Cornell University

Harnessing High Frequency Data To Inform Development and Humanitarian Interventions

Christopher B. Barrett

Keynote address to the World Bank conference
The Pulse of Progress: Harnessing High-Frequency Survey Data for Development Research in the Polycrisis Era
Washington, DC
December 17, 2024

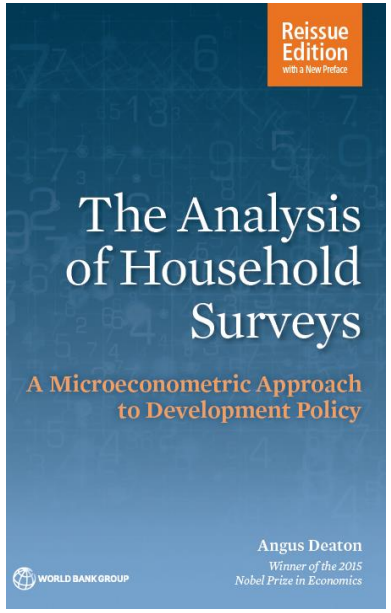
Cornell University

Why national survey data?

Statistically representative observational data essential for accurate descriptive/predictive analysis. Often useful for inferential analysis.

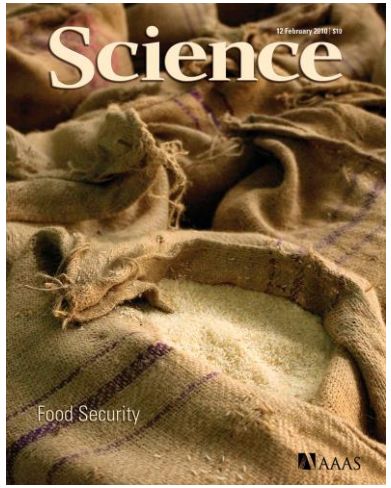
World Bank established LSMS >40 years ago for comparable measurement of living standards defined broadly. With improved measurement came improved analysis.

– Angus Deaton 1997 (and 2018)



"To direct scarce resources to where they can do the greatest good, actions must be guided by reliable information … Measurement drives diagnosis and response. "

– Barrett (*Science* 2010)



Cornell University

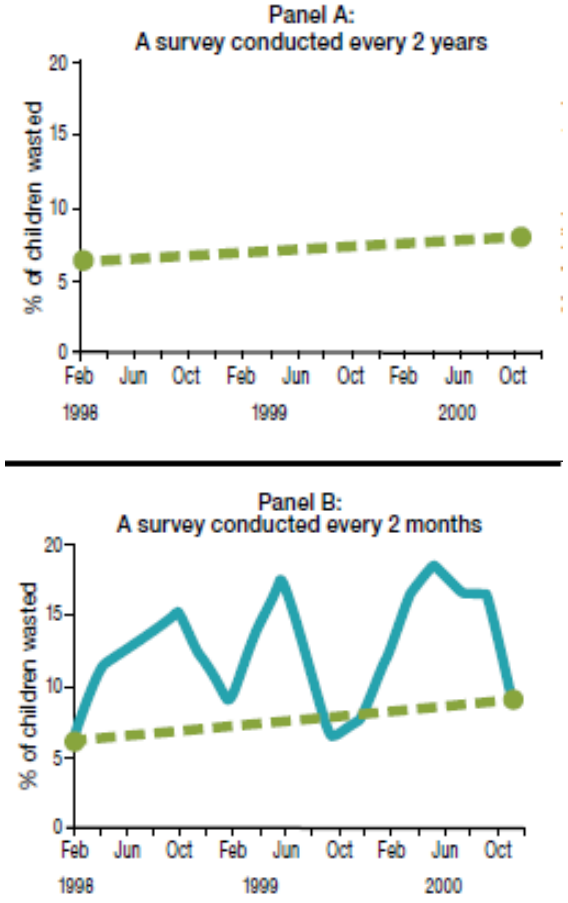
Why high-frequency longitudinal (HFL) national survey data?

