2012) original $\tilde{r}_i$ ranks.[22] To eliminate variation in sample size across Enumeration Areas (EAs), both measures are normalized (i.e., divided) by the total number of ranked households in the village plus 1. We find a correlation of 0.96 between the two measures. 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. The high correlation in aggregate orderings obtained using the Borda count method and our pairwise method matches the theoretical predictions of (Tangian, 2000) when rankable sets overlap across observers, as is the case here.

We then compare (similarly normalized) household rankings based on reported consumption expenditures to those obtained from $\tilde{r}_i$ and $\widetilde{z}_i$. We find that rankings constructed from $\tilde{r}_i$ predict consumption ranks with a R^2 of 0.1456. For the rankings obtained by our ma-

[21]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.

[22]These estimates differ from the original Alatas et al. (2012) measures because we normalize the ranks reported by individual observers before averaging them. This corrects for variation across EAs and observers in the number of households they rank.

trix method, the corresponding R^2 has a value of 0.1426. This shows that, in this case, the two methods, on their own, perform equally well in terms of predicting consumption. We also regress consumption ranks on *both* estimated ranks and find that the R^2 only rises marginally—to 0.1472. 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 nearly as good as rank averaging, even though they are not identical. This is remarkable since, as show in Table 1, the absence of correlated rankable sets gave the Borda count method an advantage when, as here, S is large.

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 different conditions 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 less knowledgeable about their neighbors. As a result, there is less overlap in the set of households that are ranked by individual observers in a particular location. It immediately follows that averaging rankings to construct a $\tilde{r}_i$ index is unlikely to be reliable in the context of our data. But, as we have shown in Section 2, the pairwise approach still applies.

4.1 Sampling frame

We conducted a peer ranking exercise as part of 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.

For the full AUDRI study, the National Statistical Institute (INS)'s enumerations areas (EAs) were used as 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. The EA geographic delimitation as described in the 2014 database was used to infer the total rural and urban population. About 83% of the AUDRI 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 blue areas in Figure A2.

The ranking exercise used for this paper focuses on a subset of the AUDRI EAs, namely 20 urban EAs and 14 rural EAs among those in the AUDRI sample.[23] The 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.[24] These study areas are named *ranking areas* hereafter.

4.2 Household Sampling and Data

Listing and Individual Surveys The AUDRI project undertook two distinct surveys: a household listing survey and a more detailed individual survey for a selected subset from the listing. The household listing survey was conducted in July-August 2019 in all 706 EAs.[25] In the *ranking areas*, for the sake of this study, enumerators were instructed to list the names of up to 14 *consecutive* households.[26] Having been identified, these households were then asked to answer a short listing questionnaire. Not all of them agreed to do so: in total, 207 households across 34 EAs (i.e., about 6 per EA) answered the listing survey, which collected information about each member of the household and basic dwelling characteristics and asset ownership.

From the listing survey, 70% of households were sampled for the individual survey that was conducted between December 2019 and early March 2020. This survey covers a wide range of topics about the individuals' labor activities, transport habits, health conditions, and public service access (4-hour long questionnaire).[27] In ranking areas, 119 individuals participated in this survey, with a completion rate of 84%.[28] The 34 *ranking areas* therefore

[23] We initially selected 20 "semi-rural" EAs and 20 urban EAs among those in the AUDRI sample, but six villages could not be reached by the team of surveyors because the village chief did not allow our enumerators access to the village.

[24] The slum definition we use is that of UN-Habitat 2006, i.e., areas lacking access to improved water, improved sanitation, sufficient living area, durable housing, and secure tenure.

[25] Enumerators started counting households from the centroid (barycentre) of the EA and moved in circles of increasing radii around the centroid, knocking on doors. Only one member of the household (above 18 years old) was surveyed and asked about other members.

[26] 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. In practice, we listed fewer of these in the target number in 8 EAs, due to a combination of logistical difficulties and low population density in rural EAs.

[27] 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.

[28] 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

contain 207 households who responded to the questions necessary to construct their PMT index: 119 from the individual survey, and 88 from the listing survey only. These are the households we wish to rank to compare this survey-based measure of material welfare with the rankings reported by neighbors.

Ranking Survey The ranking survey was administered in the ranking areas in early March 2020. Each respondent was asked to rank at least 5 and up to 14 households in their neighborhood in terms of their material well-being.[29] Half of the sample was also asked to include themselves in the ranking. As in other surveys of relative rankings, we did not tell respondents whether their rankings would be used to target benefits to certain households. In practice, they were not. Respondents were free to include any immediate neighbor they knew by name and enumerators were instructed to identify—and confirm with the respondent—the names of the target households so that they could be matched with the data we collected on them.[30]

A total of 507 respondents answered the ranking survey (see Table A1):

1. *Individual survey 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* among those who 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.[31] Individuals from the individual survey and the listing survey are identified by the letter A in the graphical analysis.

3. *Additional respondents selected on the spot* (N=230): These are members of households

in Dupas et al. 2021.

[29]The precise wording of the question was: "Maintenant que vous avez identifié les ménages que vous reconnaissez [dans notre liste], pourriez-vous me donner un classement de ces ménages, du plus pauvre au plus riche selon vous." (Now that you have identified the households that you recognize [in the list], please rank these households from the poorest to the richest in your opinion).

[30]The SurveyCTO questionnaire helped enumerators scroll through the precoded names of the target households, i.e., those in the listing data. Based on timestamps data, the ranking module took 13 minutes to administer (median). We did attempt to include the names of households other than the target households, in the hope of improving the quality of reconstructed rankings. But there were very few of them and we cannot be sure these names identify the same households.To avoid identification errors, we drop from the ranking analysis those households that could not be matched by the enumerator with a target household.

[31]For logistical reasons, we did not randomly select the gender of the respondent for this group. The survey includes a subset of the modules in the Individual Survey, notably consumption information, and the rankings.

who were identified at the beginning of the listing exercise (i.e., we have their names), but were absent during the listing survey and hence could not be interviewed. We revisited them and, when we found someone available, we administered the ranking questionnaire. These respondents were also asked the questions from the consumption module of the individual surveys, as well as questions needed to construct their PPI index. These households are identified by the letter B 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 instead survey traders operating in the vicinity of the surveyed dwellings, based on the assumption that they could observe the consumption pattern of their customers and thus have some knowledge about the material well-being. Those individuals are identified by the letter C in the graphical analysis.

Some evidence suggests that perceptions of relative poverty may diverge from objective rankings (Cruces et al., 2013). To investigate this possibility in our setting, we asked respondents about: (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. The results, summarized in Table A5, provide some support for this possibility: while 29% of respondents see themselves as poor, only 21% believe they are poorer than their neighbors and only a fifth of them think others view them as poor. Furthermore, among those who self-identify as poor, 53% do not believe others see them that way.

As noted in the introduction, income orderings reported by individuals may reflect different views about what constitutes poverty (e.g., Alatas et al., 2019; Galasso and Ravallion, 2005; Ravallion, 2008). To delve deeper into perceptions of poverty, we asked observers to describe "in their own words" the criteria used to classify households as poor, the vast majority of respondents refer to poverty as "food deprivation" (80%) and some mention "unresolved health problems" (43%). In terms of ranking, 49% take into consideration the occupation of the household head job and another 49% base it on the household's known financial struggles. Over half of the participants report frequently visiting the neighbors they listed and around half reporting having sought financial or health advice from them.

4.3 Poverty measures

Since the purpose of the ranking exercise is to compare rankings reported by neighbors to survey-based measures of material welfare, we collected extensive information on income and wealth, allowing us to compute various poverty measures. There is indeed disagreement in the literature regarding which proxy for material welfare is least subject to measurement error. Some authors believe that recall data on consumption is a good proxy for short-term material welfare while PMT and PPI are of a proxy for long-term welfare. Trachtman et al. (2022), for instance, shows that orderings reported by neighbors in rural Indonesia correlated more with wealth differences than PMT or PPI. To account for this possibility, we examine several poverty measures and proxies in our analysis. For all measures, a lower value indicates *greater* poverty.

Consumption / expenditures Three main household consumption measures were collected on most of the target households[32] and some of the additional respondents:(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[33]. (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 the survey: these include expenses for clothing, shoes, furniture, school fees.[34]

Proxy Means Test (PMT) The government of Côte d'Ivoire introduced its own PMT index 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 predicts the log(food consumption per capita) using about 25 predictors that include assets and house characteristics. We use these same weights to build our PMT index for the target households that responded to the individual survey. Since the PMT index weights were estimated separately for urban and rural EAs, we treat peri-urban EAs in our sample as rural for the purpose of their PMT. We also cross-validated the methodology used by the Government of Côte d'Ivoire with our own data. We obtain a similar fit when predicting log(food consumption per capita) and relatively similar coefficients in terms of magnitude (see the left-hand panel of Table A2). The sample distribution of our PMT estimates is shown in the top panel of Figure A3.

[32] 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

[33] 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.

[34] 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.

Poverty Probability Index (PPI) Innovations for Poverty Action (IPA) introduced the PPI index in April 2018 using Côte d'Ivoire's 2015 living standards measurement study. The PPI aims to predict the probability of the household falling below the National Poverty Line and, unlike consumption expenditure data, it is considered a measure of long-term material welfare that is unaffected by transient shocks in consumption. The PPI is constructed from answers to ten questions covering geographic location, household characteristics, and living conditions—see Table A3 for the full list of questions. We use these same ten questions to construct a PPI index for our sample of households. The sample distribution of the index is shown in the bottom panel of Figure A3. We see that our sample households are widely distributed according to their poverty status—and slightly wealthier than the average household in the country. This is expected given that our sample households live in and around the capital city of the country where the cost of living is higher.[35]

Table A4 shows summary statistics from the various measures, separately for urban and peri-urban (rural) EAs. Figure A4 shows how the measures of poverty are correlated with each other. We see that the PPI index is positively correlated with the PMT index, food consumption, conspicuous consumption, and expenditure on durables.[36]

5 Poverty rankings

5.1 Sample statistics

By design, in each of the 34 EAs 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 households. This represents 91 household pairs per EA, or 3094 pairs in total. We also surveyed 1-3 key informants per area as described previously, and thus we obtained 14.9 observers in each EA on average—507 respondents in total.

Each observer is allowed to rank up to 14 households.[37] 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 pairwise approach. We do not, however, expect all observers

[35]In Table A2, Column (6) and (7), we report the fit from the PPI regression which regresses log(food consumption per capita) on the variables used to build the PPI index. We obtain a reasonably large R^2.

[36]We use all three poverty assessments since there is disagreement in the literature regarding which proxy for material welfare is least subject to measurement error. For instance, Trachtman et al. (2022) shows that orderings reported by neighbors in rural Indonesia correlated more with wealth/consumption differences than PMT or PPI.

[37]15 for the half of the respondents who were asked to rank themselves; 14 for the others.

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. This is also why, in the 34 EAs covered by this study, the 14 target households were selected among close neighbors, in the expectation that this would facilitate rankings. In practice, respondents often refrained to rank even their immediate neighbors, arguing that they did not know enough about them.

Overall, we only collected 1820 distinct pairwise rankings—3.9% of the maximum achievable figure of 46,137.[38] 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 3.6 pairs and the maximum number of ranked pairs is 36.[39] 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. As noted in Section 2, with such low level of overlap across observer rankings, averaging ranks across observers would be very imprecise. We also find that, when a pair of target households is ranked by multiple observers, there is considerable disagreement among them: only 24% of multi-ranked pairs have full agreement among observers (Table A6).

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 pairwise method can recover useful ranking information from such sparse and noisy data.

5.2 Reconstructing ranks using the pairwise method

We now combine the pairwise rankings that we collected to construct poverty ranking estimates using the methodology described in Section 2. The results reported here rely on formula (**??**. Similar, albeit slightly less precise estimates are obtained using formula (**??**).

[38] A pairwise ranking is a report by one observer, say k, who ranks household i relative to household j. Another observer, say m, may also rank household i relative to household j, in which case we have multiple rankings on the same ij household pair.

