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

Kibrom A. Abay
University of Copenhagen

Gashaw T. Abate
International Food Policy Research Institute

Christopher B. Barrett
Cornell University

Tanguy Bernard
University of Bordeaux and International Food Policy Research Institute

February 2018

Abstract

We show analytically and empirically that non-classical measurement errors (NCME) in the two key variables in a hypothesized relationship can bias the estimated relationship between them in any direction. Furthermore, if these measurement errors are correlated, correcting for either NCME alone can aggravate bias in the parameter estimate of interest relative to ignoring mismeasurement in both variables, a ‘second best’ result with implications for a broad class of economic phenomena of policy interest. We use numerical simulation to illustrate the parameter space over which a second best approach of not correcting one variable’s NCME dominates correcting it. We then illustrate these results empirically by demonstrating the implications of mismeasured agricultural output and plot size for the long-debated (inverse) relationship between size and productivity. Our data from Ethiopia show large discrepancies between farmer self-reported and directly measured values of crop output and plot size. These NCME are strongly, negatively correlated with the true variable values and strongly, positively correlated with one another. In these data, correlated NCME generate a strong but largely spurious estimated inverse size-productivity relationship. And in line with our analytical result, correcting for just one source of NCME aggravates the bias in the parameter estimate of interest.

JEL Codes: C81, O12, Q12, Q15

Keywords: Agricultural policy, measurement, bias, Ethiopia.

1 Introduction

Measurement drives analysis. The quality of descriptive and predictive evidence is only as good as the underlying measures used to test key hypotheses. In recent years, empirical researchers have begun to devote considerably more effort to careful measurement and to explore the consequence of different measurement methods for key variables of direct policy relevance.[1] Of particular concern is non-classical measurement error (NCME), which occurs when the error in measuring a variable of interest is correlated with the true value of that variable, with the true values of other variables in the model, or with the errors in measuring those values (Bound et al. (2001)).[2] Many papers have clearly demonstrated the widespread prevalence of NCME and its relevance for policy inference in a range of fields, especially labor (e.g., Borjas (1980); Bound and Krueger (1991); Bound et al. (1994); French (2004); Kim and Solon (2005); Arthi et al. (2018)), consumer behavior (Gibson and Kim (2010); Gibson et al. (2015)), development (Baird and Özler (2012); Beegle et al. (2012); Chao et al. (2012); Desiere and Jolliffe (2018)), health (Das et al. (2012); Larsen et al. (2017)), and agriculture (De Groote and Traoré (2005); Carletto et al. (2013, 2015); Gourlay et al. (2017)). The sensible guidance provided by that literature is to employ better measurement methods so as to reduce error. The rise of improved techniques based on high resolution remote sensing, mobile phone, imagery, global positioning system (GPS) and biomarker data, along with electronic survey data entry, steadily opens up new possibilities for reducing policy-relevant measurement error.

Yet in many domains, multiple variables fall prey to NCME. Moreover, mismeasurements may often be correlated between variables, for any of several reasons. For example, survey respondents might consciously and systematically underreport assets and earnings in order to reduce prospective tax liabilities or to increase the likelihood of being deemed eligible for some benefit. Or unconscious error may arise from rounding (sometimes known as 'focal point bias') so that variables that naturally exhibit positive skewness, such as asset holdings or earnings, will commonly exhibit upwardly biased and positively correlated measurement error as a result. Or one mismeasured variable might be used by a respondent to generate an optimal prediction of another variable (Hyslop and Imbens (2001)), resulting in correlated

[1] The special issue of *Journal of Development Economics* on measurement and survey design, introduced by McKenzie and Rosenzweig (2012), was a watershed event pushing more careful measurement in development economics. Ozler (2013)'s Development Impact blog entry helped call the development communitys attention to these important issues more broadly.

[2] As is well known, classical measurement error is just a special case of the more general NCME form we study. Classical measurement error generates attenuation bias in parameter estimates and artificially inflates variance that may provide misleading description of, for example, income inequality or mobility (Gottschalk and Huynh (2010)). Such bias declines as panel survey intervals increase (Naschold and Barrett (2011)).

measurement errors.