Living standards are dynamic. A solid understanding of dynamic living standards requires HFL data.

1. **Differentiating chronic vs. transitory state** (poverty/food insecurity/ill health) matters fundamentally to policy design and evaluation.

2. **Seasonality** looms large for the poor (esp. rural poor). Surveys every few years either miss seasonality or create seasonal mismatch issues in intertemporal comparison.

Child wasting in Bangladesh



Source: Barrett & Headey (IFPRI 2014) using data from Bloem et al. (HKI 2003)

Why high-frequency longitudinal (HFL) national survey data?

3. **Dynamics introduce risk and uncertainty**. It's hard to understand human behavior and well-being without explicitly considering risk.

- Risk aversion and defensive actions/investments
    - moments beyond the mean matter to behavior and well-being.
- Non-stationarity and shocks
    - inference commonly assumes stationary DGPs

4. **Improved inference** about what causes change in living standards.

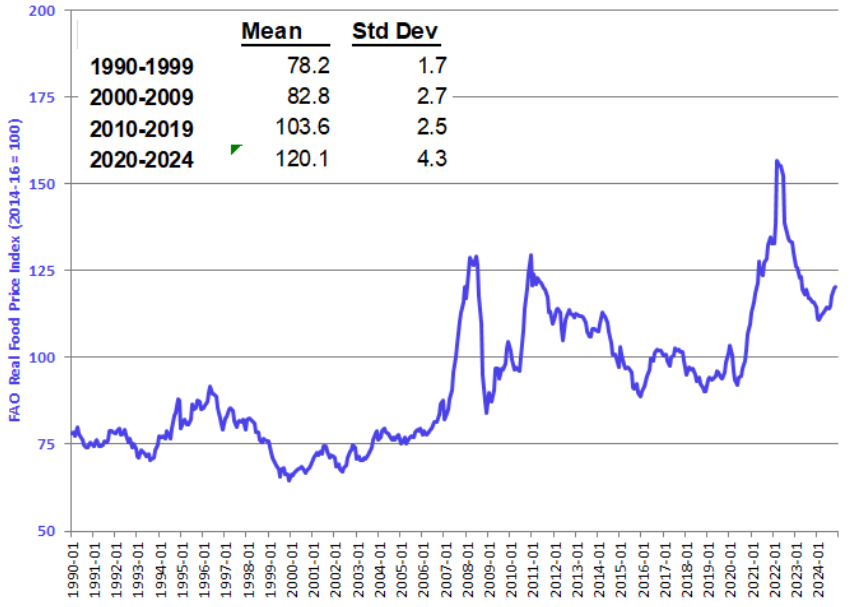
- FE estimators –true intertemporal change; time invariant unobservables.
- Improve external validity – Rosenzweig & Udry (*REStudies* 2020)