[39] If we omit self-ranks, the average number of pair rankings by observer is 2.5 and the maximum number of ranked pairs is 28 across the sample.

Ranking graphs for the 34 locations in which relative rankings information was elicited are presented in Appendix C (see Figures C1 and C2 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.[40] 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 at least one (directed) cycle involving a subset of nodes—which are thus all tied together; while the others are transitive.

In Table A7, we report, for the 14 locations with a single component and at least five ranked households, $r_i^{up} \equiv sum_j \eta_{ji}$, the number of households who are ranked richer than household i, and $r_i^{down} \equiv \sum_j \eta_{ij}$, the number of households who are ranked poorer than i We immediately notice that, contrary to equation (**??**) for the complete ranking case, r_i^{up} is not the same as $m - r_i^{down} - 1$. 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 C1), 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 unranked 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:

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

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 cycles/ties. 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 tied. This is clear in the directed graph of relative rankings for location 6 (Figure C1). One extreme case of this is location EA 30, in which all households but two are located on a set of large cycles, meaning that they are all tied: pairwise rankings exist, but they do not induce a transitive aggregate ranking.

As indicated in Section 2, we use a cluster bootstrap to estimate the precision of our rank estimates. We find that index $\widetilde{z}_i$ for each household i varies a lot across simulated samples of observers. This is illustrated in Panel (a) of Figure 1 which shows, for each estimated index value $\widetilde{z}_i$, the distribution of $\widetilde{z}_i$ 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 $\widetilde{z}_i$ is large, reflecting the lack of agreement in rankings among observers. The bimodality in the Figure arises from sampling-with-replacement: some households that are ranked using the full sample end up not being ranked when using only a subset of observers.

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 set of ties, the rankings data is contradictory. Constructed *pairwise* rankings nonetheless remain potentially informative: as explained in Section 2, cycles in the directed graph can be caused by a single misreported link by a single respondent. All the other directed links (i.e., inequality relationships) in a 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 Comparison with other estimators

Next we compare the performance of the pairwise and Borda count methods regarding the precision in our Côte d'Ivoire sample. To facilitate comparison between the two, we turn both indices into normalized ranks. Turning them into ranks eliminates any cardinal difference that may arise between $\widetilde{z}_i$ and the average of Borda counts. It also allows for ties in a comparable manner. Normalization by equation (3), with m_k the number m_k of ranked households in each EA, makes rankings comparable across EAs that have different numbers of ranked households.

We start by noting that both estimates have the same number of unranked households. The coefficient of correlation between the two normalized ranks is 0.867, which is lower than in the Indonesian data. This is unsurprising given that the Borda count method is penalized by the combination of partial ranking with possible correlation in rankable sets. Bootstrapped confidence intervals for the two sets of ranks are shown in Panels (b) and (c) of Figure 1 using 100 bootstrap replications of each of the 34 separate EAs. We see that estimated ranks are imprecise, with the possible exception of households that are ranked either very low or very high, 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 with either of the two methods.

We also compare the performance of our pairwise estimator with an estimator that is commonly used to construct relative rankings based on pairwise comparisons, as arises in sports tournaments. Like our method, this approach relies on pairwise comparisons. But it makes a number of functional assumptions that could potentially increase precision. To investigate this possibility, we applied to the Côte d'Ivoire data the Newman version of the well-known Bradley and Terry (1952) estimator, both with ties and without ties (see Appendix B for details). Since the results are quite similar for the two methods, we focus here on the simpler version without ties. As for Borda counts, we normalize the Newman rank estimates using formula (3) to ensure comparability.

We first note Newman rank estimates are correlated with the ranks based on $\widetilde{z}_i$: the correlation coefficient is 0.731. However, the Newman method only ranks 207 (57%) of the 364 ranked households with the pairwise and Borda count methods. 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 above. Results are shown in Panel (d) of Figure 1. We do not observe any shrinkage in the confidence intervals: if anything, they are much larger than in Panel (b). Based on this, we conclude that, for our data, our estimator does better than the 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.

6.1 Predictive power

We start by regressing the difference in poverty index 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 by the pairwise method.

Since, in the presence of ties, it is possible that j is ranked richer than i *and* i is ranked richer than j, all differences in outcomes are taken as j's value minus i's value to net them. Results are presented in Table 3. 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 index; and (4) PPI index.[41] Except for food shortages, these measures are constructed such that a positive 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, then j's consumption, PMT and PPI indices 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 sometimes have 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 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 household. Here the dependent variable is the consumption level. Results, presented in Table 4, show that estimated coefficients are not statistically significant. These results are perhaps not surprising, given the findings from Table 3—and the fact that the information content of P_i is more affected by the presence of ties than reported ranks. We also see from Table A9 that Borda counts are not better at predicting consumption and wealth: if anything, the use of Borda count estimates produces smaller R^2 values.

To understand why individuals do not seem to make accurate rankings, we examine which household characteristics are predictive of reported ranks. The results are shown in Table 5. Regarding the PPI index, the pattern is consistent with expectations: if i has a lower PPI index than j, k is more likely to report that i as poorer than j. For consumption

[41] In Table A8, we run the same regression for an extensive set of additional outcomes, e.g., different consumption variables, shortfall in consumption, and an index of improvement in consumption relative to last year.

variables, j is ranked richer than i (positive coefficient) if j reports higher food consumption or higher spending on durables. We see that the variables "months of food shortages", "expressed food worries in the last 12 months", and"received gifted food in the past week" consistently predict that an observer ranks a household as poorer. Contrary to our initial expectations, we do not find that conspicuous consumption expenditures—as measured in our survey—predict reported rankings.

6.2 Ranking Accuracy

Next, we examine the accuracy of reported ranks by comparing them with rankings obtained using the PMT and PPI indices constructed with the survey data and 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% compared to PPI rankings.[42] Ranking accuracy is even lower if we use food expenditure per capita as comparison.

These average levels of accuracy are only barely above what could be achieved by random guessing. Could this be because the indices themselves are very noisy? To investigate this possibility, we examine whether PMT and PPI predict consumption per capita rankings well. We find that they do: 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. When the PPI indices or food consumption levels of i and j are close, 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 around 58 to 60%).[43] It follows that, with a sufficiently large number of reports on such ij pairs, our methodology should yield a correct ranking with a reasonably high probability. To illustrate, if we assumed that the probability that each k respondent ranks correctly an ij pair is 60%, the probability of mis-ranking would fall from 40% for a single report to 35.2%, 31.7%, and 24.7% for 3, 5 and 11 reports, respectively.[44] This only applies

[42]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.

[43]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.

[44]Calculations 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.

to ij pairs that are sufficiently 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.[45] To conduct this counterfactual experiment, we again use as 'truth' the PMT and PPI indices constructed from the data. 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.[46]

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' material welfare of individuals i and j. But it gives an idea of the magnitude of the observation errors 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 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 either their PPI, PMT, or household food expenditure per capita. We then regress this variable on observer characteristics. The results are shown in Table 6.[47] We do not find any convincing evidence that observer characteristics predict ranking accuracy, leaving little scope for improving the estimation by over-sampling observers

[45] 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.

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

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

with certain characteristics.

6.3 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 7, 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.

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 7 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.

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 (see, however, Cruces et al. (2013)). 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 2021). 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 7 excluding self-ranks from the data. Results, shown in Table A10, show that dropping self-ranks improves targeting accuracy somewhat. Those who rank below 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). These differences are significant at the 10% and 5% level, respectively, and they are larger than those reported in Table 7 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 (e.g., Cruces et al., 2013). 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.[48] The experiment is not designed to identify which is the most likely explanation.

Results, presented in Table 8, 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

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

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 8). 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 9.

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.[49]

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 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 based on their PPI 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

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

poor.

9 Conclusion

This paper introduced a new method for aggregating partial poverty rankings reported by multiple observers. The method we propose relies on the same data as the commonly applied Borda count method, but it averages pairwise ranks instead of reported ranks. We show that the method outperforms the Borda count method when observers only rank a subset of target households and the subset of households they rank are more similar than random—a phenomenon we call correlated rankable sets.

We then compare the performance of our pairwise method to that of the Borda count method in two datasets. In the first dataset coming from the rural Indonesia setting of Alatas et al. (2012), the overwhelming majority of observers all rank the same households. In such setting, the pairwise approach offers no a priori advantage but is it outperformed by the Borda method? We find that it does not: normalized rankings from both methods are highly correlated and they are equally good at predicting normalized income rankings based on survey data.

In the second data, the urban setting of Abidjan, observers appear less knowledgeable of their neighbors and they all report partial rankings—a situation which raises the possibility of correlated rankable sets. Due to incompleteness in reported rankings, both the pairwise method and the Borda count method fail to rank a large proportion of target households. For those they are able to rank, however, the two method produce highly correlated normalized rankings. These rankings, however, are not precisely estimated. A more accurate picture would require increasing the density of reporting. The pairwise and Borda method do, however, outperform a commonly used pairwise ranking algorithm used for sports competition: not only does this algorithm rank much fewer target households, it also produces rank estimates that are much less precisely estimated.

We also find that, in the Abidjan data, pairwise rankings reported by respondents are only mildly correlated with various poverty measures collected on households targeted by the ranking exercise. 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—but does not require the extensive overlap in comparison sets across observers that the Borda count method 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.

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 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.

Baland, Jean-Marie, Catherine Guirkinger, and Charlotte Mali, "Pretending to be poor: Borrowing to escape forced solidarity in Cameroon," *Economic development and cultural change*, 2011, *60* (1), 1–16.

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.

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.*

Burke, Marshall, Anne Driscoll, Stefano Ermon, and David Lobell, "Using satellite imagery to understand sustainable development," 2021, *371.*

Cruces, Guillermo, Ricardo Perez-Truglia, and Martin Tetaz, "Biased Perceptions of Income Distribution and Preferences for Redistribution: Evidence from a Survey Experiment," *Journal of Public Economics*, 2013, *98*, 100–112.

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

Diamond, Alexis, Michael Gill, Miguel Rebolledo Dellapiane, Emmanuel Skoufias, Katja Vinha, and Yiqing Xu, "Estimating Poverty Rates in Target Populations: An Assessment of the Simple Poverty Scorecard and Alternative Approaches," Technical Report, Poverty Equity Global Practice Working Paper 080, The World Bank 2016.

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

_ , _ , and Laura Hernandez-Nunez, "Keeping Up Appearances: An Experimental Investigation of Relative Rank Signaling," Technical Report, National Bureau of Economic Research 2024.

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

_ , 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, *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.

Gehrlein, W. V., "Condorcet's paradox," *Theory and Decision*, 1983, *15*, 161–197.

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.

_ , Phillippe Leite, Matthew Wai-Poi, and Emil Tesliuc, *Revisiting Targeting in Social Assistance: A New Look at Old Dilemmas. Human Development Perspectives*, The World Bank, 2022.

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.

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.

List, C., "Social Choice Theory," in E. N. Zalta U. Nodelman, ed., *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University, Winter 2022.

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

Premand, Patrick and Oumar Barry, "Behavioral change promotion, cash transfers and early childhood development: Experimental evidence from a government program in a low-income setting," *Journal of Development Economics*, 2022, *158*, 102921.

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.

_ , "Poor, or Just Feeling Poor? On Using Subjective Data in Measuring Poverty," in "Happiness and Economic Growth: Lessons from Developing Countries," Oxford University Press, 11 2014.

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

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

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

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.

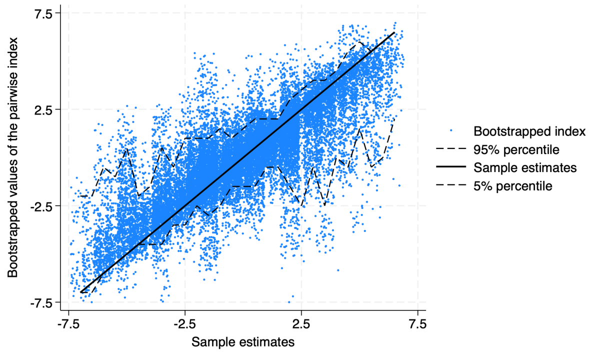
Tangian, A. S., "Unlikelihood of Condorcet's paradox in a large society," *Social Choice and Welfare*, 2000, *17*, 337–365.

