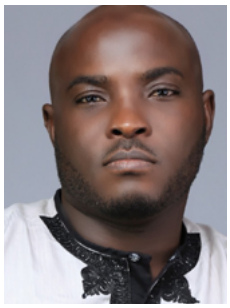
SAMPLING METHODS FOR INFORMAL URBAN POPULATIONS

by Hakeem Bishi and Shelby Grossman



Hakeem Bishi

is a Research Manager, Lagos Trader Project. His email is hakeem.bishi@gmail.com.

Shelby Grossman

is a Research Scholar, Stanford Internet Observatory. Her email is shelbygrossman@stanford.edu.

For comparativists interested in the determinants of political behavior and the consequences of politics on ordinary citizens, capturing the experiences of urban informal workers is critical. Informal workers encompass 62% of global employment and generate one-third of GDP, and more than half of the world's population lives in cities (Kok and Berrios, 2019; World Bank 2019). There are also important questions specific to this population: How are urban informal populations politically mobilized? Under what conditions will they shift to more formal work? How does politics shape their business environment?

However, conducting representative surveys of informal urban populations such as market traders, slum residents, or informal workers poses many challenges. The set of areas is not always obvious; identifying the universe of slums or markets can be difficult. Further, it is difficult to identify the population of interest, such as residents or traders or workers. For example, government agencies rarely keep lists of markets, let alone lists of traders. Private associations are common, but they closely guard their membership lists. Rapid population movement in and out of cities and within-city migration cause further challenges (Auerbach et al. 2018).

Quick solutions can damage survey representativeness. For example, one strategy sometimes used for trader surveys is a city block sampling strategy, where enumerators are instructed to stand at the corner of a city block, face in a specific direction, and approach the Nth shop on the right or left. This strategy imposes a level of order on urban areas that is not always realistic; many urban areas do not have a Manhattan grid-style structure to their streets. Some areas might have multi-story buildings, and it is not clear how this sampling strategy handles such idiosyncrasies. Importantly, the city block sampling strategy permits enumerator discretion. Enumerators could count in such a way that allows them to approach a friendly or less busy trader.

How can researchers define a target population and create a sample frame for informal urban populations? How can we select respondents from a sample frame in such a way that minimizes bias? Researchers commonly use a more sophisticated approach, what we term the "grid & random walk" approach. Here researchers use satellite imagery to impose a grid over an area, and sample cells from the grid (Auerbach 2017, Hummel 2017, Paller 2015). Enumerators then start from a randomly sampled point in the cell

and follow a random walk protocol. Variants on this include sampling buildings from cells using satellite imagery. This method can allow the researcher to cover larger geographic areas, but minimizing enumerator discretion in which individual is surveyed can be resource intensive.

Conducting representative surveys of informal urban populations poses many challenges.

In what follows we describe two strategies that we think are relatively new to simplify the process of reducing enumerator discretion. We have employed these strategies in eight surveys of 3,208 traders in Nigeria and Benin between 2015 and 2019. We term these the "full census" and "grid & partial census" approaches, and describe the conditions under which each is more suited. Last, we discuss how ignoring shop sharing (when two distinct businesses share the same shop) and traders owning multiple shops can bias samples.

Sampling strategies for informal populations

Full census

The first approach, the full census approach, involves creating complete listings for sampling units (like individuals or shops) in a sample frame, randomly sampling from these listings, and directing enumerators to specific respondents.[1] This method minimizes enumerator discretion, but the census itself can be resource intensive. If the population of interest covers a geographically large area, boundaries may need to be imposed to conduct the census, sacrificing some breadth of coverage. Put differently, there may be less overlap between the population of interest and the sample frame.

For one of our projects, the Lagos Trader Survey, the population of interest was in a dense commercial area, making the full census appropriate. The Lagos Trader Survey is a panel survey, now in its fourth round, focusing on wholesale traders of manufactured goods. We deemed a census of traders in plazas (small multi-story buildings) to be our best approximation of the population of interest. A common starting point for creating a sample frame of informal populations is to combine government data with qualitative on-the-ground research to identify the boundaries of slums or other types of communities, and we did just this. We started with a list of plazas from the Lagos Waste Management Authority. While using government data to create a sampling frame can be undesirable for studies of informal populations, in this case, because the agency collects trash from markets, and because the state government subsidizes the collection, its lists included even poorer markets. We then supplemented this list substantially with additional scouting trips around the city, having research assistants map out plazas not already in our listing. (See *Figure 1 f*or an example of what these maps looked like for a similar survey.) Next, a team of research assistants counted all shops in every building from the trash agency list and our newly mapped areas (N=52,830 shops). They also noted what type of product the shop sold. After excluding shops selling food, or providing services and office spaces, we had a listing of 24,159 shops.