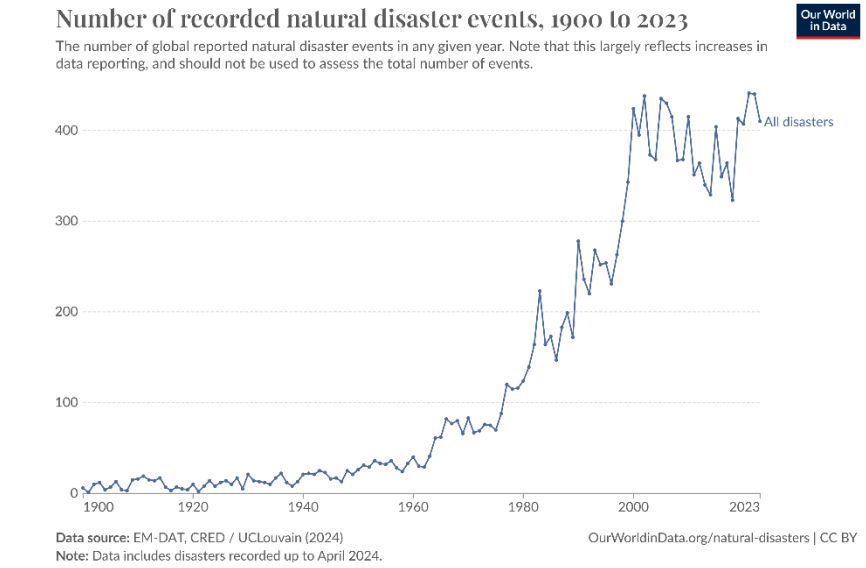
Polycrisis era

HFL data esp. important amid polycrisis. Multivariate risks growing: weather, conflict, prices, pandemics, etc. Solid inference around effective interventions crucial. Must parse chronic/seasonal/transitory effects of shocks.

Distinction b/n s-run humanitarian and longer-run development interventions is increasingly blurred, esp. in low-income settings.

Hence attention to 'resilience' = shock-proofing continuous improvement in living standards.