If multiple variables are measured with error but only some are amenable to correction, does correction for just one, but not both, otherwise-mismeasured variables reduce bias and improve inference, especially if those measurement errors are correlated? To the best of our knowledge, this important question has not yet been explored in the literature. Yet correlated NCMEs matter for the same reason that omitted relevant variables matter because each NCME is, by definition, correlated with a relevant variable.With multiple NCME, the possibility of biases of opposing signs with positive correlation between the measurement errors, or of biases of the same sign with negative correlation between the measurement errors, implies that correcting for one source of NCME does not necessarily move the resulting estimate closer to the true parameter value; indeed, it could increase bias. Hence, our central analytical finding that correcting for one source of measurement error might magnify bias in a parameter estimate of interest and its corollary, that if one cannot correct for both sources of measurement error, a second best estimate based on multiple NCME may be preferable in the sense of reduced bias.

This problem arises for a wide range of economic questions. For example, estimates of the wage elasticity of labor supply may be subject to error in measures of earnings and hours worked, the latter of which serves as both the dependent variable and the denominator of the standard wage measure, leading to division bias (Borjas (1980)). Correlated errors in nominal output and price measures may similarly bias the estimated relationship between real output or total factor productivity and inflation (Diewert and Fox (1999)). And measurement error in childrens ages, which are likely correlated with errors in height or weight measures used jointly to construct standard anthropometric indicators such as height-for-age, can significantly bias estimates of the determinants of child health (Larsen et al. (2017)).

In this paper we present quite general analytical findings, and explore their implications numerically. But we frame the initial analytical portion in relation to our empirical part, which explores the consequences of correlated NCME in crop output and land area in the long-studied size-productivity relationship (SPR) in agriculture. The SPR has been studied extensively because of its considerable implications for agricultural development policy. For decades, findings of an inverse relationship were widely invoked to support land reform programs and substantiate claims of widespread factor market failures that justify interventions. Earlier studies typically found an inverse relationship between farm size and crop output per unit cultivated area (i.e., yield, a partial productivity indicator), attributing this empirical regularity to factor market imperfections (e.g., Sen (1966); Feder (1985); Barrett (1996)) or omitted land attributes, including soil quality (Benjamin (1995); Assunção and

Braido (2007); Barrett et al. (2010)).[3] Recently, improvements in agricultural data collection have allowed researchers to explore the implication of measurement errors in self-reported production and farm or plot size.[4] Some papers have examined the implication of improved area measurement for estimation of the SPR using GPS measures of the surface area of plots (Carletto et al. (2013); Holden et al. (2013); Carletto et al. (2015)). Most recently, a few papers have explored the implication of measurement errors in farmer self-reported crop output on the estimated SPR using crop-cuts as a more objective measure of production (Gourlay et al. (2017); Desiere and Jolliffe (2018)). Those papers find that, in their data, the inverse relationship is essentially driven by measurement errors associated with self-reported production. The relationship disappears upon using crop-cuts in place of self-reported production.

While these few, recent studies explore the implication of measurement error associated with either area or production, no study has yet considered both measurement problems in a unified framework, much less generalized them beyond the specific case of the SPR. This is particularly crucial if both size and production suffer NCME and these measurement errors are correlated, a problem that bedevils a range of important empirical questions in economics. As we demonstrate, when both production and farm size are inaccurately measured, correcting for measurement error in just one variable is not sufficient to generate a consistent and unbiased estimate of the SPR. Furthermore, while previous studies show similar features of measurement errors in self-reported area and production, we know little as to why they generate conflicting empirical implications as to the effects on the estimated SPR.[5] By studying correlated NCME in a more general setting, we can reconcile these findings. More importantly, we analytically and empirically establish that correcting for mismeasurement of just one variable can aggravate rather than attenuate bias in the SPR estimate.

In what follows, we analytically and empirically characterize the implication of various forms of NCME in self-reported crop production and cultivated area. We first set up a general framework that allows NCME in both output and area as well as potential correlations in these errors. We analytically characterize the implication of alternative features of NCME in output and land area on the estimated SPR. We then empirically demonstrate our analytical

[3] Bevis and Barrett (2017) provide another new behavioral explanation for the inverse relationship between farm size and productivity. They argue that productivity may be higher around the periphery of plots, partly for biophysical edge effect reasons (e.g., improved access to sunlight) but mainly for behavioral reasons (e.g., greater observability of edges). As smaller plots have a greater ratio of edges to interior area, this can explain the commonly observed inverse relationship between productivity and farm size.

[4] We will use the terms output and production synonymously and similarly area and size.