The next step was more complicated. We could sample textile shop "number 2" on the third floor of a plaza, but "number 2" meant nothing, substantively. We only knew, for example, that that floor had two textile sellers. We wanted to direct the enumerator to a *specific* one to reduce

1. For other examples of this approach see Aker (2010), Hardy and Mccasland (2019), and Thachil (2017).

enumerator discretion. But we could not provide additional information on the shop. We decided to randomly sample a number between 1 and the total number of textile shops on that floor, and then randomly sampled "left" or "right."

We then instructed enumerators to enter the third floor of the plaza, turn left, and approach the second textile shop. However, these instructions were imprecise; two enumerators could follow the instructions and end up at different shops if they arrived at the floor using different entrances. Nevertheless, this did not matter, as the goal was for the enumerators to *think* we had pre-selected a shop. When enumerators asked about the multiple entrance issue, we advised them to use the main entrance. Ultimately, however, we were indifferent to which entrance they used, as long as they thought we had a pre-selected shop.

For this reason, policing the sampling strategy was not simple. We had auditors assess whether the interviewed trader *could have been selected* by following the sampling strategy. Could this shop have *possibly* been the second textile shop from an entrance? This was not always straightforward. Sometimes shops are closed, and therefore it's not possible to determine what they are selling. An enumerator might thus identify what she perceives to be the second textile shop, while the auditor counts it as the fifth if more shops were open when she visited. We therefore asked auditors to be generous in their assessment. But if the shop did not sell textiles or was on the wrong floor, for example, that indicated a violation of the sampling protocol. Because the sample frame was in a relatively dense area, auditors could audit sampling for five surveys a day.

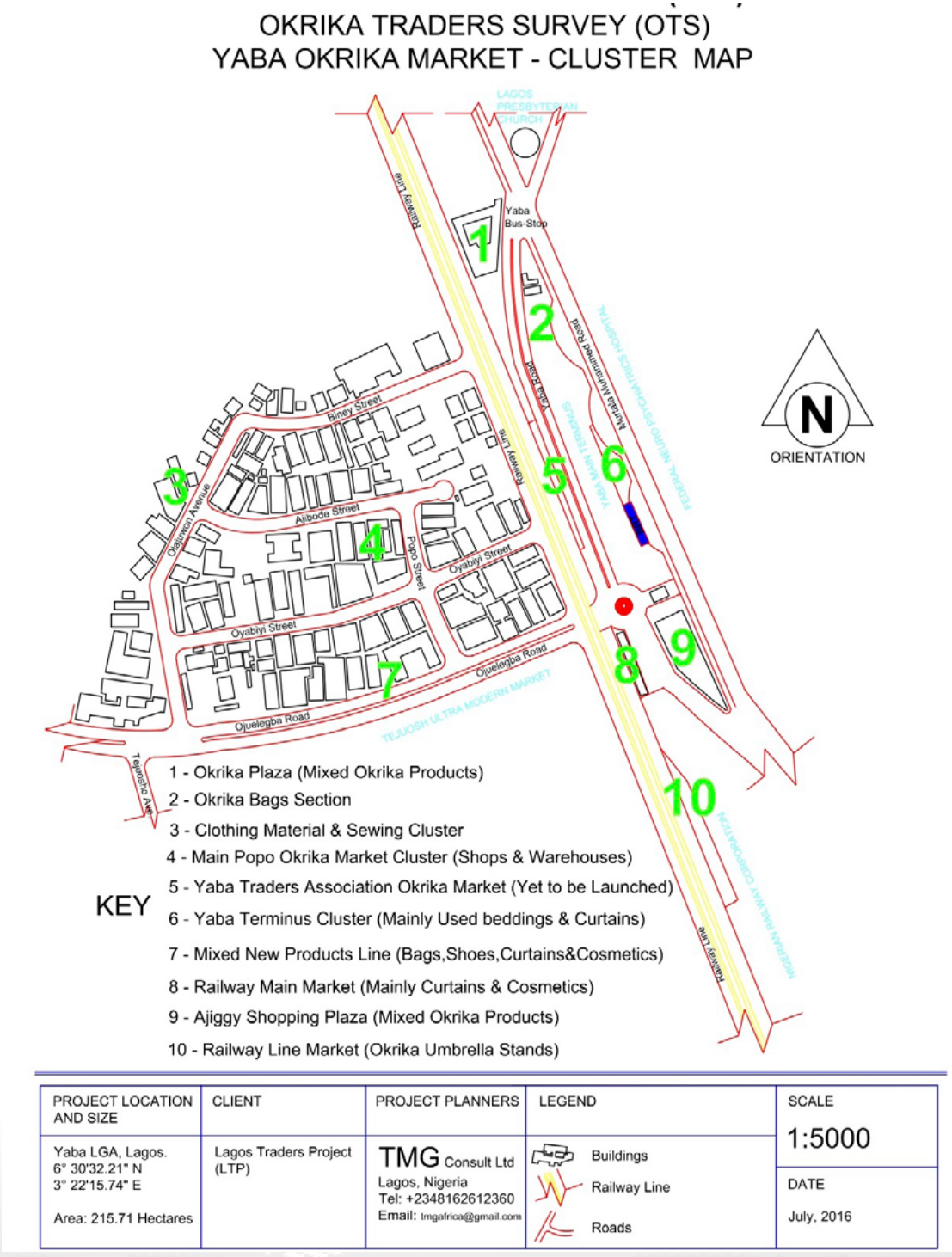
By reducing enumerator discretion, we are minimizing several data challenges. When enumerators approach traders who seem easier to interview, this could cause researchers to underestimate firm size, as less busy traders may have smaller businesses. Conversely, other forms of discretion could result in overestimating firm size: when sampling strategies do not account for buildings with multiple floors, enumerators may be more likely to survey traders on the ground floor, where rent is more expensive. In some contexts firm size may correlate with partisanship, biasing political data. Further, by listing out the full sample, we reduce the likelihood of two enumerators surveying the same respondent, which could be a challenge with random walk protocols.

