Measurement error, validation data, and program evaluation

Joshua W. Deutschmann[1] and Emilia Tjernström[2]

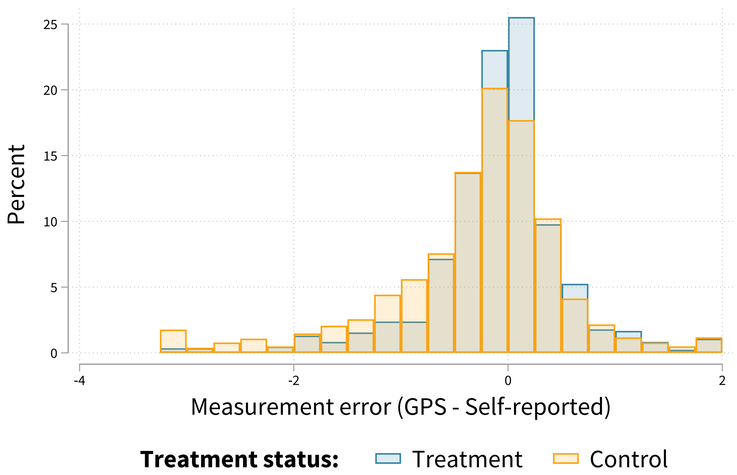
[1]University of Wisconsin-Madison, [2]University of Sydney

Introduction

- Many outcomes of interest in empirical work are costly to measure objectively
- Self-reported measures from surveys are often used in place of these costly objective measures
- What if self-reported measures exhibit measurement error that is correlated with treatment status in an experiment?

Differential Misreporting

Context: field experiment in Kenya:
(Deutschamnn et al. 2019).
Farmers misreport their cultivated acreage compared to GPS measurements, but misreporting is smaller on average among treated farmers.



Econometric setup

Two possible measures of an outcome:
(1) Self reported (Y);
(2) Objectively measured (Y^*)
Goal: Estimate treatment effect

$$Y^* = \beta_0 + \beta_1 T + \epsilon \quad (1)$$

Measurement error could be differential by treatment status:

$$Y - Y^* = \gamma_0 + \gamma_1 T + \mu \quad (2)$$

Estimating (1) with Y would yield a biased estimate of the treatment effect if $\gamma_1 \neq 0$:

$$Y = (\beta_0 + \gamma_0) + (\beta_1 + \gamma_1) T + (\epsilon + \mu) \quad (3)$$

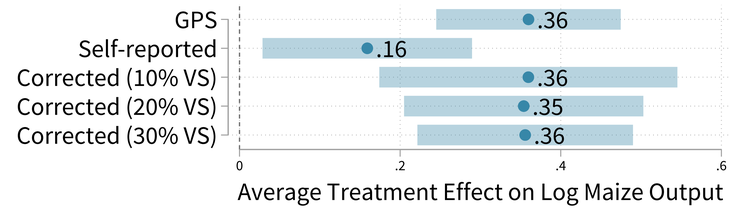
"De-biasing" with validation data

Following Buonaccorsi and Tosteson (1993), Carroll et al (2006):

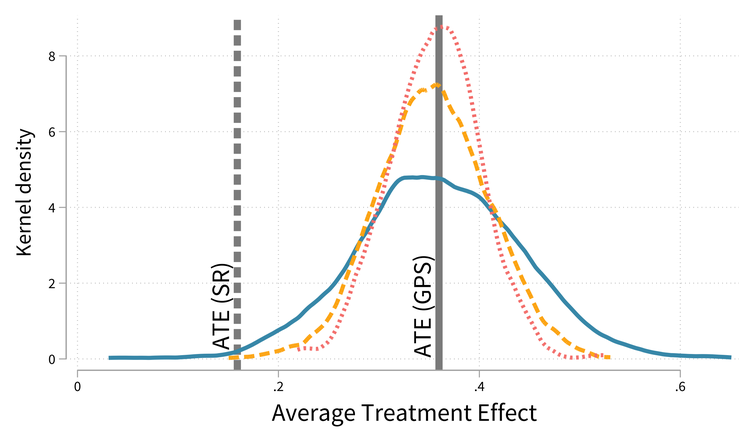
- In a subset (the *validation dataset*), collect both self-reported and unbiased estimates of the outcome of interest
- Estimate treatment effect ($\hat{\beta}^v$) and measurement error ($\hat{\gamma}^v$) using validation dataset
- Generate "de-biased" outcome ($\hat{Y}^f$) in full sample using $\hat{\gamma}^v$
- Re-estimate treatment effect ($\hat{\beta}^f$) using "de-biased" outcome
- Form best weighted combination of two estimates using joint covariance matrix and bootstrap

Method demonstration

Treatment effect estimates and mean results from simulated corrections
Outcome: log maize output



Distribution of estimated treatment effects from 2000 simulations



Next Steps

- Demonstrate with fully simulated data
- Augment with machine learning tools

Agricultural and Applied Economics
UNIVERSITY OF WISCONSIN–MADISON

Comments Welcome:
jdeutschmann@wisc.edu