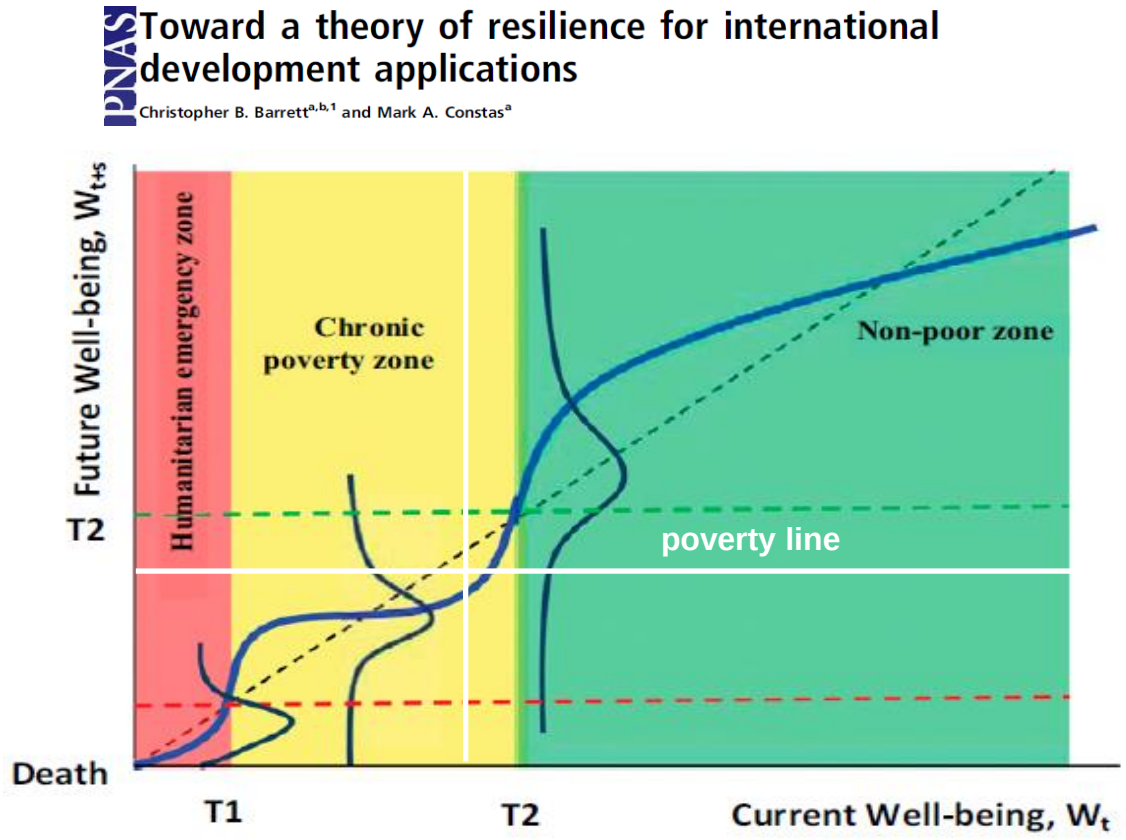


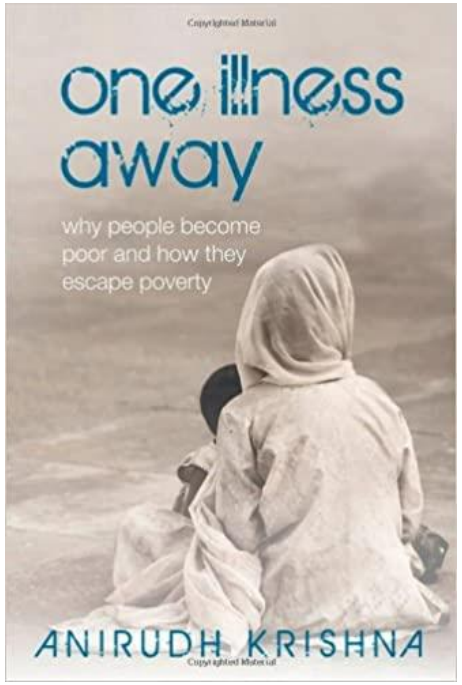


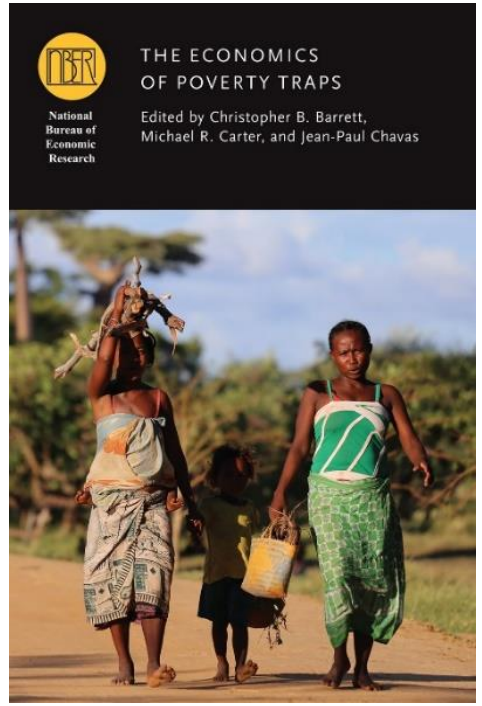
Cornell University

Polycrisis, poverty traps, and resilience

Poverty traps increasingly understood as arising due to catastrophic risk exposure and/or the experience of catastrophic shocks. Reasonable theory exists, rooting resilience in living standards.







Resilience measurement

HFL survey data especially important for studying resilience.

PNAS

Opinion: Measuring development resilience in the world's poorest countries

Derek Headey[a,1] and Christopher B. Barrett[b,1]

"our empirical understanding of development resilience in developing countries remains remarkably limited, primarily because of data shortcomings. … [The world needs] a multicountry system of high-frequency, long-term sentinel sites in the world's most vulnerable (and largely rural) regions. … In an increasingly volatile world, good data are essential. Such data will help the global community build development resilience and eliminate hunger, extreme poverty, and vulnerability in the generation ahead."

Yet very scarce. Scoping review of 'resilience' studies 2008-2020 found just 16% used panel data at all, much less HF! (Barrett et al. *WD* 2021)

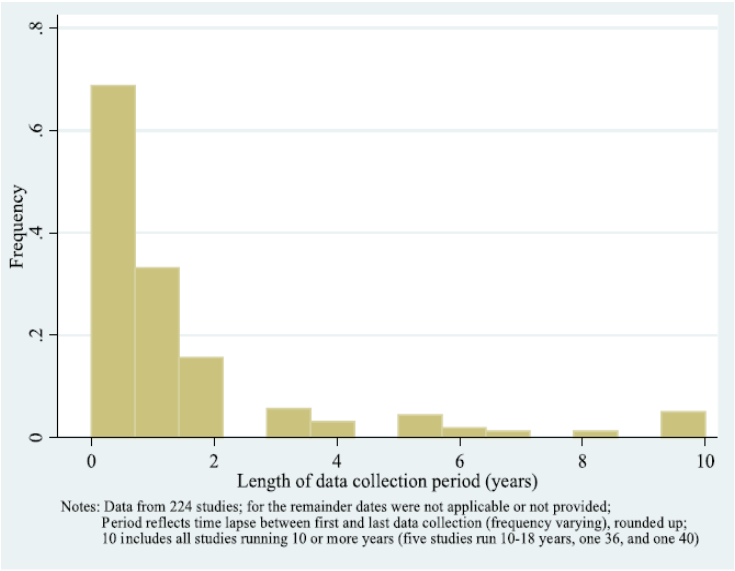


Fig. 5. Distribution of empirical studies, by duration of data collection.