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.

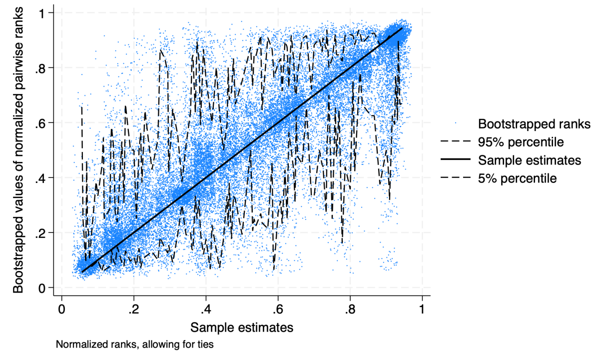
UN-Habitat, "State of the World's Cities 2006 and 2007.," *Nairobi: United Nations Human Settlements Programme.*, 2006.

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

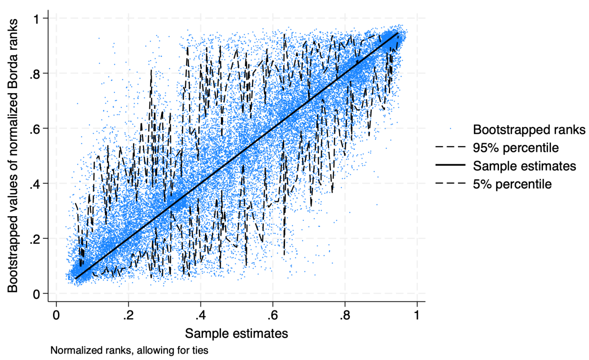
Figure 1: 90% Confidence Intervals Across Methods



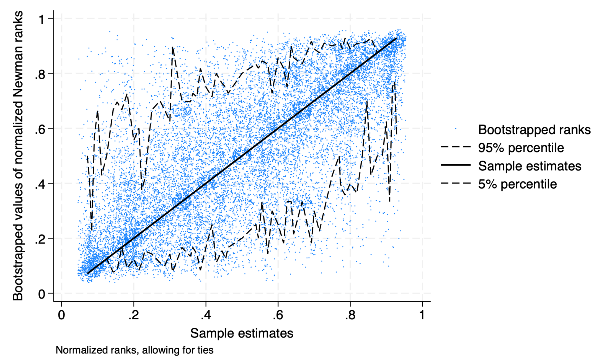
(a) Estimated Pairwise Index



(b) Normalized pairwise ranks



(a) Normalized Borda ranks



(b) Normalized Newman ranks

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

Table 1: Mean Square Error - uncorrelated rankable sets

		Using all alternatives Size of rankable set S =				Using only ranked alternatives Size of rankable set S =							
Estimator:	V =	0.1	0.2	0.4	0.7	0.1	N	0.2	N	0.4	N	0.7	N
Pairwise	0	0.059	0.020	0.001	0.00009	0.042	1,801	0.009	2,579	0.000	2,969	0.00001	2,999
Borda count		0.060	0.026	0.004	0.00021	0.046	1,849	0.016	2,580	0.002	2,969	0.00013	2,999
Scoring		0.041	0.015	0.001	0.00006	0.015	1,849	0.003	2,580	0.000	2,969	0.00000	2,999
Pairwise	0.1	0.059	0.018	0.003	0.00087	0.043	1,847	0.010	2,623	0.002	2,973	0.00087	3,000
Borda count		0.061	0.024	0.004	0.00093	0.047	1,905	0.016	2,623	0.004	2,973	0.00093	3,000
Scoring		0.040	0.012	0.002	0.00060	0.014	1,905	0.003	2,623	0.001	2,973	0.00060	3,000
Pairwise	0.3	0.063	0.026	0.013	0.00541	0.049	1,847	0.019	2,623	0.012	2,973	0.00541	3,000
Borda count		0.064	0.029	0.008	0.00316	0.052	1,905	0.022	2,623	0.008	2,973	0.00316	3,000
Scoring		0.044	0.017	0.005	0.00263	0.022	1,905	0.009	2,623	0.005	2,973	0.00263	3,000
Pairwise	0.5	0.072	0.040	0.038	0.01663	0.064	1,847	0.034	2,623	0.037	2,973	0.01663	3,000
Borda count		0.071	0.037	0.014	0.00660	0.063	1,905	0.031	2,623	0.014	2,973	0.00660	3,000
Scoring		0.052	0.025	0.011	0.00610	0.034	1,905	0.018	2,623	0.010	2,973	0.00610	3,000
Pairwise	0.9	0.090	0.072	0.066	0.04783	0.094	1,847	0.066	2,623	0.065	2,973	0.04783	3,000
Borda count		0.085	0.056	0.031	0.01574	0.086	1,905	0.053	2,623	0.030	2,973	0.01574	3,000
Scoring		0.070	0.045	0.028	0.01731	0.062	1,905	0.040	2,623	0.027	2,973	0.01731	3,000

Notes: This Table reports the Mean Square Error of the pairwise, Borda count, and scoring methods applied to the same simulated data. Simulations are based on 100 sets of 30 alternatives ranked by 9 observers per set—3000 alternatives in total. V is the standard deviation of the noise added to the 'true' value of a log(income) variable with mean zero and unit variance. Rankable sets are uncorrelated by construction: each observer sees each of the 30 alternatives with equal probability S. It follows that S is the average share of the 30 alternatives ranked by each observer. N is the number of ranked alternatives. We fix the random seed across simulations to ensure that the realizations of income are identical across all parameter values. The Mean Square error is calculated as the square of [(estimated rank minus the true rank) divided by 30]—where estimated rank and true rank are both a number from 1 to 30 and the division by 30 is used to normalize the MSE estimates. With 30 alternatives, the MSE of no information (all ranks tied at 15.5) is 0.0832 and the MSE of the reverse ranking (the worst possible outcome) is 0.333. MSE estimates increase as we move down (more observation noise) and left (less coverage). The pairwise method and the Borda count methods rely on the identical reported rank data. We show in yellow those simulations in which the pairwise method performs better than the Borda method, and in green those in which the Borda method performs better than the pairwise method. The scoring method relies on income reported, possibly with error, by each observer. Reported incomes are averaged across observers within each set to compute ranks in a set. The scoring method performs better than the pairwise and Borda count methods, but it requires observers to report quantitative income data, not just ranks. In the left-hand panel, MSE calculations includes all alternatives. Unranked alternatives receive a median rank. This serves to compare methods that generate differences in the number of ranked alternatives. In the right-hand panel, MSE's are calculated using ranked alternatives only. There is no difference between the two panels when the number of ranked alternatives is 3000 (the maximum).

Table 2: Mean Square Error - correlated rankable sets

		Using all alternatives Size of rankable set S =				Using only ranked alternatives Size of rankable set S =							
Estimator:	V =	0.1	0.2	0.4	0.7	0.1	N	0.2	N	0.4	N	0.7	N
Pairwise	0	0.127	0.061	0.022	0.00713	0.136	2,111	0.043	2,563	0.004	2,737	0.00076	2,914
Borda count		0.136	0.102	0.034	0.00752	0.149	2,111	0.091	2,563	0.017	2,737	0.00120	2,914
Scoring		0.036	0.028	0.020	0.00700	0.012	2,111	0.006	2,563	0.004	2,737	0.00075	2,914
Pairwise	0.1	0.135	0.070	0.023	0.00825	0.148	2,079	0.053	2,537	0.005	2,734	0.00150	2,908
Borda count		0.143	0.109	0.038	0.00903	0.159	2,084	0.099	2,537	0.021	2,734	0.00221	2,908
Scoring		0.040	0.029	0.020	0.00750	0.015	2,084	0.008	2,537	0.005	2,734	0.00131	2,908
Pairwise	0.3	0.145	0.088	0.031	0.01058	0.163	2,061	0.072	2,535	0.013	2,734	0.00397	2,908
Borda count		0.152	0.125	0.052	0.01191	0.171	2,084	0.119	2,537	0.037	2,734	0.00505	2,908
Scoring		0.045	0.033	0.022	0.00926	0.023	2,084	0.014	2,537	0.007	2,734	0.00303	2,908
Pairwise	0.5	0.150	0.101	0.048	0.01875	0.171	2,057	0.088	2,533	0.029	2,734	0.01140	2,908
Borda count		0.155	0.135	0.070	0.01585	0.176	2,084	0.131	2,537	0.057	2,734	0.00898	2,908
Scoring		0.053	0.041	0.028	0.01219	0.034	2,084	0.022	2,537	0.013	2,734	0.00611	2,908
Pairwise	0.9	0.154	0.120	0.076	0.05172	0.176	2,050	0.109	2,529	0.055	2,734	0.04311	2,908
Borda count		0.157	0.145	0.099	0.02805	0.179	2,084	0.143	2,537	0.088	2,734	0.02159	2,908
Scoring		0.074	0.061	0.045	0.02395	0.063	2,084	0.046	2,537	0.031	2,734	0.01776	2,908

Notes: This Table reports the Mean Square Error of the pairwise, Borda count, and scoring methods applied to the same simulated data. Simulations are based on 100 sets of 30 alternatives ranked by 9 observers per set—3000 alternatives in total. V is the standard deviation of the noise added to the 'true' value of a log(income) variable with mean zero and unit variance. Rankable sets are correlated by construction: each observer sees S*30 consecutive alternatives, with alternatives ranked by their true income value. A set has 30*(1-S) different sequences of consecutive alternatives. Each of the 9 observers is randomly assigned, with equal probability, to one of these sequences. It follows that S is the share of the 30 alternatives that is ranked by each observer. In the Table, N is the number of ranked alternatives. We fix the random seed across simulations to ensure that the realizations of income are identical across all parameter values. The Mean Square error is calculated as the square of [(estimated rank minus the true rank) divided by 30]—where estimated rank and true rank are both a number from 1 to 30 and the division by 30 is used to normalize the MSE estimates. With 30 alternatives, the MSE of no information (all ranks tied at 15.5) is 0.0832 and the MSE of the reverse ranking (the worst possible outcome) is 0.333. MSE estimates increase as we move down (more observation noise) and left (less coverage). The pairwise method and the Borda count methods rely on the identical reported rank data. We show in yellow those simulations in which the pairwise method performs better than the Borda method, and in green those in which the Borda method performs better than the pairwise method. The scoring method relies on income reported, possibly with error, by each observer. Reported incomes are averaged across observers within each set to compute ranks in a set. The scoring method performs better than the pairwise and Borda count methods, but it requires observers to report quantitative income data, not just ranks. In the left-hand panel, MSE calculations includes all alternatives. Unranked alternatives receive a median rank. This serves to compare methods that generate differences in the number of ranked alternatives. In the right-hand panel, MSE's are calculated using ranked alternatives only. There is no difference between the two panels when the number of ranked alternatives is 3000 (the maximum).

