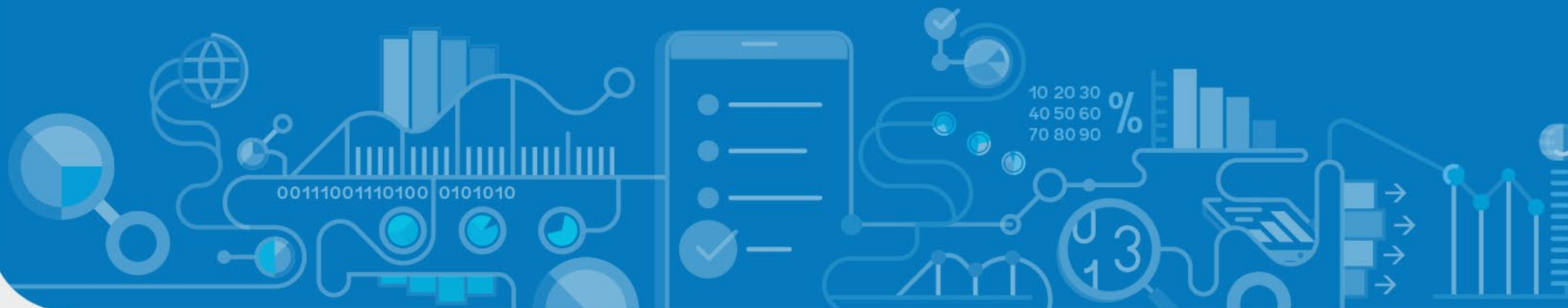




Using Household Surveys and Specialized Enterprise Surveys to Measure Informal Enterprises:

Methodological Evidence from Ghana

Akuffo Amankwah – Hibret Belete Maemir – Pauline Castaing – David C. Francis –
Amparo Palacios-Lopez – Diego Zardetto – Richmond Atta-Ankomah

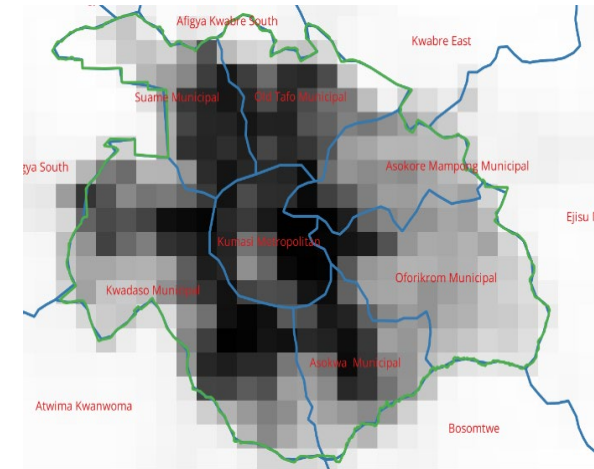


Background

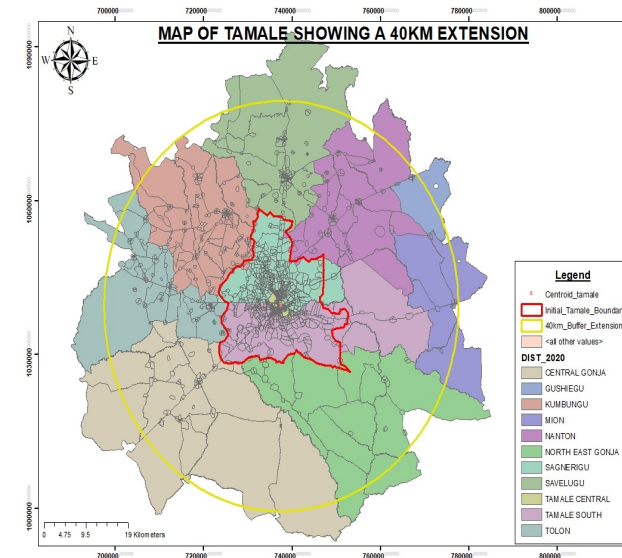
- In LMICs, the informal sector is a prominent feature and means of livelihood, contributing 30%-70% of GDP and employing 20%-80% of the labor force (Ulyssean, 2020).
- However, informal businesses are mostly missing from official records, listings of active businesses, and firm-level surveys: **hard-to-reach population**.
- Accurately measuring the informal sector presents significant challenges, hindering research and limiting the scope for policy recommendations.
- **Defining informality** has been contentious: **distinction** enterprise informality and employment informality (Swaminathan, 1991; Ulyssea, 2018; Vanek et al., 2014; Hussmanns, 2004).
- Other studies highlight methodological issues in estimating the size of the informal sector in LMICs (Charms, 2002; Hussmanns, 2004; Maligalig and Guerrero, 2008).
- Two survey approaches – Informal Sector Enterprise Surveys (ISES) and Household Surveys (HS)
- Our study emphasizes the importance of sampling and survey methods and their implications for characterizing and estimating informal enterprises using experimental data from Ghana.