Resilience measurement

Resilience – like poverty, food security, women's empowerment – is a latent variable … inherently hard to measure. Takes time to create/validate good measures of latent variables. Not yet there for resilience measurement.

Core issues:

1. Interventions aim to '*build resilience*', thus measures must allow *resilience* to be a dependent variable and to carry normative implications (e.g., monotonicity).
2. Must work with multiple perils, given multivariate risk exposure.
3. Must be decomposable since subpopulations may be heterogeneous.
4. Given pronounced behavioral response to risk, must account for *ex ante* risk exposure not just *ex post* experience of shocks.
5. Measures remain contested, imprecise, and inconsistent (Upton, Cissé & Barrett *Ag Econ* 2016; Cissé & Barrett *JDE* 2018; Barrett et al. *WD* 2021; Upton, Constenla-Villoslada & Barrett *JDE* 2022)

More and better resilience measurement research is needed.

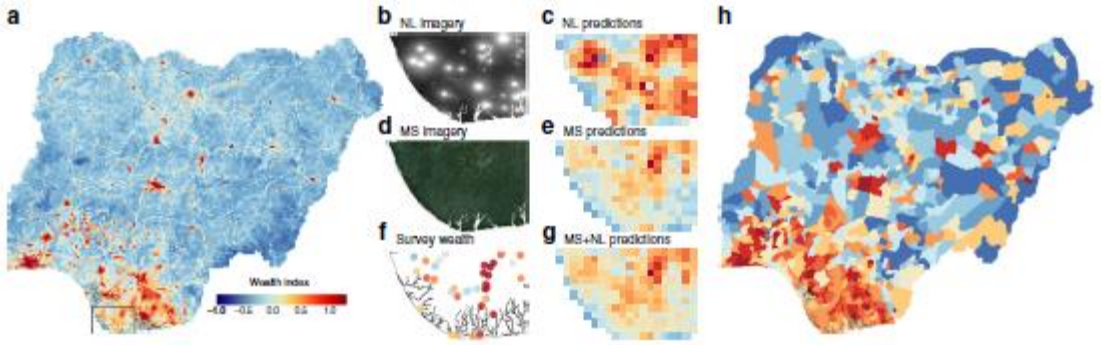
Cornell University

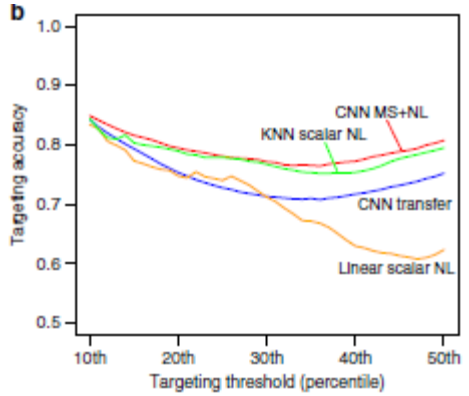
Benefits/Limits to Machine Learning

The dearth of HFL socioeconomic survey data motivates enthusiasm for HFL geospatial data and ML algorithms to generate nowcasts/forecasts:
" a given African household will appear in a household well-being survey less than once every 1000 years" – Yeh et al. (*Nature Communications* 2020)

Machine learning is indeed very promising, esp. for data fusion.
(Yeh et al. *NC* 2020; McBride et al. *AEPP* 2022; Gualavisi & Newhouse *WBER* 2024; Newhouse *CSAB* 2024)

Example: spatial imagery + survey data permit pixel-scale wealth predictions at scale and downstream targeting. (Yeh et al. *NC* 2020, Figs 5-6 below)





Benefits/Limits to Machine Learning

But HFL EO data complement, not substitute, for HFL SE data. Opportunities arising from ML and VHR/HF geospatial data boost need for HFL SE data.