Table 3: Relative Rank vs. Relative Survey-Measured Poverty: Pair-Level

$j - i$ **difference in:**	(1) Food exp per ca	(2) Months food short	(3) PMT index	(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
Number of bservations	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
Number of bservations	1014	951	570	1099

Notes: We report two 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) by the pairwise method. A negative difference indicates that j is ranked poorer than i. All dependent variables are calculated as the value for household j minus the value for household i: a positive difference means that i is poorer than j. *Food expenditures:* Total of consumption expenditures on staples, meat, vegetables, fruits, drinks, and alcohol reported for a one week recall period: *Months food short:* Number of months the household experienced a food shortage over the last twelve months. PPI and PMT Scores are computed following the methodology described in Section 4.3. Robust standard errors are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Aggregate Rank vs. Survey-Measured Poverty Level : Individual Level

Levels of:	(1) Food exp per ca	(2) Months food short	(3) PMT index	(4) PPI index
P 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

Notes: We report a regression for the constructed variable of aggregated rankings, as described in Section 6. The relative position of household i is variable P in equation (2.3.2), that is, the difference between the number of households below i minus the number of households above i. The difference is 0 when all households a tied, meaning no one is ranked above anyone else. A high score means the individual is ranked richer. All dependent variables are levels of consumption variables. *Food expenditures:* Total of consumption expenditures collected with a one week recall period: staples, meat, vegetables, fruits, drinks, alcohol. *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 . *Months food short:* Number of months the household experienced a food shortage over the last twelve months. *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 respondenteś household has improved relative to previous year. PPI and PMT Scores are computed following the methodology described in Section 4.3. Only the surveyed respondents for which we recovered a rank are included in this regression. Robust standard errors are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Predictors of rankings: Pairwise comparisons

	Dep. var. =1 if resp. k reported that i is poorer than j		
	(1)	(2)	(3)
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 per capita in the last week		0.010***	
		(0.004)	
Value of conspicuous consumption expenditures in the last month		-0.006	
		(0.004)	
Spending on durables per capita in the last year		0.018	
		(0.019)	
Received gifted food last week (yes=1)			-0.118***
			(0.043)
Food worries during last 12 months (yes=1)			-0.044
			(0.034)
Months with food shortages in the last year			-0.004
			(0.006)
Days with skipped meals in last 3 months			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 is a triad. The outcome variable is an indicator equal to 1 if k ranked j poorer than i and 0 otherwise. It is missing k did not rank j relative to i. ij pairs including the respondent k are dropped. The three columns contain different types of regressors, e.g., assets, expenditures, and experienced poverty. Each regressor is the difference in the value of the variable between j and i. Any 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 6: Predictors of ranking accuracy

	PMT index (1)	PPI index (2)	Food expenditure per capita (3)
Respondent is a woman	0.044	-0.079	0.091*
	(0.056)	(0.054)	(0.052)
Respondent is a migrant	0.022	0.003	-0.068
	(0.059)	(0.057)	(0.055)
Respondent is Non-Ivorian	0.022	-0.027	-0.012
	(0.060)	(0.059)	(0.054)
Respondent is an informant	0.063	0.080	-0.062
	(0.085)	(0.095)	(0.065)
Respondent a hh head	0.036	-0.025	0.092
	(0.057)	(0.056)	(0.056)
Semi-Rural EA dummy	-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)
Food expenditure p.c. in the last week	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	-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)
Belongs to a 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: Each column is a regression of the propensity for an observer k to rank accurately two households i and j according to the variable listed at the top. The number of observations is the number of observers for whom we can 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 indices. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Can peer rankings identify those below the median?

	Ranked Below the Median According to:		
	PMT Index	PPI Index	Food Expenditure Per Capita
Panel A.			
Household ranked below median by pairwise method	0.105	0.019	-0.006
	(0.085)	(0.053)	(0.060)
Constant	0.395***	0.411***	0.453***
	(0.033)	(0.029)	(0.026)
Observations	172	311	311
Panel B.			
Household not ranked by pairwise method	0.167	-0.023	0.035
	(0.101)	(0.056)	(0.054)
Constant	0.444***	0.459***	0.469***
	(0.013)	(0.014)	(0.013)
Observations	207	507	507

Notes: In the first two panels, we create a dummy equal to 1 if the aggregate pairwise 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 by the pairwise method. We construct the aggregated ranking as the relative position of individual i in the constructed network, i.e., how many people can be ranked as poorer than i, minus how many can be ranked richer. Only individuals ranked by at least one other respondent are considered. We run OLS regressions of this dummy on a dummy for whether the household is below the median based on one of the survey measures (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 of the aggregate peer ranking are 10.1 percentage points more likely to be below the median of the PMT score (Column 1). In the bottom panel, the dependent variable is a dummy equal to 1 if the household was not ranked by any observer. Robust standard errors are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Testing for Self-Ranking Bias

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ranked i poorer than j						
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
Sample Restriction	*30 EAs*	*All 34 EAs*	*Bottom Cons*	*Bottom PMT*	*Bottom PPI*	*Men*	*Women*

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 the bottom 50% within their EA across different wealth measures. Sample sizes vary because of ties and missing observations for some respondents who were not asked the PMT questions, as described in the data section. 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 9: 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

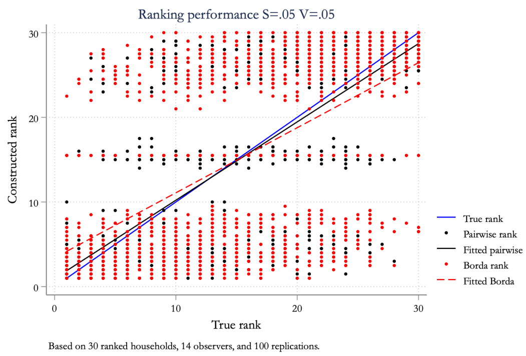
Notes: The unit of observation at the dyad level. The outcome variable is a dummy equal to 1 if k report a ranking for the individual i, 0 otherwise. Pairs i-k involving the respondent k are dropped. Missing distance is replaced by the average distance in the EA and we control for such a case in the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

Online Appendix

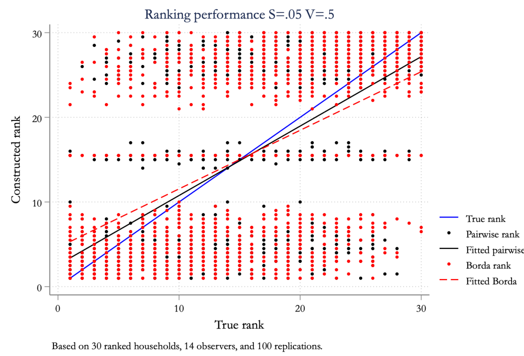
Table of Contents

A Additional Figures and Tables

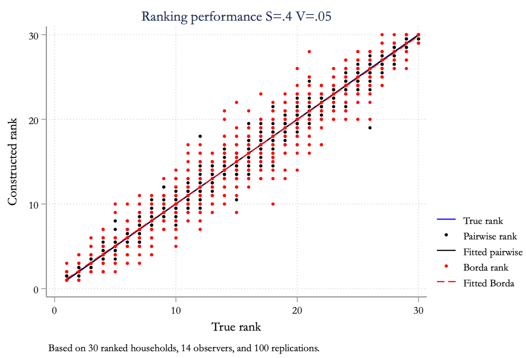
Figure A1: Simulated Performance of Ranking Methods



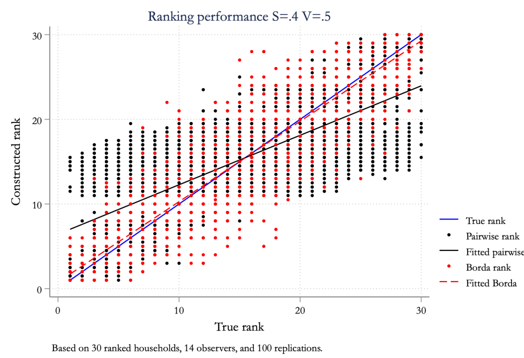
(a)



(b)



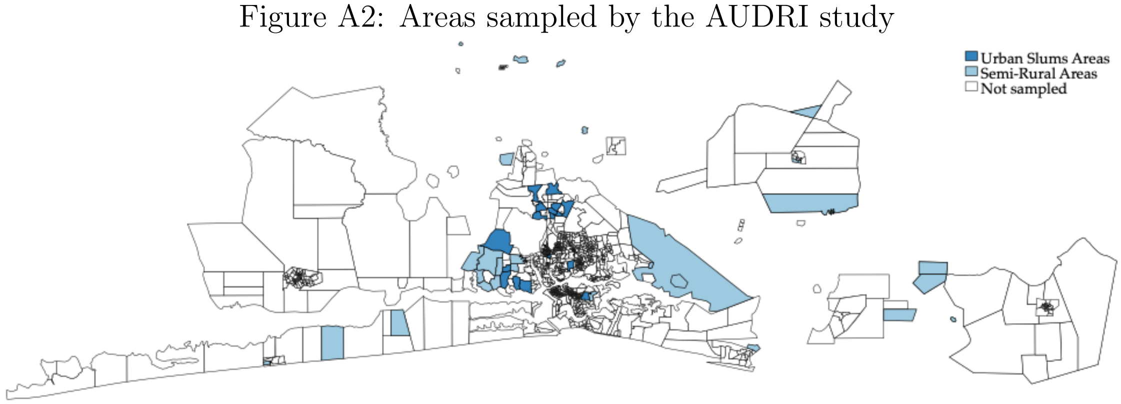
(c)



(d)

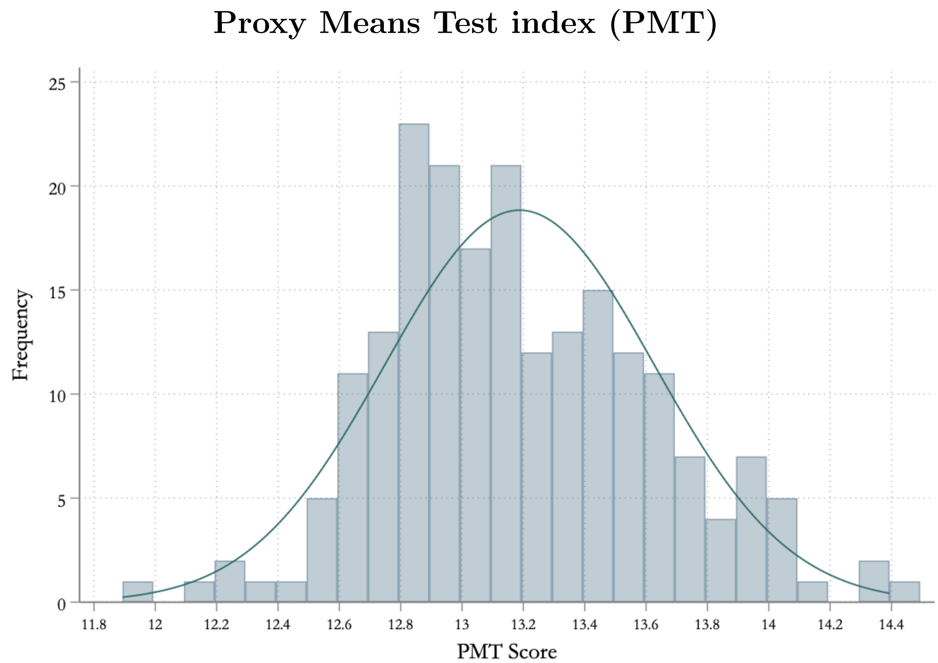
Notes: These Figures show the simulated performance of the pairwise and Borda count methods. Each Figure shows results for a given average proportion S of the sample that is observed by each observer and a given variance V of the mistakes observers make in estimating other households' income. The distribution of true income is log-normal with variance 1. The number of observers and ranked households is the same across the Figures and shown under each of the them. The number of simulation replications are indicated below each Figure.

