Missing(ness) in Action:
Selectivity Bias in GPS-Based Land Area Measurements

Talip Kilic*[†], Alberto Zezza[†], Calogero Carletto[†], Sara Savastano[ǂ]

* Corresponding Author: tkilic@worldbank.org
[†] Development Research Group, World Bank, Washington, DC, USA.
[ǂ] Department of Economics and Finance, University of Rome - Tor Vergata, Italy.

Land area is a fundamental component of agricultural statistics, and of analyses undertaken by agricultural economists. While household surveys in developing countries have traditionally relied on farmers' own, potentially error-prone, assessments to collect land area information, the availability of affordable and increasingly reliable Global Positioning System (GPS) units has recently made GPS-based area measurement a practical alternative. Nonetheless, survey implementing agencies typically require interviewers to record GPS-based area measurements *only* for plots within a given radius of dwelling locations, in order to reduce survey costs, keep interview durations within reasonable limits, and avoid the difficulty of asking respondents to accompany interviewers to distant plots. It is, therefore, common for as much as a third of the sample plots not to be measured, and available research has not shed light on the possible selection bias in analyses that rely on partial data due to gaps in GPS-based area measures. We explore the systematic patterns of missingness in GPS-based plot areas, and investigate their empirical implications in the context of the inverse scale-land productivity relationship. Using Multiple Imputation (MI) to predict missing GPS-based plot areas in nationally-representative survey data from Uganda and Tanzania, we highlight the potential of MI in reliably simulating the missing data, and document a stronger inverse scale-land productivity relationship with the complete data. Our study demonstrates the use of judiciously reconstructed GPS-based areas in alleviating concerns over potential measurement error in farmer–reported area assessments, and with regards to systematic bias in plot selection for GPS-based area measurement.

Keywords: Global Positioning System, Land Area Measurement, Land Productivity, Multiple Imputation