Study Design

- Conducted an ISES and a multi-topic HS with integrated module on household enterprises (HS-IME) **simultaneously in the same target areas** in Ghana
 - Two secondary cities – Tamale and Kumasi
 - Data collection spanned September to December 2022
 - The HS-IME by LSMS team and ISSER of UG; ISES by DECEA and Esoko.
 - ISES was based on a **spatial sampling approach (Adaptive Cluster Sampling)**
 - HS-IME was based on a standard **list-based sampling approach (two-stage stratified cluster sampling)**.
- Doing the two surveys in the same location simultaneously allows us to
 - Compare estimates (both counts and profile of informal enterprises) in the two surveys.
 - Explore potential differences in the drivers of performance between the two approaches.
 - Provide guidance on methodological improvements to both surveys to better capture informal enterprises.



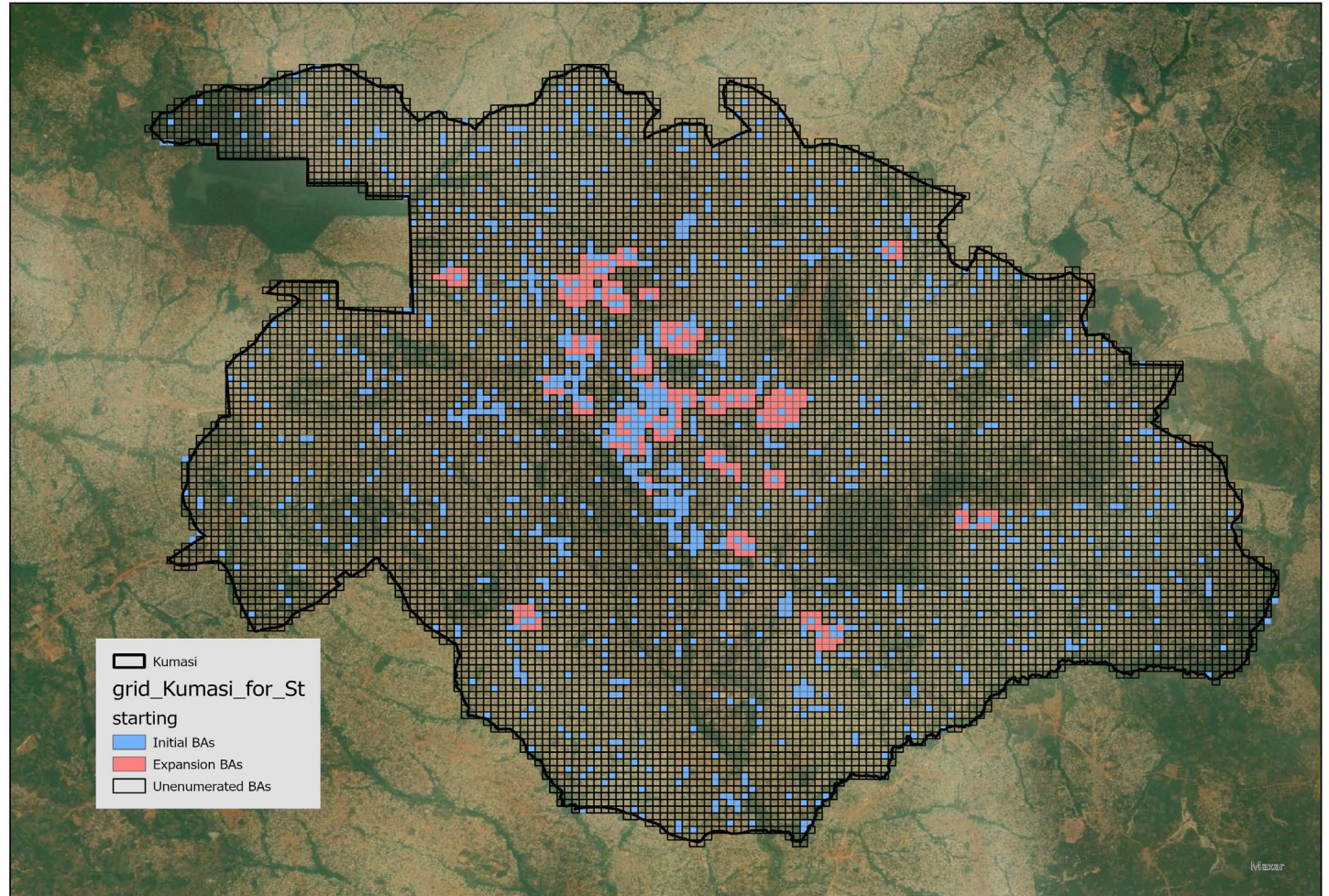
Kumasi



Tamale

Sampling procedure: ISES

- ACS employs area-based frame instead of frame of establishments.
- Requires a **well-defined geographic area**.
- Each city was overlaid with a spatial grid, dividing the area into squares of equal size (150mx150m) -> **Block Areas (BAs)**