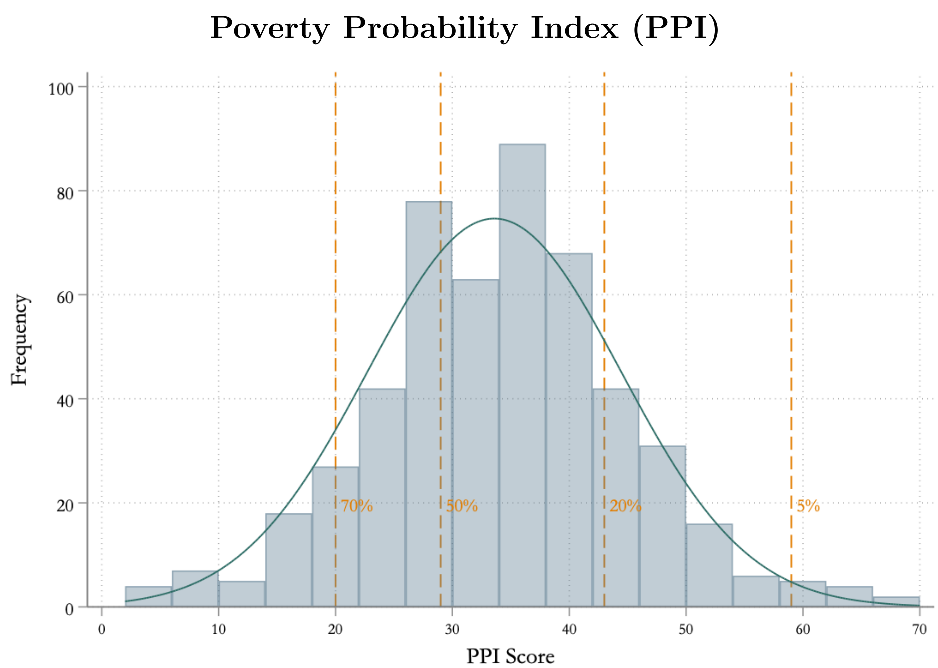
Figure A2: Areas sampled by the AUDRI study



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

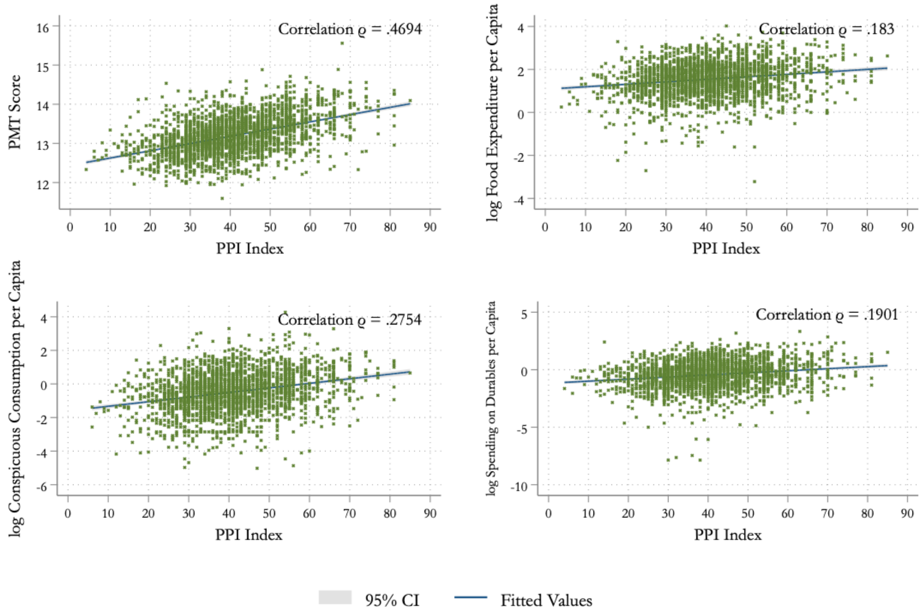
Figure A3: Sample distribution of the PMT and PPI indices in the AUDRI Abidjan sample





Notes: We plot the distribution of the PPI (poverty probability index) and PMT (proxy mean test index). In the top Figure, we show the PPI in our sample, using weights estimated by Innovations for Poverty Action (IPA) in April 2018 on the basis of Côte d'Ivoire's 2015 Enquête sur le Niveau de Vie des Ménages (Household Living Standard Measurement Survey). The "Poverty Likelihood", i.e., the probability to be below the National Poverty Line, is indicated in orange. The bottom Figure shows the PMT index developed by the Ivorian Government. The two indices are described in more details in Section 4.3.

Figure A4: Correlation across poverty measures in the Abidjan sample



Notes: Each Figure shows the sample correlation between pairs of the poverty measures discussed in Section 4.3. To increase power, we use the full AUDRI study sample, irrespective of whether they participated in the ranking exercise or not.

Table A1: Breakdown of the sample of observers by origin

Sample origin:	**#**	**%**	**% of Household Head**	**% of Women**
Individual survey (A)	119	23.47%	47.06%	56.30%
Listing survey only (A)	88	17.36 %	38.64%	69.32%
Selected on the spot (B)	230	45.36%	46.52%	46.96%
Informants (C)	70	13.81%	25.71%	71.43%
Total	**507**	**100%**	**42.41%**	**56.41%**

Notes: In EAs selected for the ranking exercise, we sought to interview all respondents to the Individual survey (December 2019 to March 2020) and to the Listing survey (July to August 2019). The Table shows those who could be found and surveyed for the ranking exercise. A number of additional households were recruited on the spot as observers to increase sample size. Informants were also recruited on the spot among traders and shopkeepers operating in the area.

Table A2: Estimated Fit of the PMT and PPI indices w.r.t. log(consumption per capita)

	PMT Index					PPI Index	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	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

Notes: 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. Column 4 reports the R^2 from the same regression run on the households involved in the ranking exercise while Column (5) includes 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: PPI Scorecard for the Côte d'Ivoire 2015 National Poverty Line

Indicators	**Responses**	**Points**
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 Index		Sum of points

Notes: The points provided here are those for the National Poverty Line.

Table A4: Summary Statistics - Poverty Measures

	Urban (1)	Rural (2)
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 expenditures (concert, bar, cinema, games)	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)
- HH expenditures on clothes/shoes	1.10	1.34
	(1.09)	(1.48)
- HH expenditures on furniture	0.39	0.27
	(1.16)	(0.84)
- HH expenditures on school fees	0.89	0.52
	(2.80)	(1.35)
Value of food expenditure per capita in the last week	3.76	3.74
	(2.89)	(3.39)
Spending per capita on durables in the last 12 months	0.53	0.61
	(0.67)	(1.46)
Indices		
PMT Index	13.23	13.13
	(0.43)	(0.44)
PPI Index	37.13	28.70
	(9.55)	(10.64)
Other variables		
HH's head unemployed or inactive	0.20	0.15
	(0.40)	(0.36)
Number of mobile phones per capita	0.81	0.82
	(0.47)	(0.48)
Number of observations	294	213

Notes: Consumption expenditures reported in 1,000 CFA per week.

Table A5: How do Respondents Think About Poverty?

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

Notes: Respondents to the March 2020 ranking survey were asked to define poverty in their own words. Their answers were subsequently turned into categories by the research team. The Table displays the proportion of respondents who mention each of the listed items as associated with poverty.

Table A6: EA-level summary statistics in the Côte d'Ivoire ranking exercise

	(1)	
Number of intended targets per EA	14.00	
	(0.00)	
Number of observers per EA	8.65	
	(4.26)	
Overlap in coverage between observers >= 2	0.31	
	(0.23)	
Overlap in coverage between observers = 1	0.15	
	(0.16)	
Overlap in coverage between observers = 0	0.54	
	(0.23)	
% with full Agreement on ij pairs	0.24	
	(0.43)	
Observations (EAs)	34	1

: The extent of overlap in coverage between observers refers to the share of target households that are ranked by at least 2 observers, 1, or none. Full agreement across observers is the proportion of responses in which both observers state that i¡j among all pairwise ranks reported by individual observers.

Table A7: 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												

Notes: 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 A8: 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

Notes: We report two 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). A negative difference indicates that j is ranked poorer than i. All dependent variables are calculated as the value for household j minus the value for household i: a positive difference means that i is poorer than j. *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. Robust standard errors are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: Do estimated ranks predict survey-measured poverty?

Levels of:	(1) Food exp per ca	(2) Months food short	(3) PMT index	(4) PPI index
Ranks obtained by pairwise method	0.015	-0.032	0.012*	0.167
	(0.050)	(0.051)	(0.007)	(0.162)
log(distance)	0.019	-0.117	0.010	-0.283
	(0.092)	(0.080)	(0.015)	(0.348)
Constant	3.824***	2.019***	13.114***	35.315***
	(0.585)	(0.528)	(0.088)	(1.972)
R2	0.001	0.011	0.058	0.013
Observations	223	215	155	223

Levels of:	(1) Food exp per ca	(2) Months food short	(3) PMT index	(4) PPI index
Ranks obtained by Borda count	0.464	0.011	0.146	3.914
	(1.037)	(0.666)	(0.111)	(2.745)
log(distance)	0.019	-0.121	0.012	-0.279
	(0.092)	(0.080)	(0.015)	(0.347)
Constant	3.600***	2.056***	13.025***	33.379***
	(0.630)	(0.622)	(0.089)	(2.212)
R2	0.002	0.008	0.040	0.017
Observations	223	215	155	223

Notes: Each column is a regression of a survey-measured poverty variable on estimated ranks. The log(distance) is used as control. The top panel uses ranks obtained using the pairwise method, the bottom panel uses ranks obtained using the Borda count method, as described in Section 2.3. Only the surveyed respondents for which we estimated a rank are included in these regressions. A high relative rank means the individual is ranked richer. The dependent variables are as follows: *Food expenditures:* Total of consumption expenditures on staples, meat, vegetables, fruits, drinks, alcohol over a one week recall period: . *Months food short:* Number of months the household experienced a food shortage over the last twelve months. PPI and PMT indices are computed following the methodology described in Section 4.3. For comparability purposes, the estimation samples are restricted to individuals ranked by both methods. Robust standard errors are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A10: Can peer rankings identify those ranked the median? Omitting self-ranks)

	Ranked Below the Median According to:		
	PMT Index	PPI Index	Food Expenditure Per Capita
Panel A.			
Household ranked below median by pairwise method	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.			
Household not ranked by pairwise method	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: In the first two panels, we create a dummy equal to 1 if the aggregate pairwise 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 by the pairwise method. We construct the aggregated ranking as the relative position of individual i in the constructed network, i.e., how many people can be ranked as poorer than i, minus how many can be ranked richer. Only individuals ranked by at least one other respondent are considered, and we here exclude the self-ranks. We run OLS regressions of this dummy on a dummy for whether the household is below the median based on one of the survey measures (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 of the aggregate peer ranking are 12.7 percentage points more likely to be below the median of the PMT score (Column 1). In the bottom panel, the dependent variable is a dummy equal to 1 if the household was not ranked by any observer. Robust standard errors are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B The Newman Algorithm

In the empirical part of the paper, we compare how our proposed pairwise method performs to another pairwise algorithm commonly used, outside of economics, to rank alternatives when only some pairs have been ranked relative to each other. Of particular relevance to us is the Zermelo (1929) algorithm, which 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, which should thus be seen as a parametric estimator. Here we use an improvement to the algorithm introduced by Newman (2022) because of its simplicity and speed.

