

When Labor Counts:

Measurement Timing, Agricultural Productivity, and Gender Gaps in Nigeria

Paula Gonzalez

World Bank & Northwestern University

8th December 2025



GENDER
INNOVATION
LAB



THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

Contents

- 1 Motivation & Measurement Problem
- 2 Data & Measurement Design
- 3 Results: What Measurement Changes
- 4 Implications & Future Work



Why Nigeria? Why Agriculture? Why Gender?

- **Nigeria matters:** Home to over 87 million poor people - central to global poverty reduction. (World Poverty Clock, 2020)
- **Agriculture is critical:** Employs the majority of workers and contributes about 24% of GDP.
- **Gender inequality matters:**
 - Women manage only ~25% of plots but supply about 37% of agricultural labor.
 - Female-managed plots produce 24–47% less than male-managed ones.
 - This gap represents \$2.3 billion in forgone earnings annually. (World Bank, 2022)

Framing the Problem: Outcomes vs. Measurement

Decades of work show:

1. The agricultural gender productivity gap is real, but its drivers are still not fully understood (Udry, 1996; O'Sullivan et al., 2014).
2. Labor and time use are likely central, yet evidence on their returns is thin and context-dependent (Palacios-Lopez et al., 2017; Pierotti et al., 2022; Seymour et al., 2020; Van Aelst & Holvoet, 2016).

A key challenge:

- Capturing agricultural labor in multi-topic surveys is inherently difficult, particularly where tasks are seasonal, irregular, or gender-differentiated. This can affect how labor and productivity are recorded.

What This Paper Contributes to Measurement

Goal: Quantify how measurement strategies for labor and time use shape what we learn about productivity and gender gaps.

Measurement contributions:

- First side-by-side comparison of **GHS recall-based labor** and **high-frequency ALS labor** for the *same plots and season*.
- Direct quantification of measurement error by **region, crop, season, and manager gender**.
- Evidence on how **timing of interviews** interacts with agricultural calendars to bias labor measures.
- Concrete implications for **LSMS-ISA survey design**: when standard modules are sufficient and where targeted improvements matter most.

How GHS Measures Agricultural Labor

General Household Survey (GHS), 2018–2019

- Nationally representative multi-topic survey.
- Agricultural module fielded **post-planting** and **post-harvest**.
- Labor is collected using recall questions:
 - “In the last X days, how many hours did [person] work on this plot?”
 - Separate questions for family and hired labor.

Measurement challenges:

- Interviews occur at fixed points in time that may not align with peak activity.
- Long recall windows and complex, intermittent tasks.
- Gendered division of labor (e.g., processing, late harvest) is especially vulnerable to underreporting.

How ALS Re-Designs Labor Measurement

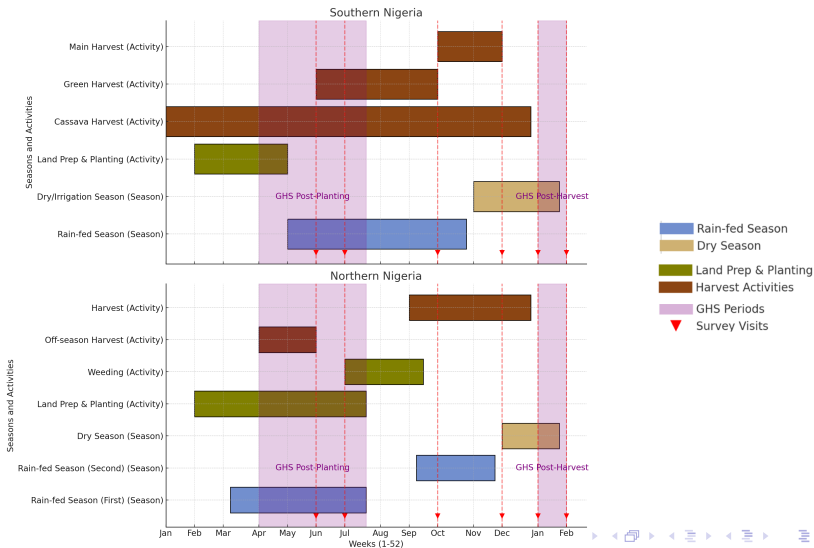
Agriculture Labor Survey (ALS), 2019

- High-frequency follow-up to the GHS for farming households.
- Seven survey rounds spanning the agricultural season.

Key measurement innovations:

- **Seasonal coverage:** Rounds aligned with planting, weeding, harvesting, processing.
- **Short recall periods:** One-day and one-week labor diaries reduce recall bias.
- **Intra-household detail:** Labor is collected for all plot workers, and one randomly selected worker is interviewed in depth on tasks, earnings, gender norms, and work relationships.
- **Comparability:** Implemented on a GHS-linked sample, so differences reflect *measurement*, not sample selection.