Grid & partial census

While the full census approach made sense for a target population in such a dense area, we use a different strategy where the population of interest is more dispersed. One of these projects is an ongoing study of used clothing traders dispersed across Abeokuta and Ibadan, two southwestern Nigerian cities. For this survey, we use a grid & partial census approach.

We began with team members conducting scouting trips to each city, and in collaboration with locals who knew these cities well, identifying the commercial areas that had the used clothing businesses we aimed to cover. Using Google Maps, we then shaded the parts of the city with these businesses. From there, we imposed a grid over the entire city, and randomly sampled an equal proportion of cells from the mapped area with a high density of eligible businesses, and from the non-mapped areas. Cells from the latter category allowed us to capture businesses in residential neighborhoods. We diverged from existing strategies in avoiding the random walk strategy; rather we asked enumer-

Figure 1:
TMG Consulting created a map for us of a used clothing market. Research assistants then used this map to create the listing by counting shops, and noting the product being sold, in each building.



ators to conduct a census of shops in each cell, and then attempt to survey 100% of traders in that cell who met the criteria defining our target population.

These cell-level censuses aimed to minimize discretion that emerges with random walk protocols that lack strong auditing on the sampling procedure, within a sampling strategy designed to cover a broad geographic area. In the pilot for this survey, pairs of enumerators both conducted the census and conducted surveys in their cell. For the full survey, one team will conduct the census, and another will complete the surveys. This will further limit discretion by disincentivizing strategically leaving certain shops off the census. Additionally, we have clear protocols for handling literal edge cases, like a shop that is between a sampled and unsampled cell.

Contextual factors may force modifications to the ideal version of the grid & partial census approach. For example, having one team conduct a census and another conduct the surveys is relatively straightforward when traders are in plazas on particular floors where shops have distinct attributes. This was more complicated for a pilot survey we conducted in a Lagos used clothing market, where some traders were in nameless single-floor buildings, and others selling from permanent metal structures with roofs, but no door. It would be difficult for the census conductors to write out information that would lead an enumerator to one metal structure versus another. We experimented with using What3Words, an app that collects GIS coordinates, and transforms the coordinates into three words (like table.apple.blanket) that can be easily shared. But when enumerators attempted to find shops with What3Words, poor cellular networks, combined with What3Words' three-meter square bandwidth (and high market density) made it difficult for enumerators to identify the unique business. In these types of markets, it may be necessary for enumerators to both conduct the census and the surveys to avoid these issues.

Without the full-cell census, enumerators may be tempted to survey traders available to be surveyed on the day they visit the area. This could lead researchers to underestimate firm size, as temporarily unavailable traders may be wealthy enough to travel within Nigeria, to other West African countries, or even to China or Dubai to purchase goods. Further, more marginalized traders may be more reluctant to be surveyed, and so to complete many surveys, enumerators may prefer to survey a different trader instead of following up with the reluctant trader later in the week. We mitigate these challenges by attempting to survey all eligible traders in each cell.

