Cornell University

Correlated Non-Classical Measurement Errors, 'Second Best' Policy Inference and the Inverse Size-Productivity Relationship in Agriculture

Kibrom A. Abay, ILRI
Gashaw T. Abate, IFPRI
Christopher B. Barrett, Cornell
Tanguy Bernard, IFPRI

30th International Conference of Agricultural Economists
Vancouver, BC, Canada
July 30, 2018

Introduction

- Measurement drives analysis
- The quality of descriptive and predictive evidence is only as good as the underlying measures used
- The "credibility revolution" may benefit from a "measurement revolution" to improve causal inference
- Much recent research on improved measurement finds widespread non-classical measurement error (NCME) in key variables relevant to agriculture and beyond
- Conventional recommendation: employ better measurement methods … We agree, of course!

Cornell University

Introduction

- Yet often ***multiple*** variables suffer from NCME
- Not all such variables are amenable to more accurate measures
- Errors may be correlated for any of several reasons, e.g.:

  - Systematic (under/over) reporting

  - Rounding of multiple variables ('focal point bias')

  - Optimal prediction errors (Hyslop & Imbens, 2001)

Introduction

In this paper, we explore:

- What are the consequences of incomplete correction for NCME for inference, especially if errors are correlated?
- Does correction for one, but not both, variables' NCME reduce bias and improve inference?
- Apply to ag size-productivity relationship (SPR) about which recent studies suggest the inverse relationship might just be an artefact of NCME in land (Carletto et al., 2013, 2015) or output (Gourlay et al., 2017; Desiere and Jolliffe, 2018)

Prior lit focused on univariate NCME problem, despite the possibility of two-sided NCME problems, e.g., estimating SPR, elasticity of labor supply, etc.

Preview of Key Results

- Bias due to NCME is analytically ambiguous
- With correlated NCME, correcting for measurement error in just one variable can **aggravate** the bias: a novel and general ‘second best’ result.
- In our data, NCME in farmer self-reported plot size (area) and output is considerable, patterned, and correlated.
- NCME generates a spurious inverse SPR. The naïve estimate (i.e., no correction for NCME in area or output) is statistically equivalent to the true (i.e., fully corrected) estimate and significantly **less biased** than are estimates based on correction for just one source of NCME.

Analytical Framework

- Consider the relationship between log output and log area:

$$Y^* = \theta X^* + \varepsilon$$

- **The size-productivity (i.e., yield) relationship is estimated**:

$$Y^* - X^* = \beta X^* + \varepsilon; \beta \equiv \theta - 1$$

- But we observe neither Y^* nor X^*
- Rather we observe, with possible NCME (u and v), Y and X
- This yields a **two-sided** measurement problem
- Quite generally, β is the true SPR parameter.

Core Analytical Results

We analytically derive the effects of measurement error

Let - δ be NCME parameter in output on output (*Y** on *Y*)

- λ be NCME parameter in output on area (*Y** on *X*)
- α be NCME parameter in area on area (*X** on *X*)
- Φ be V(*X**)/V(*X*)
- π be COV(*u*,*v*) in ME terms