Timing Matters: Seasonality and Data Collection



ALS Sample and Link to GHS

ALS sample:

- 973 farming households, 1,060 plots, followed across 7 rounds.
- Drawn from 15 GHS-linked states in 4 agro-ecological zones.
- 61% male-managed plots, 39% female-managed.
- 99% of plots successfully tracked across the season.

Why this matters for measurement:

- For each plot we can construct:
 - GHS-style recall labor inputs.
 - ALS-based high-frequency labor inputs (with/without seasonal accumulation).
- This allows a clean comparison of **measurement strategies holding the underlying production process fixed.**

Empirical Strategy

Baseline specification:

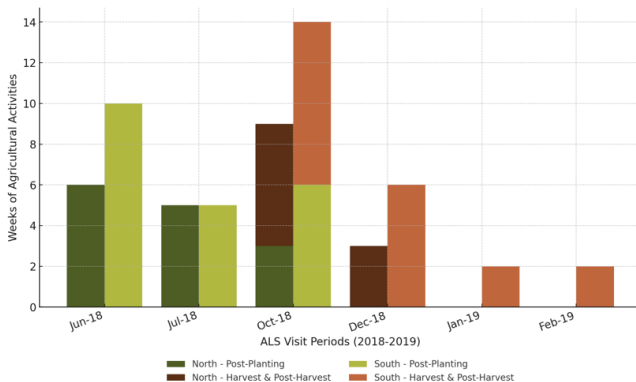
$$Y_{ij} = \alpha + \delta G_j + \omega C_j + \theta P_i + \lambda_z + \varepsilon_{ij}$$

Measurement variation:

$$Y_{ij} = \alpha + \delta G_j + \beta L_{ij}^m + \Gamma X_{ij} + \lambda_z + \varepsilon_{ij}$$

- L_{ij}^m : labor input under measurement strategy m .
- Compare estimates using:
 - No labor controls
 - GHS recall labor
 - ALS labor (single round)
 - ALS accumulated labor

Agricultural Activity Timeline by Region



Region	Final Adjusted Active Weeks	Post-Planting Weeks	Post-Harvesting Weeks
North (NC, NE)	23 weeks	14 weeks	9 weeks
South (SE, SS)	39 weeks	21 weeks	18 weeks

Weeks of reported agricultural activity during ALS visits. The table aligns cumulative weeks with GHS post-planting and post-harvest rounds. Northern seasons are shorter and more concentrated; Southern seasons are longer and more staggered.

Gender Gap in Output under Different Labor Measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Actual value harvested (IHS)							
	SOUTH				NORTH			
	No Labor	GHS	ALS	ALS	No Labor	GHS	ALS	ALS
		Through Season	Through Season	No accumulative		Through Season	Through Season	No accumulative
Woman Manager	-0.34*** [0.11]	-0.34*** [0.11]	-0.30*** [0.11]	-0.24** [0.11]	-0.44*** [0.15]	-0.44*** [0.15]	-0.46*** [0.15]	-0.43*** [0.15]
Total Labor		0.02** [0.01]	0.03*** [0.01]	0.28*** [0.05]		0.02*** [0.01]	0.02*** [0.01]	0.11*** [0.04]
Constant	10.22*** [0.34]	10.21*** [0.33]	10.01*** [0.32]	9.96*** [0.33]	10.84*** [0.45]	10.83*** [0.44]	10.82*** [0.45]	10.85*** [0.45]
Observations	609	609	609	609	451	451	451	451
Adjusted R-squared	0.26	0.28	0.30	0.34	0.40	0.41	0.41	0.41
Mean Male	62638	62638	62638	62638		117155	117155	117155
<i>P-value Across Regressions</i>		0.78	0.06	0.08		0.76	0.44	0.02

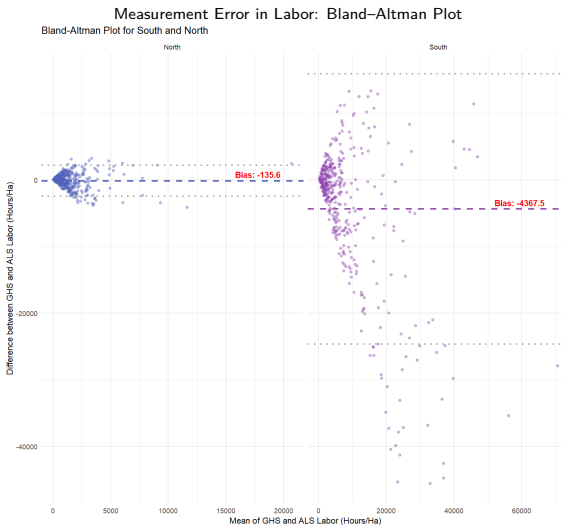
The value displayed for t-tests are the differences in the means across the groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels. (1) All regressions include geopolitical zones fixed effects. (2) P-values are provided at the end of the table. Panel A displays p-values used to compare coefficients between consecutive regressions; for example, a p-value of 0.05 compares the coefficient of the female manager in column 3 with that in column 2. (3) The agricultural value of production was measured during the last wave of the ALS data collection in July 2019, following the post-harvest period. (4) Data were sourced from two surveys in Nigeria: the 2018-2019 General Household Survey Data (GHS) and the 2018-2019 Agricultural Labor Survey (ALS).

