Comments on “Comprehensive Data Quality Studies as a Component of Poverty Assessments” by Brooks, Lippe and Waibel

Alemayehu Ambel
LSMS, World Bank
Study Summary

• **Objective:** Determinants of non-sampling errors in income related variables

• **Model:**
  - 5 Dependent Variables: Missing Values, Refusals, Implausible Values, Outliers, & Any of these (Error Value)
  - Independent vars: Enumerator, Respondent, Interview/ Survey Environment

• **Data:** 2017 Thailand and Vietnam – long –term panel survey

• **Methodology:** Probit and OLS (% of errors)

• **Results:** Enumerator and respondent characteristics matter

• **Excellent work!** Key contributions include:
  - innovations in para data collected at different stages,
  - focus on important variables, and
  - useful lessons for a CAPI environment.
Questions/Suggestions

Dependent Variables: More information / explanation in the dependent variables:

• Difference between implausible values and outliers.
• Outliers – coverage or legitimate (sampling error)?
• Target variables (income related)- how are they aggregated and how is the relative importance of each element. For example, is a missing on main income source the same as a minor source?
Questions/Suggestions

Independent Variables

• What is the effect of pre-fieldwork preparation/ questionnaire design, translation, & training. We do these to prevent enumerator and respondent effects.
  • We can shift responsibilities from interviewers and respondents to designers in CAPI environment
  • The implication- with adequate preparation – the effect will be minimal

Conceptual Framework

• The current framework explains more of the motivation than the actual exercise in the study. What about using the Total Survey Error Framework (e.g. Groves & Lyberg 2010 in public opinion quarterly)?
Questions/Suggestions

Estimation Methodology
• Not clear how the zero values were discarded/censored in the OLS – The results do not mirror the results in the previous tables- tobit?
• Combined sample? I think, the similarity of the two surveys is an opportunity to do so.

Results:
• Conflicting results in different specifications
• Not significant – in so many cases for some factors- but still included in the findings as relevant
• When significant – not in the right model - intuitiveness
Thank You!