The bias arising from ME depends on the source

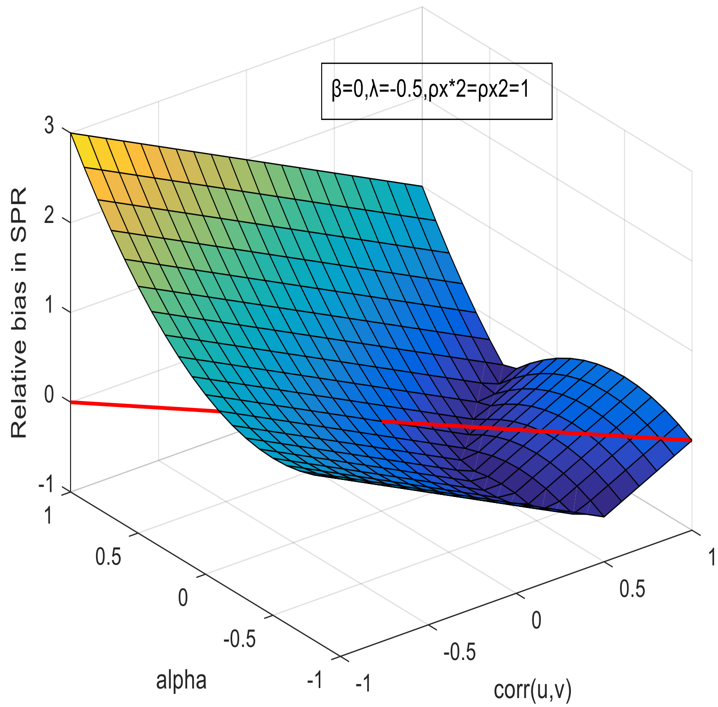
Table 1: Summary of Analytical Results

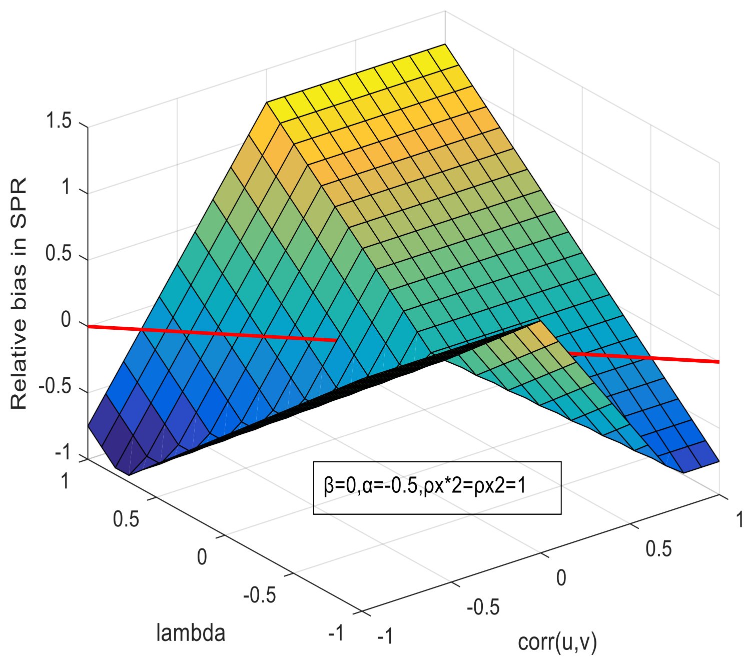
Source of non-classical measurement error	Key Parameters δ	λ	α	π	Estimated SPR	Direction of bias on the SPR
No error	0	0	0	0	β	No bias
Error in production	<0	0	0	0	$(1+\delta)\beta$	Underestimation of ISPR
Error in production	*	<0	0	0	$\beta+\lambda$	Overestimation of ISPR
Error in plot size	*	0	<0	0	$\beta(1+\alpha)\Phi-\alpha(1+\alpha)\Phi$	Ambiguous
Error in both	*	<0	<0	0	$\beta(1+\alpha)\Phi-\alpha(1+\alpha)\Phi-\lambda\Phi$	Ambiguous
Error in both	*	<0	<0	>0	$\beta(1+\alpha)\Phi-\alpha(1+\alpha)\Phi-\lambda\Phi+\pi\Phi$	Ambiguous

Correcting just one source of NCME can aggravate rather than reduce bias in estimates of β!

A numerical illustration:

RB>(<)0 means correcting DV only reduces (increases) bias





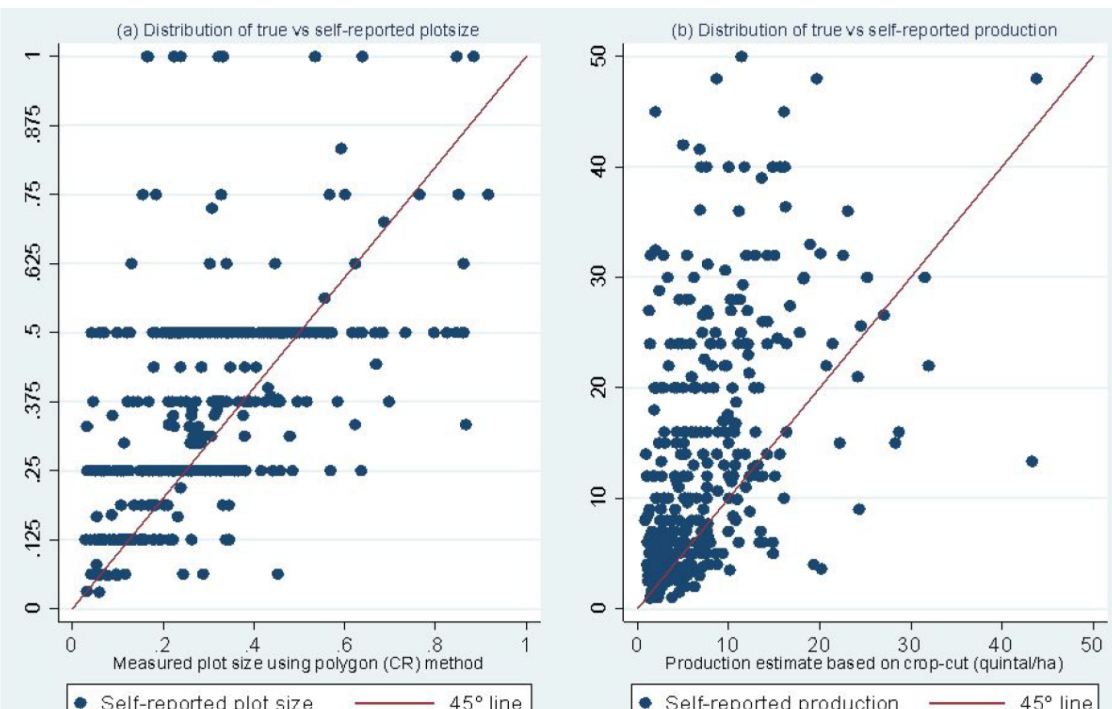
Data

- Data collected from rural wheat farmers in Ethiopia
- 36 villages from 18 *woredas* in Oromia, Amhara, Tigray
- Random sample of 1wheat plot/farmer 2013/14 *meher* season
- Crop-cut measurement in Nov/Dec 2013 by experts from the Central Statistical Agency.
- Household survey conducted in February and March 2014
- Land area measures: self-reported, compass-and-rope
- Production measures: self-reported, crop-cut

Characterizing Measurement Errors

Considerable measurement error exists in both area and output. Both exhibit NCME. Lots of focal point bias. And NCME are correlated.

In the terms of the analytical model, in these data: α, δ, λ<0, π>0, and Φ>1



Correlation b/n NCMEs

Expl vars	ln (SR output/ crop-cut output)
ln (land area bias)	0.377***
	(0.131)
ln (CR Plot size)	-0.363***
	(0.110)
r^2	0.55

Hh/plot chars, village dummies not shown. N=365

Consequences of NCME for SPR Estimates

Benchmark – i.e. fully corrected – (naïve, uncorrected) results in red (green)

Key explanatory variable	Dependent variable			
	CCO	**CCO**	**SRO**	**SRO**
CRA	**-0.104**		-0.668***	
	(0.063)		(0.072)	
SRA		-0.579***		**-0.187*****
		(0.079)		**(0.077)**
r^2	0.56	0.53	0.60	0.48

CCO = cut output, SRO = self-reported output; CRA = compass & rope area, SRA = self-reported area, all in natural log terms. HH/plot characteristics, kebele FEs included. SEs clustered by kebele.

In our date, true yield is statistically invariant to true area with small magnitude point estimate. NCME generates a biased estimate of an inverse SPR (like most of the lit).

Correction only for NCME in just one variable yields *<u>most biased</u>* SPR estimates.

Use of two erroneous measures yields biased (exaggerated) SPR estimate but *<u>statistically significantly lower bias</u>* than when correct just one variable. Indeed, point estimates not statistically significantly different from those based on correction for both NCMEs!