1. **ML models need lots of good, reasonably current training data**, ideally HFL SE data. Nonstationarity is a fundamental challenge, even to nowcasting. (Browne et al. *PLoS ONE* 2021; Constenla et al. in review; Tennant et al. in review).
2. **Prone to atheoretic modeling** (McBride et al. *AEPP* 2022)
3. **Frequent, large bias** – focus on OOS predictive skill overall. Misses significant spatial/livelihood heterogeneity, esp. related to unobservables and non-classical measurement error. (Ru et al. in review; Tennant et al. in review)
4. **Reproducibility harder** than conventional econometric/statistical work.
5. Often **too technical and high-cost** for operational agencies.

Cornell University

What role for HFL survey data in polycrisis era?

Rapid response ≠ researchers' comparative advantage. We must improve our ability to respond to sudden changes.

HFL SE survey data can not only:

- improve ability to describe dynamics and risk
- do better inference
- add value to HFL geospatial data and ML

HFL SE survey data can also enable analytical tasks that researchers undersupply currently.

Cornell University

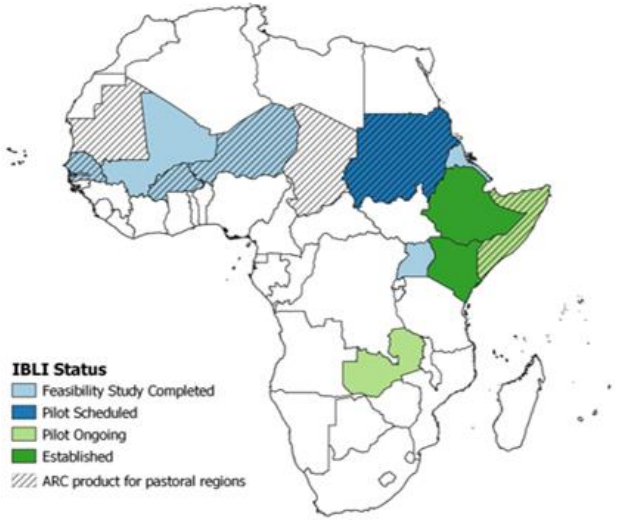
What role for HFL survey data in polycrisis era?

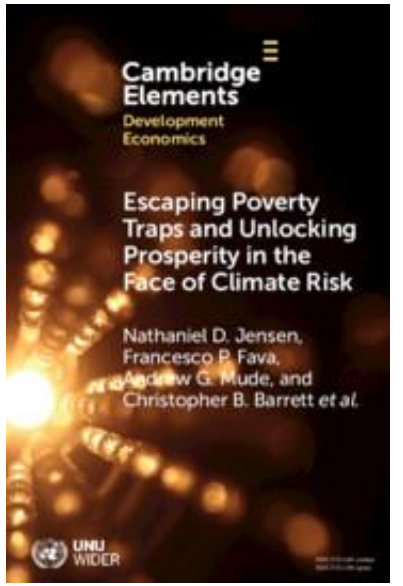
Use to design interventions that can build resilience.

Example: index-based livestock insurance (IBLI)

- HFL survey data enabled original IBLI design and piloting. (Chantarat et al. *JRI* 2013)
- Enabled quasi-experimental comparison of IBLI vs. cash transfers. (Jensen et al. *JDE* 2017)
- End result: IBLI scaled from pilot in small parts of 2 countries to broader use in 4 (and growing: Kenya Livestock Insurance Program, DRIVE). ~600K people covered to date, expected/targeted >1.5M by end-2025.

