

High-frequency survey data for development research: Human capital and wellbeing

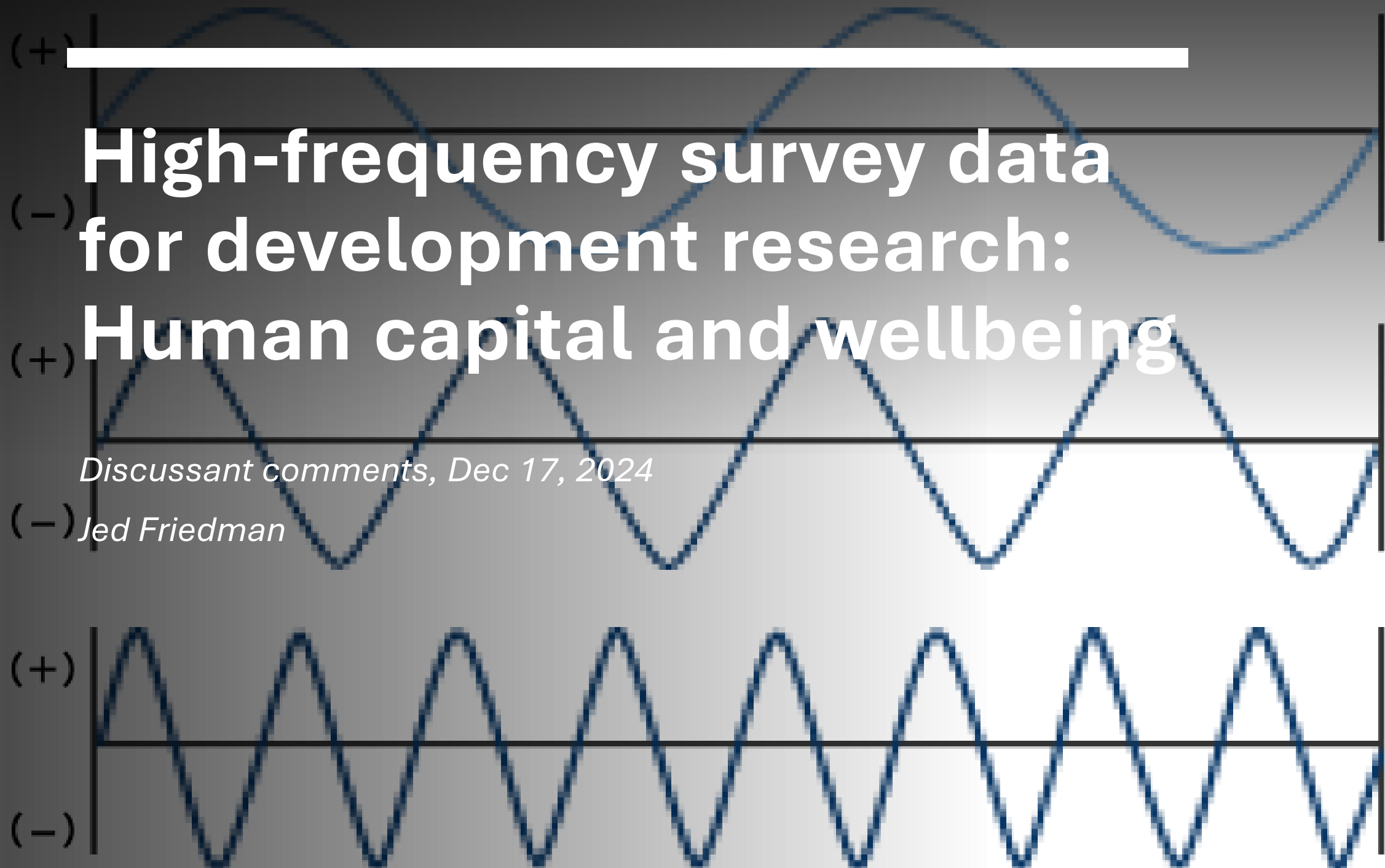
Discussant comments, Dec 17, 2024

Jed Friedman

Low

Frequency

High



Promises and Pitfalls with HFPS (high-frequency phone surveys)

Lower Costs, Faster Data Collection, Real-time Data

High-frequency phone survey data is less expensive than traditional survey methods - more accessible to organizations with limited resources. High-frequency of collection can identify dynamics missed by less-frequent traditional surveys

Unknown interview mode effects in development settings

Some evidence that complex measures such as household consumption are under-reported over phone (e.g. [Abate et al. 2024](#)), but what about mental health, student learning, and health expenditures?