Consequences of NCME for SPR Estimates

Summary of Empirical Relationships

Source of non-classical measurement error	Key empirically estimated parameters				Estimated SPR	Relative implication on the SPR
	δ	α	λ	π		
No error	NA	NA	NA	NA	-0.104 (0.062)	Insignificant ISPR estimated
Error in production	-0.671*** (0.054)	-0.558*** (0.080)	NA	NA	-0.659*** (0.074)	Strongest ISPR estimated
Error in plot size	NA	NA	-0.532*** (0.042)	NA	-0.578*** (0.077)	Strong ISPR estimated
Error in both	-0.670*** (0.055)	-0.558*** (0.080)	-0.532*** (0.042)	>0	-0.204*** (0.073)	Weaker ISPR estimated

Compared to the benchmark:

- **Strong ISPR estimated when we ignore either error**
- **Weak ISPR estimated when we ignore both errors**
- **Results match our general, analytical predictions**

Conclusions

- Much conventional survey data is measured with (non-classical) error.
- Improving measurement is essential to good description and inference.
- But beware hubris of correcting measurement of just one variable.
- We analytically and empirically study correlated NCME in the context of the longstanding size-productivity relationship (SPR) puzzle in developing country agriculture.
- We show that analytically the signs and magnitude of bias in SPR estimates – and other, analytically similar estimates (e.g., avg wage rates as functions of hours worked) – are ambiguous when NCME exists in both output (dependent variable) and area (independent variable)
- We show that accounting for measurement error in only one of the two mismeasured variables may (counterintuitively) *worsen* the bias in key parameter estimates, as compared to correcting neither … a novel and general 'second best' result.
- In our data, the inverse relationship appears very weak in corrected data, but a strong inverse relationship exists in the presence of NCME.

Cornell University

Thank you for your interest and comments

Variable	Description	Mean	Std. Dev.	Min	Max	Obs.
Area SR	Self-reported area size for reference plot (ha)	0.42	0.36	0.03	4.00	488
Area measured	Measured area size during crop-cut (ha)	0.37	0.39	0.03	3.80	483
Production SR	Self-reported production for reference plot (qt.)	21.05	19.18	0.50	120.00	488
Production measured	Estimated production based on crop-cut (qt.)	8.98	9.91	0.81	101.5	365
Yield SR	Self-reported (production/area), (qt./ha)	30.69	18.18	1.00	96.00	488
Yield measured	Measured (production/area), (qt./ha)	28.23	15.05	2.78	95.38	366
Age of HH head	Age of the household head in completed years	45.67	10.84	20.00	77.00	488
Gender of HH head	Gender of the household head	0.86	0.34	0.00	1.00	488
HH size	Number of household members	6.79	2.39	1.00	16.00	488
Literacy of HH head	=1 if the household head is literate	0.64	0.48	0.00	1.00	488
No. of corners	Number of corners of the reference plot	8.74	4.88	4.00	23.00	484
Closure error	Closure error in plot area measurement	1.09	0.89	0.02	4.50	483
Area unit†	=1 if farmers used ha for SR area measurement	0.39	0.49	0.00	1.00	488
Total owned area†	Total farm land owned by sample farmers	2.31	2.14	0.00	20.00	488
Crop-cut to edge	Distance between the crop-cut and shortest or closest plot edge (meters)	25.83	18.57	1.40	148.00	374
Production unit	=1 if farmers used kg for SR production measurement	0.59	0.49	0.00	1.00	488
Total wheat produced†	Total wheat production during 2013/14 *meher*	46.64	75.26	0.95	755.00	488
Soil fertility†						
High	=1 if the fertility of the reference plot is high	0.44	0.49	0.00	1.00	488
Medium	=1 if the fertility of the reference plot is medium	0.49	0.50	0.00	1.00	488
Poor	=1 if the fertility of the reference plot is poor	0.07	0.26	0.00	1.00	488
Soil color†						
Red	=1 if the color of the reference plot is red	0.26	0.44	0.00	1.00	488
Black	=1 if the color of the reference plot is black	0.54	0.49	0.00	1.00	488
Grey/sand	=1 if the color of the reference plot is grey or sandy	0.20	0.40	0.00	1.00	488
Distance to plot†	Walking distance between the dwelling and the plot	30.98	9.94	0.00	120.00	488
Plot ownership	=1 if the reference plot owned by the HH	0.82	0.38	0.00	1.00	488