This algorithm works as follows. 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:

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

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:

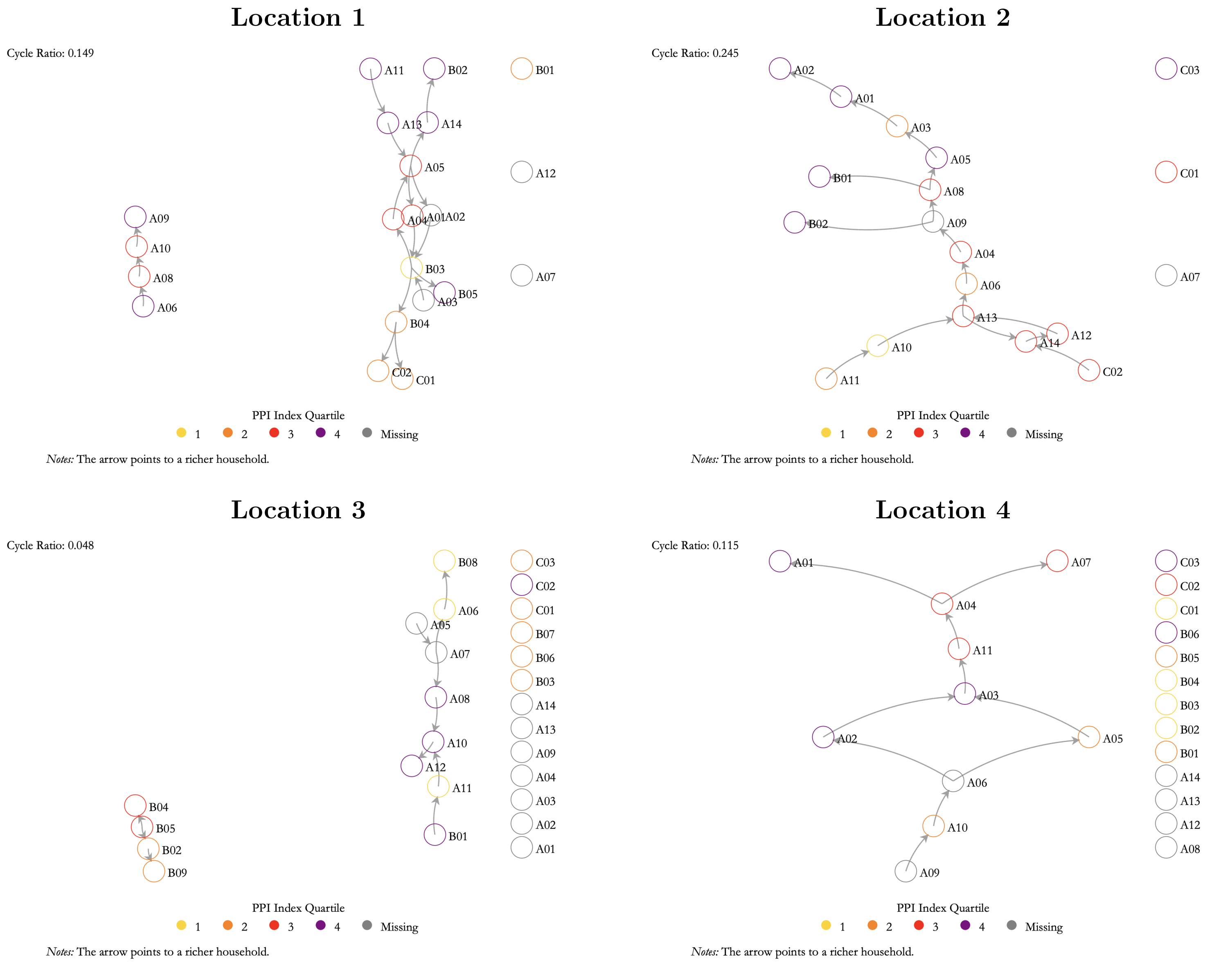
$$\pi_i' = \frac{\sum_j w_{ij}\pi_j/(\pi_i + \pi_j)}{\sum_j w_{ij}/(\pi_i + \pi_j)} \tag{13}$$

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. This algorithm, however, fails to rank all the alternatives when coverage is too sparse. It also does not provide a measure of precision of the resulting estimates.

C Network Figures

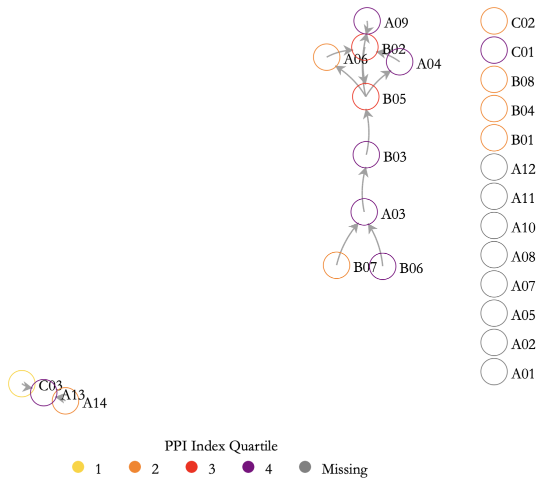
Targeted households (i.e., those on which we have listing or survey information) are identified in the graphs by the letter A. Households that were added as external observers for the sole purpose of the ranking exercise are identified in the graphs by the letter B. Key informants are identified by the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

Figure C1: Directed graph of relative rankings - Urban Slums



Location 5

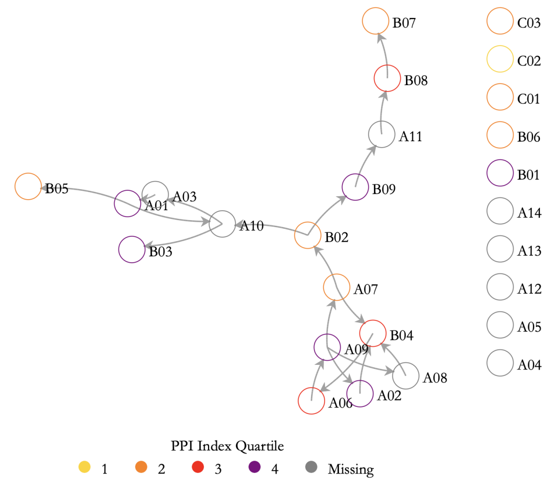
Cycle Ratio: 0.093



Notes: The arrow points to a richer household.

Location 6

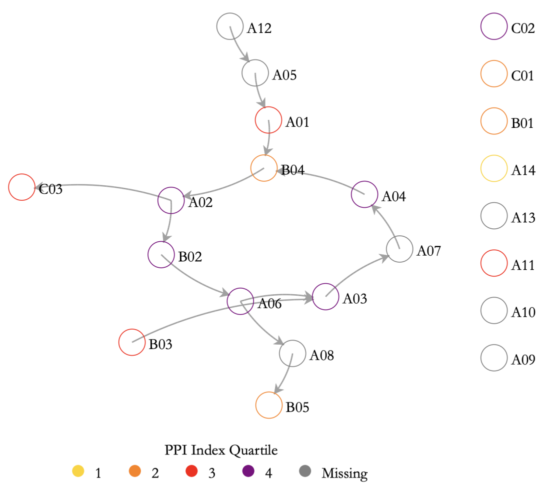
Cycle Ratio: 0.137



Notes: The arrow points to a richer household.

Location 7

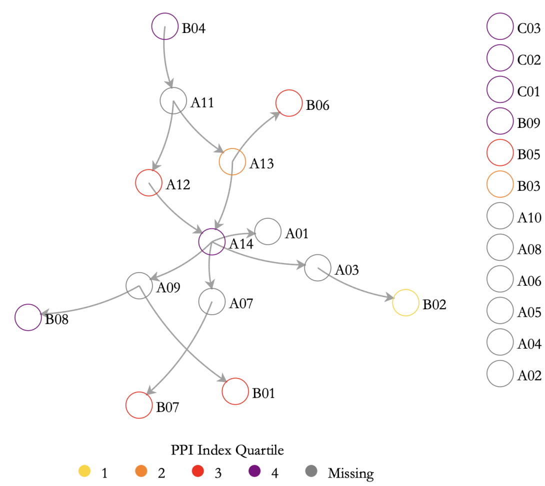
Cycle Ratio: 0.128



Notes: The arrow points to a richer household.

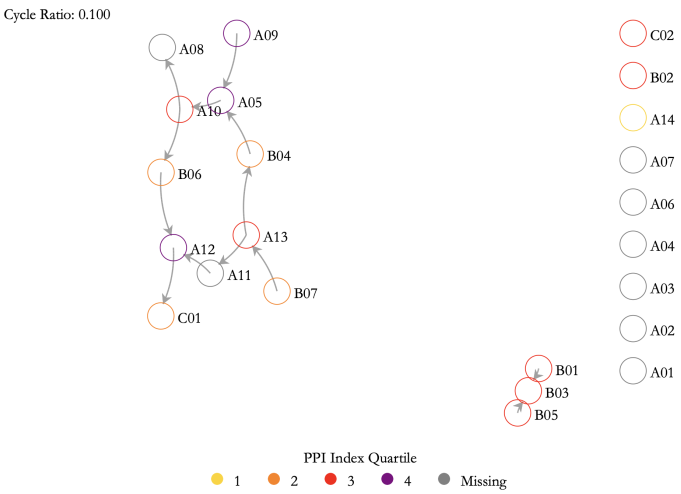
Location 8

Cycle Ratio: 0.100



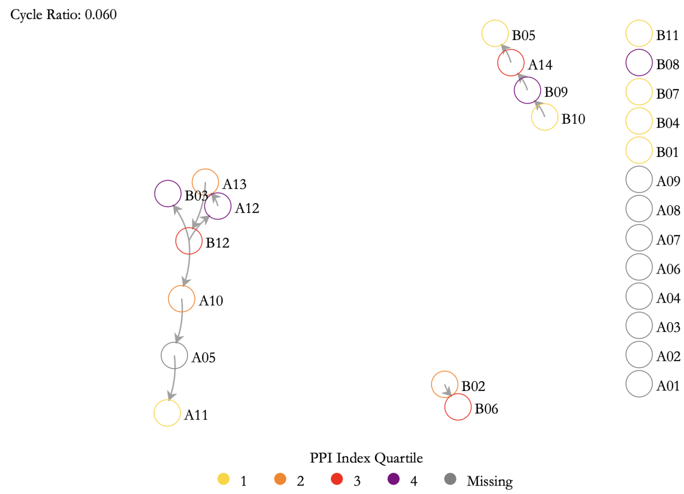
Notes: The arrow points to a richer household.

Location 9



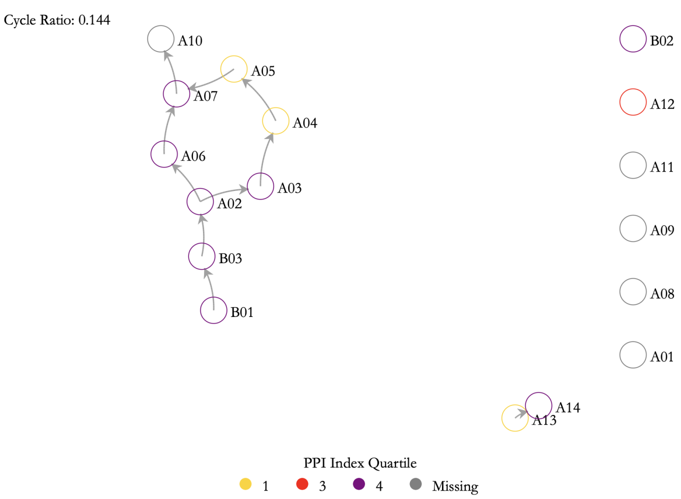
Notes: The arrow points to a richer household.

Location 10



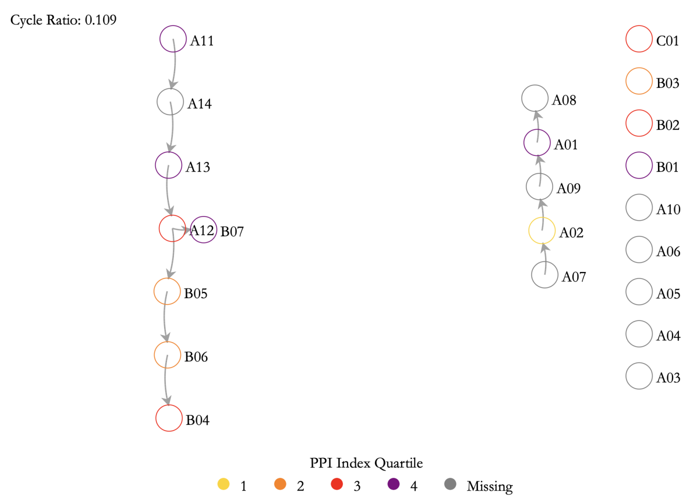
Notes: The arrow points to a richer household.

Location 17



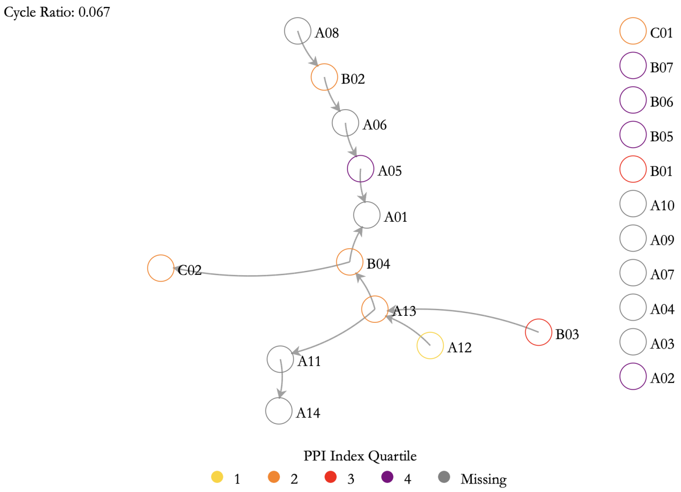
Notes: The arrow points to a richer household.