[5] Carletto et al. (2013) find that inaccuracies in land area measurement lead to underestimation of the inverse relationship between plot size and productivity while Carletto et al. (2015) show the opposite.

findings, employing both self-reported and objectives measures of output and area from an agricultural household survey in Ethiopia. For production, we compare farmers self-reported production measures and production estimates based on crop-cuts, which are widely considered the gold standard for measuring agricultural output. We similarly rely on both farmer-reported land area as well as measurements based on compass-and-rope method, also known as Polygon method.[6] Compass-and-rope is considered the most reliable method to accurately measure land area (Keita and Carfagna (2009); Fermont and Benson (2011); Carletto et al. (2015, 2016)).[7] By employing these four different measures of farm size and production, we illustrate empirically the patterns our analytical results predict regarding the long-debated SPR.

We make three contributions to the literature. First, we show that when both the dependent and a key explanatory variable suffers NCME, the effect of these measurement errors on the estimated parameter of interest is analytically ambiguous, depending on several parameters that characterize each mismeasured variable. To the best of our knowledge, this is the first paper to provide a general analytical framework for understanding the inferential implications of multiple, correlated NCME, and of incomplete correction for multiple NCME. In our data, we find that measurement errors in self-reported area and production are strongly correlated. As a result, correcting for either problem alone may not ensure unbiased estimation of the SPR. Indeed, our analytical and empirical exercises show that correcting for either measurement problem alone may even aggravate bias in the SPR estimate relative to ignoring both measurement problems. This is essentially an applied econometric analog to the 'theory of the second best' (Lipsey and Lancaster (1956)) result and serves as a useful caution against overconfidence in the gains from improved measurement of single, key variables.

Second, we empirically corroborate in a new data set the core findings of recent studies (Holden et al. (2013); Gourlay et al. (2017); Desiere and Jolliffe (2018)) that claim that measurement error can explain the inverse relationship observed in farmer self-reported area and productivity data. Our results refine these prior findings by identifying features of measurement errors that can generate a spurious inverse SPR and when correcting for measurement

[6]Also, known as traverse measurement, the method involves measuring the length of each side and the angle of each corner using a measuring rope and a compass and the surface area of the measured plot can then be calculated using trigonometry (De Janvry and Sadoulet (2000); Schoning et al. (2005); Casley and Kumar (1988)). Although the method is cumbersome and time consuming, it remains the approach of choice for specialized data collection due mainly to its accuracy compared to GPS or self-reported measures (Carletto et al. (2015)).

[7]For example, Fermont and Benson (2011) compare farm size measurement using GPS and compass-and-rope, and show that GPS estimates significantly underestimate smaller farm sizes while both methods perform comparably for larger plots (those greater than 0.5 ha).

error might fail to correct for biased SPR estimates. Importantly, that result is not an automatic byproduct of measurement error in area and output, particularly if these errors are correlated.

Third, our analytical framework and data allow us to compare the relative impact of the measurement errors in self-reported production and area on the estimated SPR. We analytically show and empirically find that when both variables suffer from similar measurement errors, inaccuracies associated with production are relatively more consequential. We also document that measurement errors in self-reported production and area may also affect parameter estimates relating productivity to other covariates of interest (e.g., soil characteristics).

Despite our emphasis on the estimated relationship between agricultural productivity and cultivated area, our analytical results have far more general implications. Not only do these findings reinforce previous concerns about recall-based and self-reported agricultural data, they also reveal the existence of an empirical equivalent to the theory of the second best, which holds that when one market failure in an economy cannot be corrected, efficiency may – perhaps counterintuitively – be maximized by introducing an offsetting market distortion. Two market failures may cancel each other out. We similarly demonstrate that when there exist NCME in both the dependent and independent variables of interest, and especially if those errors are correlated, then correcting for just one source of measurement error may, paradoxically, exacerbate the bias in the resulting parameter estimate of interest.

2 Measurement errors in household surveys

Most agricultural research relies on self-reported, recall-based data. Due to cost and logistical considerations, most data are collected through single visit household surveys perhaps repeated over time to generate longitudinal (i.e., panel) data using extensive multi-topic instruments. Respondents are asked to recall and aggregate information often over many months and, in the case of agriculture, sometimes across two or more separate harvests of multiple crop types. While recall and aggregation errors can affect many metrics, they can have especially pronounced consequences for measuring area cultivated and production (i.e., harvested output).