Pilot areas



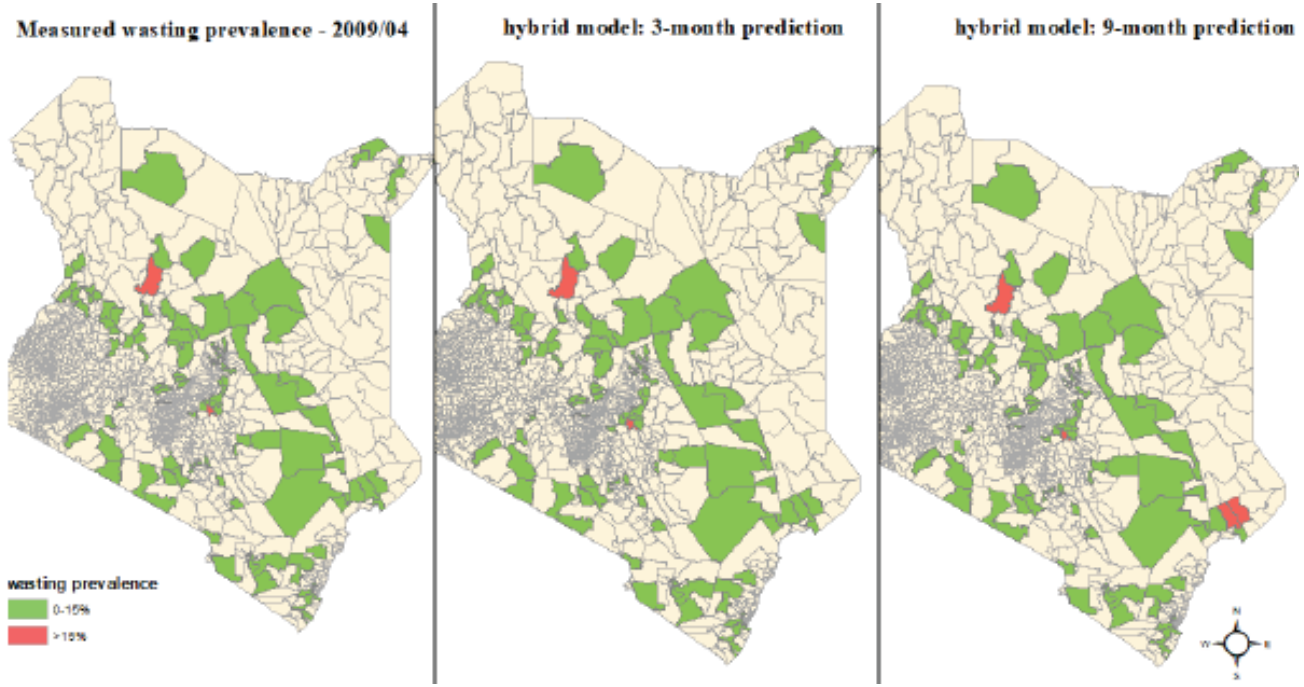


Cornell University

What role for HFL survey data in polycrisis era?

Use for early warning and geographic targeting

Most empirical economic analysis inherently backward looking; policy advice is forward-looking. HFL survey data allow for (i) rapid updating, (ii) nowcasting/forecasting for early warning and geographic targeting. High value of high frequency sentinel site data (Constenla et al. in review).



Cornell University

Why high-frequency longitudinal (HFL) national survey data?

In polycrisis era, HFL survey data are more valuable than ever. Help answer policymakers' questions in face of growing risk and shocks:

- **What gains from emergency preparedness?** Behavioral responses to risk/shocks measured right.
- **What benefits from response to shocks?** Transitory vs. seasonal vs. chronic and possibility of poverty traps. Inference using 'within' variation.
- **Which tool(s) to use?** More A/B testing, less $H_0=0$ testing. Test under varied states of nature.
- **Where and when to respond?** Nowcasting/forecasting maps
- **How to customized bundled responses?** Sufficient variation in sites to identify heterogeneous and interaction effects.

Researchers have lots to do to help build resilience to polycrisis!

Cornell University

Thank you

Thank you for your time and interest!

Follow-up questions/comments?
Email me: cbb2@cornell.edu