Selective response

Differential non-response of phone vs face-to-face (F2F). Seemingly well-studied with this HFPS collection (e.g. [Ambel et al. 2021](#)).

What are the exciting questions to investigate with high-frequency data?

What can high-frequency data reveal in the human capital and well-being space that cannot be studied with other data?

Stress under Shocks: Food Insecurity, Weather Shocks, and Mental Health in Malawi (I)

Question investigated

What are the mental health consequences of economic deprivation, as identified through weather shocks and self-reported food insecurity? Security is indeed fundamental to well-being ([Slavich et al. 2023](#)).

Related literature

Insecurity is stressful - lots of work on this, especially in conflict and disaster settings. For economic shocks, mental health detriments may persist after economy rebounds ([Friedman and Thomas, 2009](#)). Benefit in documenting the short- and medium-run dynamics.

Data treatment

Response weights used to address possible phone-survey selection bias. Are there any groups particularly under-represented? More description of data environment beneficial.

Stress under Shocks: Food Insecurity, Weather Shocks, and Mental Health in Malawi (II)

Identification

Food insecurity → poor mental health, weather shocks → food insecurity, but weather shocks → poor mental health through other channels: difficult to understand story as many sources of insecurity apart from food insecurity.

Nearby weather is arguably not valid IV for insecurity as weather is highly spatially correlated, community FEs don't help

Meaning of outcome variable

Not sure if PHQ-8 is locally validated, nevertheless should look at threshold effects around pre-specified cut-offs, for both PHQ-8 and FIES.

Additional directions

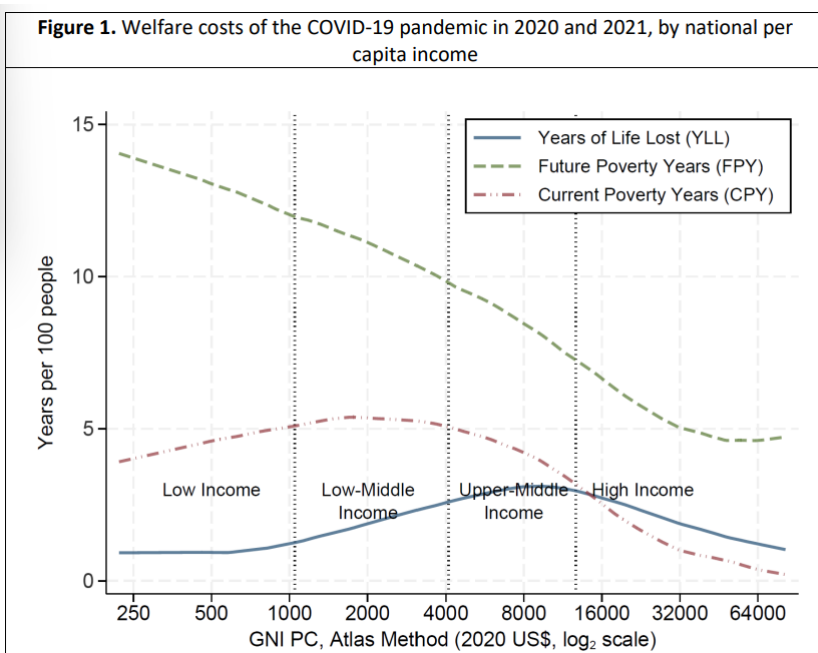
Can the authors investigate differing impacts of differential shocks? I.e. is food insecurity a particularly stressful experience w.r.t. general financial insecurity? Is it more difficult to “recover” from such shocks?

Document the short run dynamics between food insecurity and mental health. Response differentials by gender – females typically report higher adverse mental health, what extent attributed to sensitivity to shocks?

The Impact of the COVID-19 Pandemic on Students' Learning Outcomes in Uganda (I)

Question investigated

Documents COVID-19 period impacts on school enrollment class progression. Uganda experienced one of the longest formal school closures in the world, and welfare consequences of learning loss will very likely outpace losses due to mortality and contemporaneous income decline (see [Decerf et al. 2024](#)).



Related literature

Much emergent work on this general topic, especially with regards to remediation ([Singh et al 2024](#)). Need to better understand learning resilience strategies that worked over this period.

Data treatment

Assume this uses the Uganda 1st and 8th round panel – and corrects for phone-survey selection bias (?). Are any groups particularly under-represented? [I.e. concern with selection within strata]. More description of data environment (e.g. why no data from pre-pandemic?)