- All blocks that meet the threshold (29) trigger full business enumeration of all surrounding blocks
- If these newly sampled BAs meet or exceed threshold (29), **they trigger additional inclusion of adjacent BAs.**
- Dynamic sampling, selection probabilities *change to account for spatial clustering* the population.



Sample distribution of of ISES

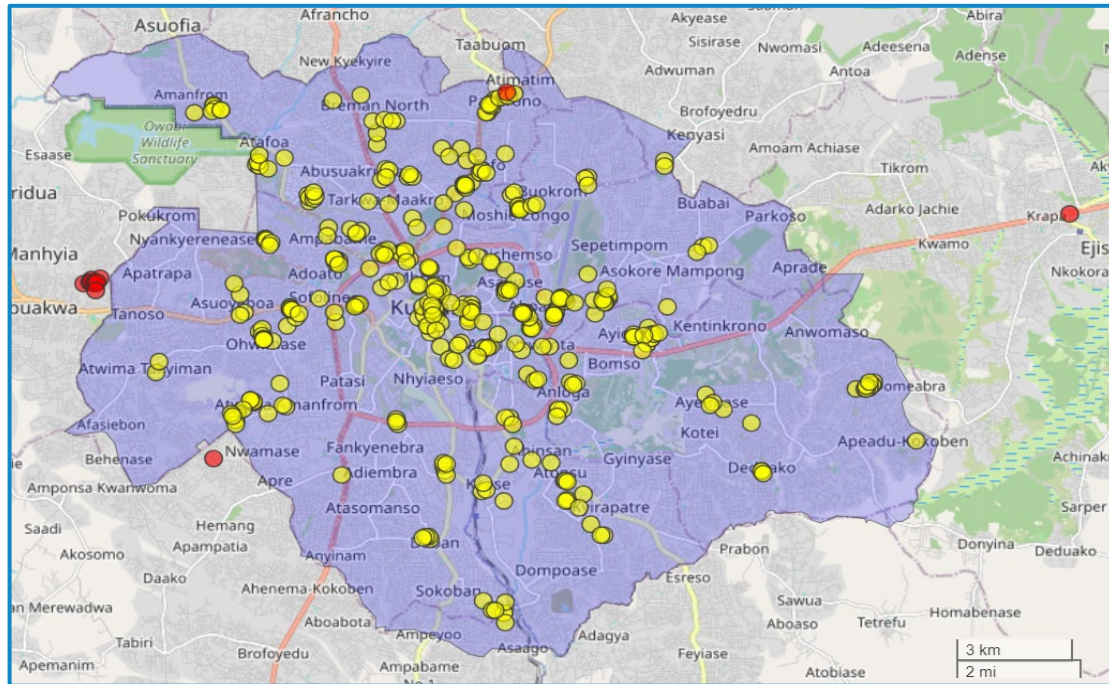
City	Strata	Universe of blocks	Starting blocks enumerated	Total number of blocks enumerated	Informal business units enumerated	Average no. of informal businesses units per block	Interviews completed
Kumasi	Residential	6400	420	461	2329	5	439
	Central Business District	109	60	81	1367	17	98
	Other Business Districts	1627	325	526	7146	14	907
	Markets	152	106	106	1891	18	136
	Everything else	1478	91	99	510	5	122
	Total		9766	1002	1273	13243	
Tamale	Residential	7921	396	431	442	1.03	114
	Central Business District	182	100	166	1569	9.45	179
	Other Business Districts	1938	387	578	2666	4.61	467
	Markets	27	27	27	418	15.48	24
	Everything else	14801	74	77	18	0.23	12
	Total		24869	984	1279	5113	