Location 18



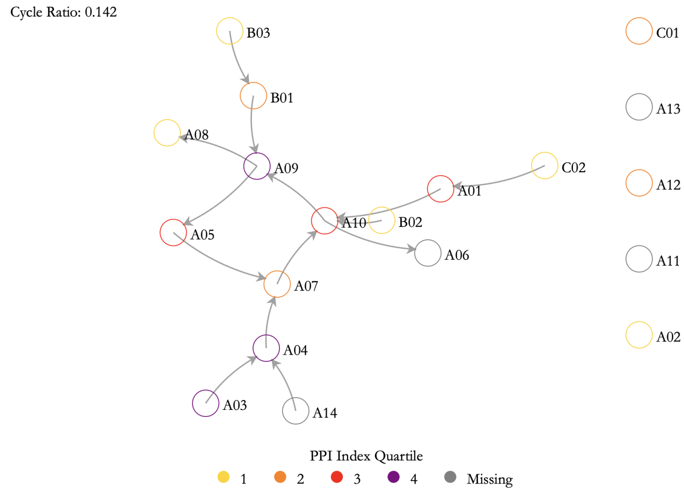
Notes: The arrow points to a richer household.

Location 19



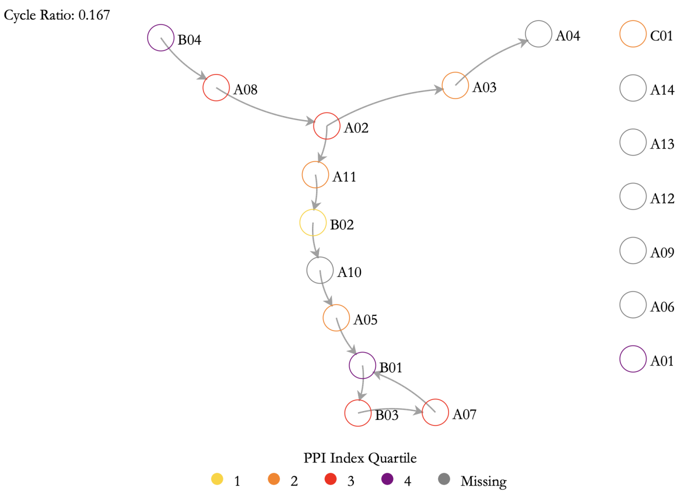
Notes: The arrow points to a richer household.

Location 20



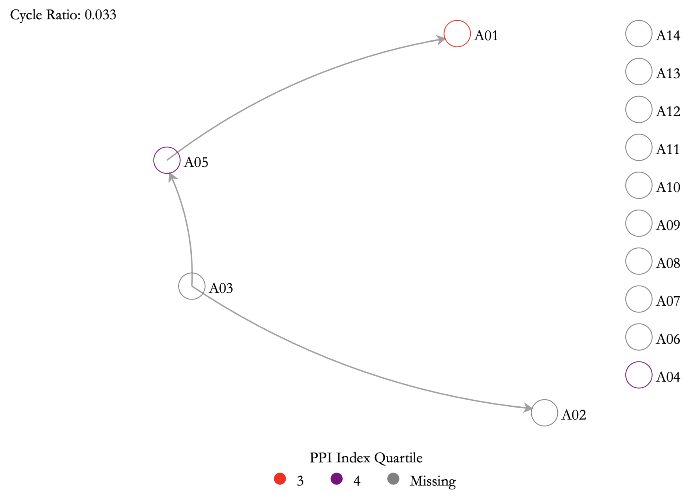
Notes: The arrow points to a richer household.

Location 21



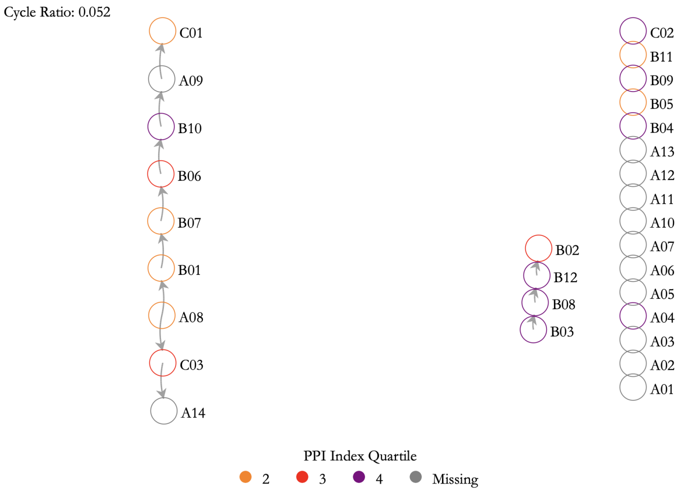
Notes: The arrow points to a richer household.

Location 22



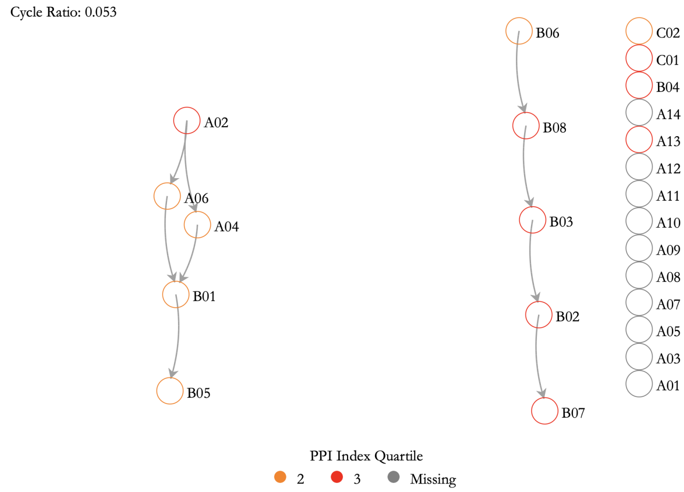
Notes: The arrow points to a richer household.

Location 23



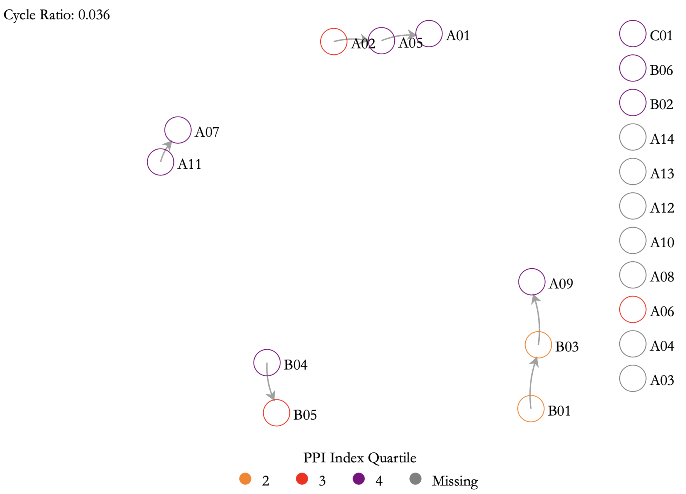
Notes: The arrow points to a richer household.

Location 24



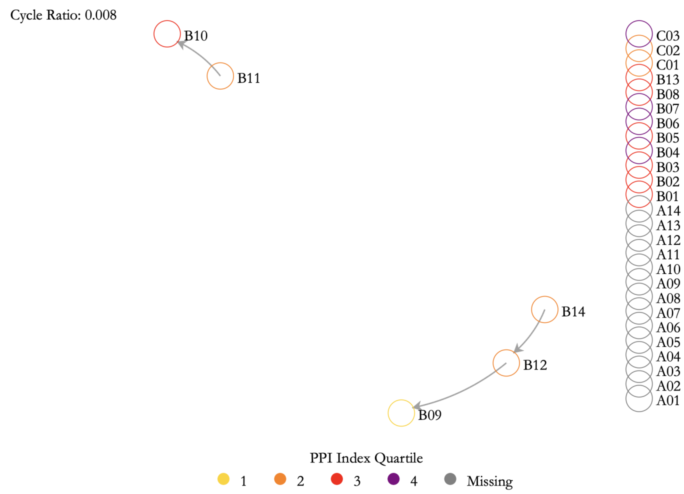
Notes: The arrow points to a richer household.

Location 25



Notes: The arrow points to a richer household.

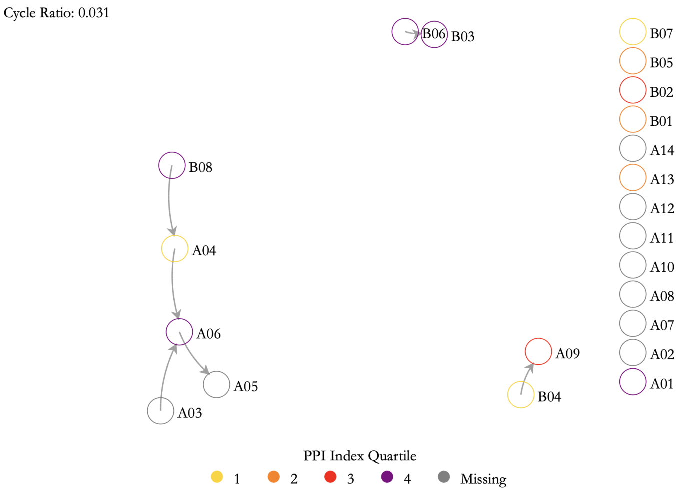
Location 26



Notes: The arrow points to a richer household.

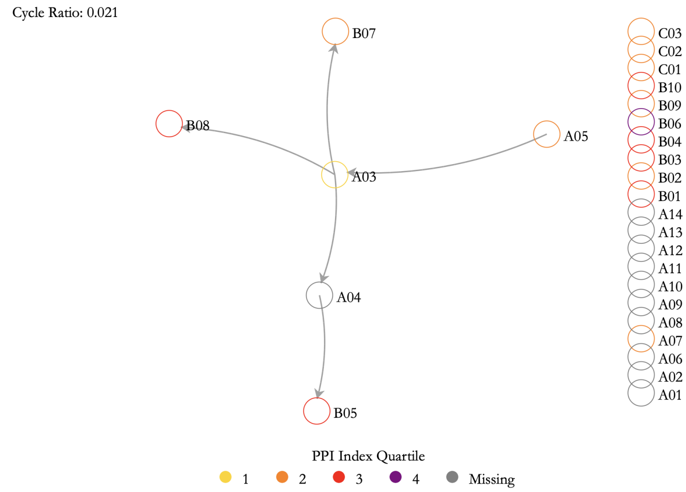
Figure C2: Directed graph of relative rankings - Rural Villages

Location 11



Notes: The arrow points to a richer household.

Location 12

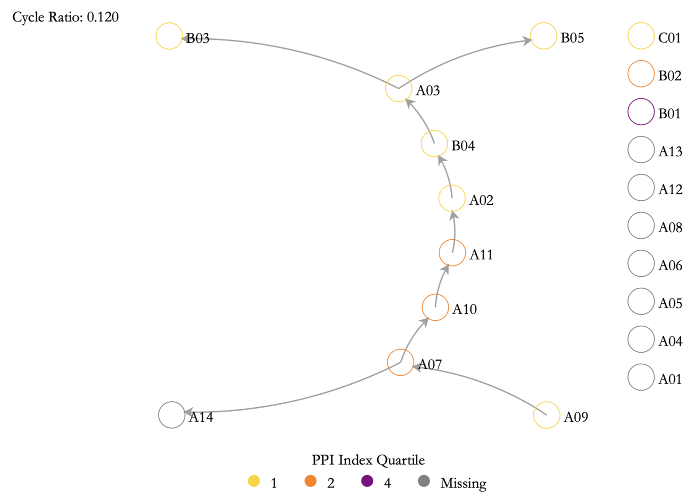


Notes: The arrow points to a richer household.