Some such error may be classical, meaning the error itself is mean zero and uncorrelated with the true value of either the dependent variable or any independent variables of interest. For example, farmers in developing countries may lack the level of literacy and numeracy needed to accurately estimate and aggregate land area and crop production measurements, leading to significant, but random and symmetric (around the true value) measurement error

(De Groote and Traor, 2005). In a regression context, it is well known that classical measurement error will under-estimate relationships: either in absolute magnitude, in case the error lies with the independent variable (through attenuation bias); or in statistical significance, if the error lies with the dependent variable (through increase in the estimators variance). In the context of the SPR, classical measurement error will naturally bias estimates towards zero, i.e., toward failure to reject the constant returns to scale null hypothesis.

Non-classical measurement error, in which the error is correlated with the true variable(s) of interest, is of considerably greater concern. Multiple mechanisms might introduce NCME in self-reported land area and crop production. First, farmers may intentionally misreport their land area and crop production so as to conceal wealth and thereby avoid taxes or be found eligible for proxy means tested benefits of various types (Diskin (1997)). This type of misreporting can vary systematically with the true value of farm size, since those with little land or output have little or nothing to hide. Second, farmers may not accurately recall information related to much earlier events; in particular, extended recall periods may cause them to forget details of past events (Beegle et al. (2012); Arthi et al. (2018)) or season-specific harvests (Ali et al. (2009); Howard et al. (1995)).[8] Third, precise and universally applied measurement units may not be widely employed in low-income rural areas where imprecise local measures are commonplace. Traditional units can vary between locations and farming systems, implying that measurement and conversion into standardized units can introduce systematic errors. Finally, respondents may tend to round off values around focal points (e.g., one hectare or one day), a problem that may be more consequential, in percentage terms, for smaller plots and harvests than for larger ones.[9]

While the inverse SPR was long observed in survey data, an emerging literature now argues that measurement errors in either land area or production may generate spurious estimates. On the land measurement side, recent studies relying on GPS devices consistently find evidence that farmers overestimate area for smaller plots and underestimate for larger ones (e.g., De Groote and Traoré (2005); Carletto et al. (2013); Holden et al. (2013); Carletto et al. (2015)). However, the implication of area measurement error on estimating the SPR varies and sometimes contradicts each other. For example, Carletto et al. (2013) document that error in land area measurement underestimates the inverse relationship between farm size and productivity, while Carletto et al. (2015) find that it leads to overestimation of the inverse relationship.

[8] Such recall bias affects many other agricultural metrics, including labor use (Arthi et al. (2018)).

[9] While most of the above reasons apply to farm size and production measurements, there are additional problems that may affect measurement of production. For example, farmers may have forgotten season-specific harvests (Ali et al. (2009); Howard et al. (1995)) and portion of their production given as gifts and/or in-kind payments (David (1978)).

On the production side, two recent studies find that the inverse relationship disappears when using crop-cuts instead of self-reported production. They conclude that the estimated inverse relationship is simply driven by measurement errors associated with production measurement (Gourlay et al. (2017); Desiere and Jolliffe (2018)).[10]

While the above few studies explore the implication of measurement error associated with either production or size, no study has analyzed the implication of measurement errors in both metrics. In many situations both area and output are measured with errors and this may have varying implications relative to the measurement errors in either one alone. This is particularly crucial if measurement errors in crop production and farm size are correlated. Intuitively, measurement errors in self-reported production and land area will often be correlated. For example, if households engage in strategic misreporting of land size, they may be more likely to do so as well for their harvests. Similarly, if rounding appears to be the main source of measurement error, rounding in both measures will naturally generate some correlation in measurement errors. For strictly positive-valued variables such as production and land area, upward rounding of production and area generates a potential positive correlation between measurement errors across both variables. The same will be true for positively skewed variables subjected to rounding around focal points, as the density in ranges beneath the focal point will typically exceed the density in the range above it. Below we analytically characterize alternative forms of measurement errors in land and production measurements.

3 Analytical framework

Consider the following relationship between a true outcome of interest Y^* and the true value of a single explanatory variable, X^* , both expressed as the log-transformation of the underlying variables:

$$Y^* = \theta X^* + \varepsilon \tag{1}$$

We assume that the regression error term,ε, is mean zero and uncorrelated with the explanatory variable. Next, assume that we do not observe the true measures of production and land area. Rather we observe error-ridden self-reported measures, Y and X (also expressed in logs), which can be expressed as combinations of true measures and measurement errors as follows:[11]

[10]Gourlay et al. (2017) also used high-resolution remote sensing-based measurements for crop yield estimation.