Estimated gender gap in plot-level output, South (columns 1–4) and North (columns 5–8), under alternative labor measurement strategies.

Interpreting the Gender Gap Table

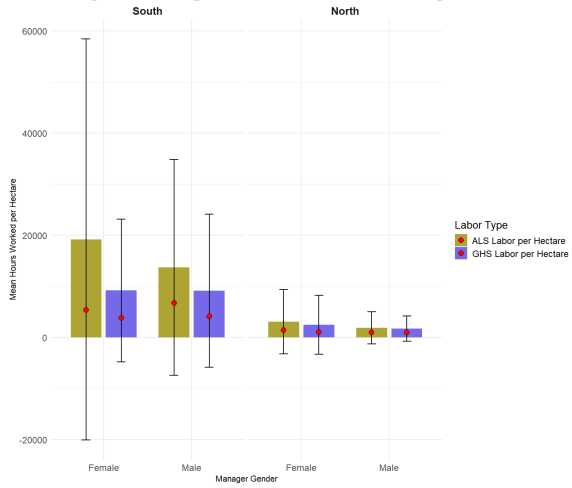
- In the **South**, moving from “No Labor” to “GHS Labor” reduces the estimated gender gap only modestly.
- Replacing GHS labor with **ALS Accumulated** labor substantially changes:
 - The magnitude of labor coefficients.
 - The residual gender gap left after conditioning on labor.
- In the **North**, where the agricultural calendar is shorter and more synchronized:
 - GHS labor is closer to ALS measures.
 - Measurement error appears smaller and less gender-differentiated.

Takeaway: Measurement timing and strategy matter most in settings with long, staggered seasons and labor-intensive crops.



GHS underreports labor relative to ALS, especially in the South, where the bias is large and systematically negative.

Labor per Hectare by Manager Gender and Region



Distribution of labor per hectare using GHS vs. ALS measures, by manager gender and region. ALS captures larger dispersion and systematically higher labor in the South, especially on female-managed plots.

Heterogeneity in Labor Contributions & Measurement

1. By labor type

- ALS distinguishes manager, family, and hired labor more clearly.
- GHS compresses family labor, especially for irregular or gendered tasks.

2. By season

- Post-harvest labor (often female-led) strongly predicts output in ALS.
- Post-planting labor contributes less, partly due to interview timing.

3. By crop

- Cassava is highly sensitive to accurate labor measurement.
- Maize shows more stable results across measurement strategies.

What Measurement Changes —Summary

- **Level of labor:** GHS underreports labor, especially in Southern Nigeria and on labor-intensive crops.
- **Variation in labor:** High-frequency ALS data reveals substantial within-season variation that GHS compresses.
- **Returns to labor:** Using ALS labor increases estimated labor returns and alters the structure of marginal returns by type of labor and season.
- **Gender productivity gap:** The “unexplained” gap shrinks in some contexts once better labor data are used, but not uniformly.

Implication: Measurement strategy does not just add precision; it can qualitatively change how we interpret gender gaps in productivity.

Final Takeaways

- Labor and time-use measurement is not neutral: it shapes what we learn about productivity, labor returns, and gender gaps.
- High-frequency designs like ALS are particularly valuable where:
 - Seasons are long and staggered.
 - Labor is highly intensive and gender-differentiated (e.g., cassava).
- Standard surveys such as GHS remain essential, but:
 - They underreport labor.
 - They blur seasonal dynamics in ways that are not random across space, crops, or gender.

Bottom line:

When labor counts in policy, measurement timing and design also have to count.

Policy & Measurement Implications

For Measurement:

- Incorporate high-frequency or time-sensitive labor modules where recall is difficult (e.g., post-harvest, cassava).
- Use GHS as a foundation—refine it where more precision is needed.

For Policy:

- Target support to high-return labor periods and undervalued crops.
- Link labor insights to structural transformation policies: identifying where agriculture can be a lever for growth.

Future Work

- **Timing of harvest and output measurement:** Study how later collection of output data affects observed gender gaps when women harvest later than standard survey windows.
- **Labor productivity vs. total output:** Separate the role of labor in intensification (yields) vs. land expansion.
- **Crop-specific labor dynamics:** Extend analysis to other crops and task-specific labor (weeding, processing).
- **Measurement design experiments:** Test low-cost tools (prompted recall, pictorial aids, phone follow-ups) to reduce recall bias in large-scale LSMS surveys.

Thank you!

Questions & comments welcome.