Sampling procedure: HS-IME

- **Standard two-stage stratified cluster sampling approach + indirect sampling of enterprises**
 - The 2021 Ghana Population and Housing Census list of Enumeration Areas (EA) served as sample frame.
 - We restricted PHC to the geographic delineation of each city's target area and then used as the frame for the first stage sampling.
 - A stratified sample of 134 EAs (67 per city/location) was randomly selected from the sampling frame with PPS (**First Stage Sample**).
 - A complete listing of all households was conducted in the selected EAs.
 - 15 households from each EA were randomly selected with equal probabilities and interviewed, resulting in 1,005 households per city (**Second Stage Sample**).
 - **Indirect sampling of household enterprises:** all enterprises operated by any respondent household were surveyed
- **Household sampling weights** calculated using standard formula and approach
 - No separate weights computed for enterprises.
- **GPS coordinates** collected both at the **household level** and at the **location of the enterprise** operation.

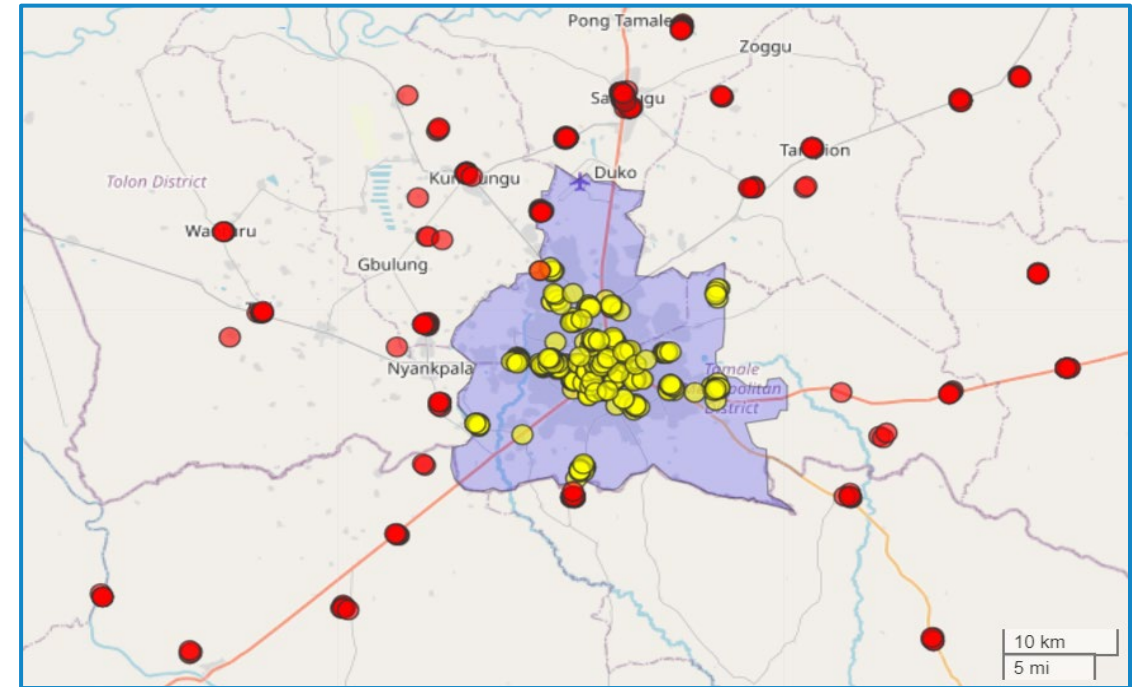
HS-IME realized sample of *operating informal* enterprises

Kumasi



- ▶ Overall, 498 *operating informal* enterprises interviewed
- 25 external enterprises
- **473 internal enterprises**

Tamale



- ▶ Overall, 659 *operating informal* enterprises interviewed
- 242 external enterprises
- **417 internal enterprises**

Analytical approach

- Estimate the **count of informal enterprises** operating in each city by survey type.
- Explore whether similar factors (x) drive enterprise performance (y) from the two surveys. Regressions are run *separately for each survey*.

$$\text{Equation 1 : } y_n = \delta + \sum_{i=1}^k \alpha_i x_{ni} + \varepsilon_n$$

- Pooled the two survey datasets to explore whether enterprises in a particular survey type are different from the other enterprises in terms log **sales** and **productivity** (log sales per worker).