The Impact of the COVID-19 Pandemic on Students' Learning Outcomes in Uganda (II)

Identification

In a “normal” two-year period there will be time effects within cohort – what is the pre-pandemic rate of grade progression and schooling interruption? Benchmark results against this “counterfactual”. Are certain age-groups more vulnerable to shock?

Analysis

Grade progression is ordered, so consider ordered probit, poisson, etc. In general, so few predictors are predictive – more descriptives before multivar analysis would help understanding.

Not clear how outcomes are constructed in relation to pre-pandemic baseline (round 1 is post-start of pandemic). Also, if some school restrictions in place in round 8, ideally can continue analysis through end of restrictive period.

Additional directions

Learning - Is the cited learning loss conditional on enrollment and grade attainment?

Meaning of (and heterogeneity in) “instructional status” – e.g. continual engagement with tutor different from presence of instructional books in household

Dynamics – exploit the high-frequency nature of the data to measure exposure duration to “instructional status”

Do We Need More Frequent Data to Measure of Health Out-of-Pocket Expenditures? (I)

Question investigated

Does the infra-annual distribution of health OOP (from panel) differ from traditionally constructed annualized estimates of health OOP (from x-section)? Exploits the high-frequency nature of data in 5 countries to construct panel and 2 x-sectional estimates.

Great use of HF data (although interpretation tricky due to lack of benchmark).

Related literature

Unaware of other work – but related work on psychology of survey response and consumption measurement is relevant. [Friedman et al. 2017](#) suggest differential response patterns across extensive and intensive margin of spending on food items, with telescoping in short recall periods. Confirmed by [Abate et al. 2022](#).

Data treatment

19 survey rounds from 5 countries. Attrition is low but non-zero. Corrections for attrition and possible phone-survey selection bias? More description of data environment – including any reported mortality and health shocks in household.

Do We Need More Frequent Data to Measure of Health Out-of-Pocket Expenditures? (II)

Findings and interpretation

Frequency of spending much higher in panel than repeated x-section, but annualized totals not very different (and sometimes lower in panel!). Is OOP value largely driven by one-off high spending events, i.e. hospitalizations, that do not repeat?

Amazing x-country variation in total spending and incidence, e.g. Nigeria much higher than most others – driven by hh vs. individual nature of questionnaire? Or difference in presence of informal health providers?

Who is spending?

The 80/20 rule (and 50/5 rule) in health spending – but the 20% is not stable over time. Who are these “20%” in the data, does profile vary across mode?

Additional directions

Implications for measurement - how much do we miss with a single x-section? Are gains in measure “worth” repeated rounds?

Repeated survey as an intervention ([Zwane et al. 2011](#)) – does varying lag between survey rounds affect reporting?

Authors of [Abate et al. 2024](#) may help understand health expenditure measured over phone vs. F2F.



Internet Access and Youth's Mental Health and Well-being: Evidence from Ethiopia (I)

Question investigated

What does internet access do to adolescents' well-being and mental health in low-income country context?

Related literature

Related work mostly from high-income countries. There is much concern around social media use, although systematic reviews find mixed results, typically of small magnitude ([Odgers and Jensen, 2020](#)). Causality is hard to establish.

Data treatment

YL round 6 separated in 2 “calls” – is attrition present? Possible to use response weights (e.g. w.r.t. YL round 5) to address possible phone-survey selection bias? More description of data environment would be helpful – distributions of outcomes, etc.

Internet Access and Youth's Mental Health and Well-being: Evidence from Ethiopia (II)

Identification

Presents both OLS and IV with TWFE. On the validity of IV (community average internet access): if internet affects behavior and views of peers, parents, etc, then adolescent mental health affected independent of individual access.

Related, the IV estimates are order of magnitude higher than OLS (i.e. moving from 0.3 to 3 SD impact) – not clear why. Doesn't appear to be weak instrument. Check measurement error in internet access, relative variability in regressor and instrument, LATE.

Meaning of outcome variables

Not sure if GAD-7 and PHQ-8 are locally validated, especially for adolescents, nevertheless good to look at threshold effects around pre-specified cut-offs.

What does Cantril ladder measure in adolescents in general? Appears to emphasize power and wealth over broader well-being and relationships in adults ([Nilsson et al. 2024](#)).

Additional directions

Important to understand what these adolescents are doing online – (i.e. replacing in-person with internet socializing, or planning with off-line friends)? Perhaps qual research in similar settings can be leveraged?

Striking that cannot replicate ID strategy for older cohort (they all have internet access at this point?). What about OLS? Can similar analysis be done on this cohort as assessed in Round 5?