Addressing sampling obstacles due to shop sharing and owning multiple shops

In our surveying, we have identified two firm-specific obstacles that we have worked to address. First, there is a phenomenon in West Africa, and likely elsewhere, of shop sharing. To avoid taxes and stay profitable during economic downturns, two or more traders may share one physical space. One does not work for the other, and they are not co-owners; they have separate businesses in the same small shop. We addressed this by having questions in the survey about shop sharing, taking time to develop a protocol for what counts as shop sharing so that the enumerator could help the trader decide if they were a shop sharer. If a trader reports shop sharing, the enumerator creates a new tracking sheet for the second individual, and aims to survey that trader as well. Shop sharing is an extremely common phenomenon and it

is important for future surveys to capture this information: in surveys of used clothing traders in Lagos and Cotonou, 26% and 32% of traders respectively report shop sharing.

The second phenomenon is of traders having multiple shops, and we have observed that this is common as well: in a representative survey of traders in plazas across Lagos in 2018 we found that 26% of traders reported having at least one other shop. Among those with other shops, 75% had one other shop and 25% had more than one other shop. It is unclear how previous firm surveys have dealt with this. Our solution has been to ask traders if they have any other shops, and if they do, to request that they answer all questions (e.g. questions about revenue) for all of their shops together. This has proven a little tricky, as it is unclear if traders think in this way. We have also had to devise strategies to ensure that we do not survey what we think is a new business but is in fact a shop owned by a respondent who we already surveyed.

Asking about shop sharing and multiple shop ownership matters for correctly measuring firm size, an outcome of interest in many studies. In the absence of information about shop sharing, one could overestimate firm size; for example, researchers may use shop rent as a proxy for firm size, but two shop sharers may split the rent. In the absence of information about multiple shop ownership, researchers could under-estimate firm size if traders respond to questions about their business by only thinking about the shop they are currently in.

Conclusion

In conclusion, we believe that our surveys have highlighted innovative tactics for sampling informal urban populations. We believe our strategies can help to minimize enumerator discretion and more accurately measure firm size. We hope that by describing the practicalities of various sampling strategies researchers will be better able to design their own.

These sampling strategies have led to several important findings. In a *World Politics* article, we look at the determinants of private good governance in Lagos market associations (Grossman 2020). While existing research often suggests that private good governance should thrive when predatory state actors keep their hands out of group affairs, we find that under some conditions the threat of state intervention motivates private group leaders to conduct in-group policing and support trade. In the absence of these state threats, private group leaders may extort from their own traders. In another paper, Meredith Startz estimates the size of contracting frictions traders face when importing their products, and shows that there would be large welfare gains to policies that facilitate market integration (2018). ●

References

Aker, Jenny C. 2010. "Information from markets near and far: Mobile phones and agricultural markets in Niger." *American Economic Journal: Applied Economics* 2(3):46-59.

Auerbach, Adam. 2017. "Neighborhood associations and the urban poor: India's Slum Development Committees." *World Development* 96:119-135.

Auerbach, Adam M., Adrienne LeBas, Alison E. Post, and Rebecca Weitz-Shapiro. 2018. "State, society, and informality in cities of the global south." *Studies in Comparative International Development* 53(3):261-280.

Grossman, Shelby. 2020. "The politics of order in informal markets: Evidence from Lagos." *World Politics* 72(1):47-79.

Hardy, Morgan, and Jamie Lee Mccasland. 2019. "Lights off, lights on: The effects of electricity shortages on small firms." World Bank Policy Research Working Paper 9093.

Hummel, Calla. 2017. "Disobedient markets: Street vendors, enforcement, and state intervention in collective action." *Comparative Political Studies* 50(11):1524-1555.

Kok, Jan de and Mario Berrios. 2019. "Small matters: Global evidence on the contribution to employment by the self-employed, micro-enterprises and SMEs." https://www.voced.edu.au/content/ngv%3A84611

Paller, Jeffrey W. 2015. "Informal networks and access to power to obtain housing in urban slums in Ghana." *Africa Today* 62(1):30–55.

Startz, Meredith. 2018. "The value of face-to-face: Search and contracting problems in Nigerian trade." Working paper.

Thachil, Tariq. 2017. "Do rural migrants divide ethnically in the city? Evidence from an ethnographic experiment in India." *American Journal of Political Science* 61(4):908– 926.

World Bank. 2019. "Global economic prospects: Darkening skies." https://openknowledge.worldbank.org/bitstream/handle/10986/31066/9781464813863.pdf