$$\text{Equation 2 : } y_n = \gamma + \beta_1 ISES_n + \beta_2 Tamale_n + \beta_3 ISES_n \cdot Tamale_n + \sum_{i=1}^k \theta_i x_{ni} + \mu_n$$

Key Finding: Count of enterprises

City	HS-IME				ISES			
	Count	CI_L (95%)	CI_U (95%)	n	Count	CI_L (95%)	CI_U (95%)	n
KUMASI								
Operating, formal	13,983	9,298	18,667	50				
Operating, informal	119,722	106,530	132,914	473	67,548	58,651	76,445	1,580
TAMALE								
Operating, formal	6,157	3,098	9,216	31				
Operating, informal	88,820	74,621	103,019	417	21,295	17,242	25,348	784

What could explain the divergence in Counts?

○ Potential undercounting of home-based and mobile businesses by ISES?

- No significant differences between ISES and HS-IME in the estimated proportions of home-based enterprises
- In both cities, the estimated proportion of mobile enterprises is much higher for the HS-IME
 - 26% for the HS-IME vs 16% for the ISES in Kumasi and 21% for the HS-IME vs 12% for the ISES in Tamale

○ Potential “double counting” of businesses by the HS-IME?

- Informal businesses could be co-owned by different households and therefore overweighted
- *Unaccounted multiplicity effect* to explain the discrepancy in the estimated counts of enterprises one should posit that each enterprise is owned on average by 1.7 hhs in Kumasi and by 2.7 hhs in Tamale.

○ Different ability of HS-IME and ISES to capture enterprises operating intermittently during the day?

- While ISES can only capture enterprises that are operating at the time of enumeration, HS-IME doesn't have this limitation, e.g., can better capture enterprises that operate at night.

Drivers of performance in each survey type: Results from equation 1

Significant correlates of **sales** by Survey Type:

- Six predictors **common** to ISES and HS-IME
 - Total workers, bank account, operating at home, being mobile, using phones, years of operation
- Three predictors only in HS-IME
 - Retail sector, manufacturing sector, and member of organization
- Three predictors only in ISES
 - Household size, Share of female workers, and access to credit

Significant correlates of **productivity** by Survey Type:

- Seven predictors **common** to ISES and HSHS-IME
 - Being in the retail sector, total workers, bank account, operating at home, being mobile, using phones, years of operation.
- Two predictors only in HS-IME
 - No education of owner, member of organization
- Two predictors only in ISES
 - Share of female workers, access to credit

Differences in performance: pooled data (equation 2)

	Log of sales			Log of sales per worker		
From ISES	-0.516*** (0.102)	-0.442*** (0.104)	-0.512*** (0.094)	-0.471*** (0.094)	-0.353*** (0.102)	-0.429*** (0.098)
Located in Tamale	-0.425*** (0.103)	-0.445*** (0.108)	-0.335*** (0.113)	-0.210* (0.109)	-0.242** (0.113)	-0.186 (0.115)
ISES_survey*tamale	-0.032 (0.179)	0.004 (0.181)	0.029 (0.162)	-0.380**	-0.322* (0.183)	-0.180 (0.186)
Month fixed effects	No	Yes	Yes	No	Yes	Yes
Covariates	No	No	Yes	No	No	Yes
Observations	2,533	2,533	2,396	2,523	2,523	2,396
R-squared	0.041	0.042	0.266	0.038	0.042	0.224

- Enterprises captured in the ISES are **smaller in size** and **less productive** than those captured by the HS-IME
- Interaction term is not significant for sales
 - the relationship between the survey type and sales does not have to do with location.

Key takeaways

On the methodological front:

- Observed divergence in the count of informal enterprises
 - Potential mechanisms: **under-counting** of mobile enterprises in ISES, undercounting of enterprises that operate by night in ISES, “**double counting**” of co-owned enterprises in HS
 - Clearly demonstrates the **importance of sampling choices in determining counts** of the informal enterprise population.
- **Recommendation for HS:** Include (i) question on co-ownership of businesses by different households; where other households reside; and (ii) elicit typical working hours (e.g., night work)
- **Recommendation for ISES:** aligning the time windows of enumeration to peak hours and methods to estimate and capture, say, mobile enterprises that moved on before they could be listed

On Enterprise performance:

- Performance tends to be consistently lower in ISES
- Yet **correlates** of performance generally remain consistent across both surveys
- **Suggest comparable policy implications.**

LSMS

Living Standards Measurement Study



The LSMS website