Location 13

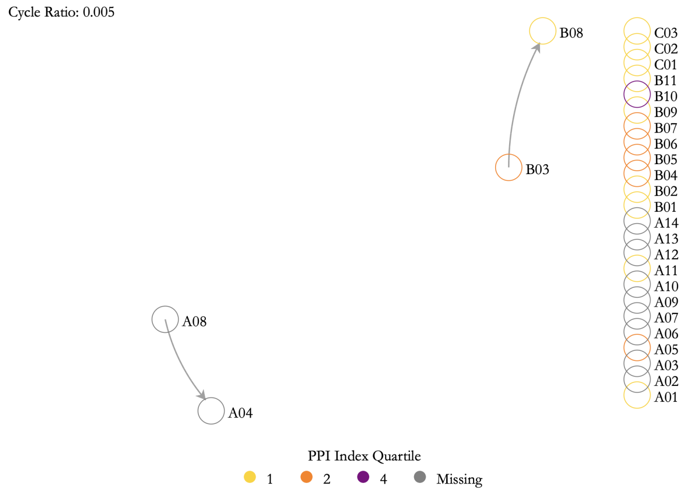
Notes: The arrow points to a richer household.

Location 14



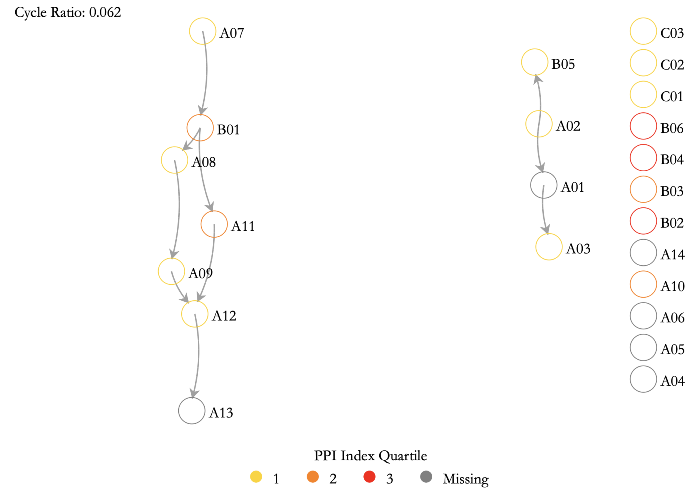
Notes: The arrow points to a richer household.

Location 15



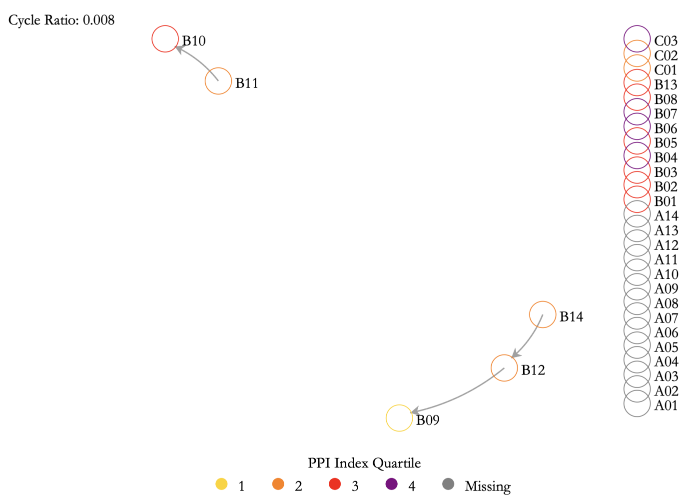
Notes: The arrow points to a richer household.

Location 16



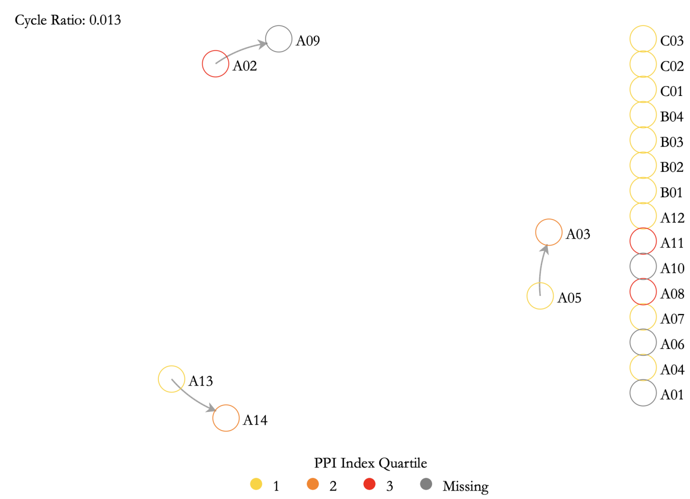
Notes: The arrow points to a richer household.

Location 27



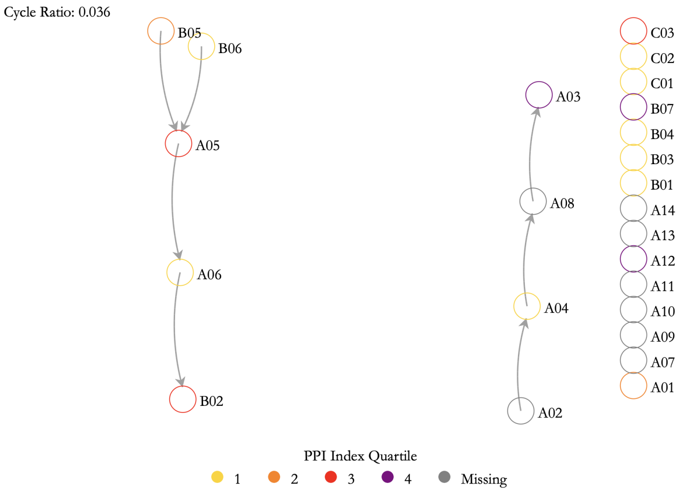
Notes: The arrow points to a richer household.

Location 28



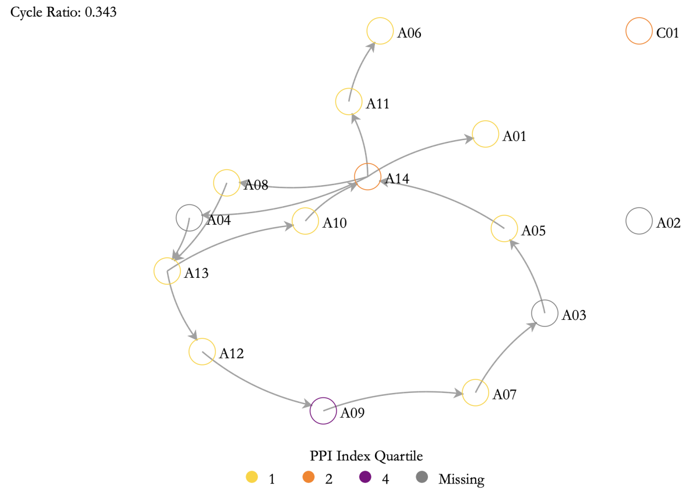
Notes: The arrow points to a richer household.

Location 29



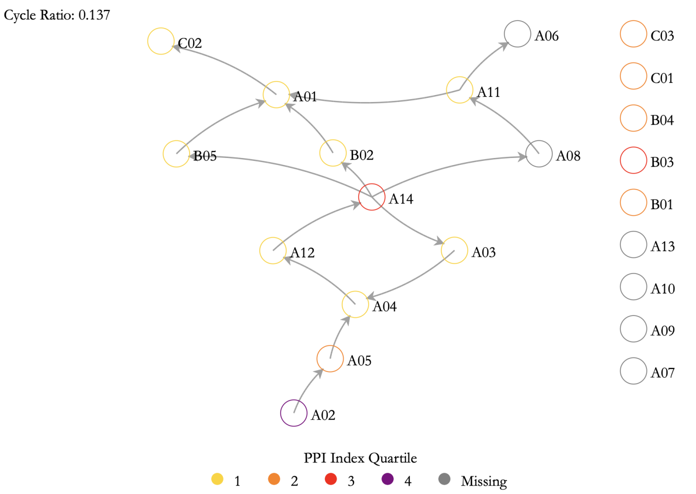
Notes: The arrow points to a richer household.

Location 30



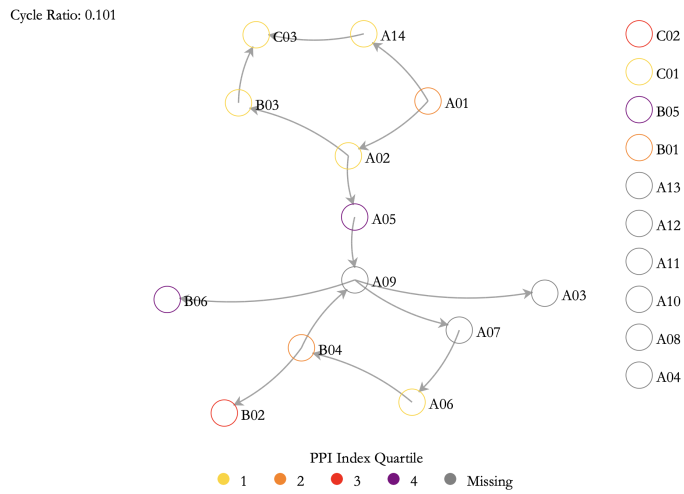
Notes: The arrow points to a richer household.

Location 31

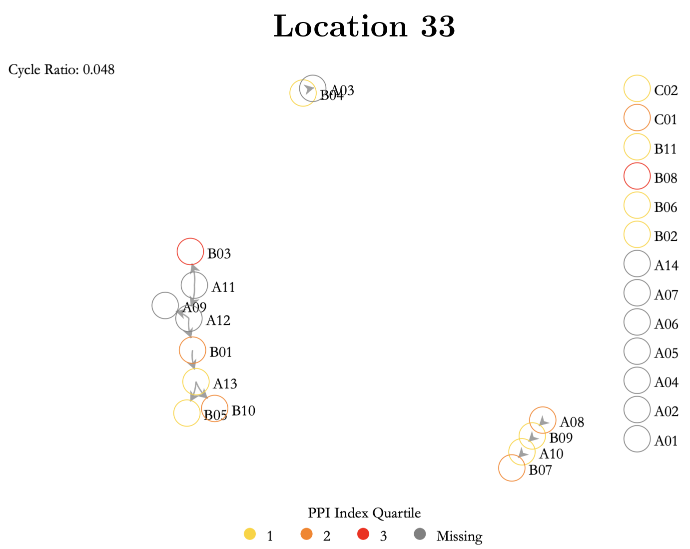


Notes: The arrow points to a richer household.

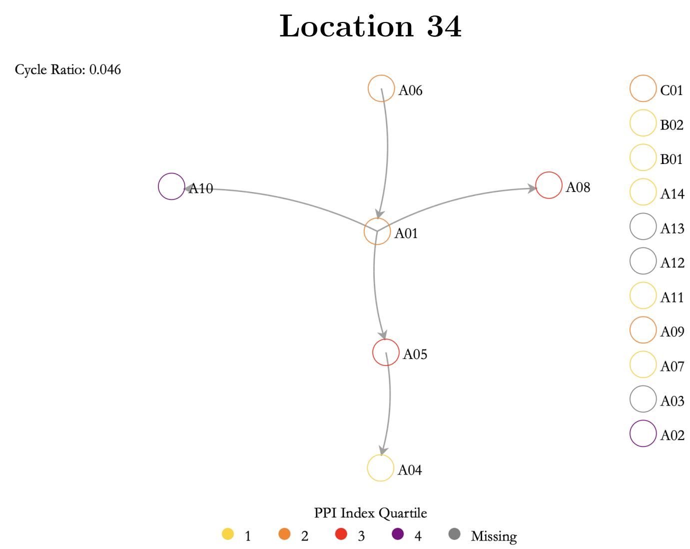
Location 32



Notes: The arrow points to a richer household.



Notes: The arrow points to a richer household.



Notes: The arrow points to a richer household.