[11]This specification implies that measurement errors are assumed to be additive in their logarithmic transformed values and hence multiplicative in their original form.

$$
\begin{aligned}
Y &= Y^* + u \\
X &= X^* + v
\end{aligned} \tag{2}
$$

In what follows, we show how the nature of the relationship between the measurement errors, u and v, and X^* affect estimates of the size-productivity relationship (SPR).

3.1 Size-productivity relationship

Letting Y^* and X^* measure production and land area, respectively, we first transform Equation (1) into the more commonly estimated relationship between yield (production/area) and land area cultivated. Recalling that both Y^* and X^* are expressed in logs:

$$
Y^* - X^* = (\theta - 1)X^* + \epsilon = \beta X^* + \varepsilon \tag{3}
$$

Equation (3) is the workhorse estimable equation used in the SPR literature. Our purpose is to analyze the effects of alternative forms of measurement errors in either *production* or *area* on the β estimate. Note that land area enters both the right and left-hand side of Equation (3); measurement error in land area therefore affects both the dependent and independent variables. However, given the relationship between Equations (1) and (2), our representation and implications remain general. In particular, one can use this basic framework to examine generic measurement problems that include four cases of NCME, wherein measurement error in the dependent variable, u, is correlated with (i) the true outcome or (ii) the explanatory variable, or (iii) the measurement error in the explanatory variable, v, is correlated with its true value, or (iv) the measurement errors u and v are correlated.

Case 1: Measurement error in the dependent variable correlated with true value of the dependent variable

$$
u = \delta Y^* + \omega
$$

where ω is a random term uncorrelated with the explanatory variable (land area) and the error term in Equation (3). This implies that:

$$
Y = (1 + \delta)Y^* + \omega
$$

With these features, ordinary least square (OLS) estimation of Equation (3) using self-reported production would result in:

$$\beta^{OLS} = \frac{cov(Y - X^*, X^*)}{var(X^*)} = \frac{cov((1+\delta)(\beta+1)X^* + (1+\delta)\varepsilon + \omega - X^*, X^*))}{var(X^*)} = (1+\delta)\beta \tag{4}$$

In the context of SPR, Case 1 implies measurement error in production that is correlated with the true production level. If we assume, as we find empirically, that the correlation is negative (i.e., that those with the lowest harvest tend to over-estimate output the most in proportion to true output), then OLS using self-reported production weakens the estimated (inverse) relationship between land area and productivity. The degree of underestimation increases with the correlation between measurement error and true production, δ (Bound et al. (2001); Gibson and Kim (2010)).

Part of the correlation reflected in δ may be driven by the correlation between measurement error in self-reported production and true land area.[12] This type of measurement error is more consequential in our context, at least in generating correlation across measurement errors in production and land area, which we analyze next.

Case 2: Measurement error in dependent variable correlated with true value of independent variable

$$u = \lambda X^* + \zeta$$

where ζ is random noise uncorrelated with the true value of farm size and the error term in Equation (1). Using similar substitutions, one obtains the following expression:

$$\beta^{OLS} = \frac{cov(Y - X^*, X^*)}{var(X^*)} = \frac{cov((\beta+1)X^* + \varepsilon + \lambda X^* + \zeta - X^*, X^*))}{var(X^*)} = \beta + \lambda \tag{5}$$

Applied to SPR, Case 2 implies that mismeasurement of production is correlated with farm size, as would occur for instance if smaller farmers were more likely to overestimate output, which appears to be the case in our data. According to Equation (5), such negative correlation induces overestimation of the inverse relationship. Following this reasoning, Desiere and Jolliffe (2018) and Gourlay et al. (2017) provide empirical evidence showing that self-reported production measures can generate a spurious inverse relationship even when productivity is invariant with respect to area.

[12] This is always the case if production is a deterministic function of land area. If production is a probabilistic function of land area, as usually specified in regression production functions, we may theoretically disentangle the correlations between measurement error in self-reported production and measured production caused by land area as well as other (unobservable) factors.

So far, we have considered two cases of measurement errors in production that may result in conflicting implications on the inverse relationship. The first case attenuates the inverse relationship while the second case amplifies it. The overall net effect depends on the relative sizes and sources of the measurement error. Considering similar levels of correlations, the overestimation caused by Case 2 dominates the underestimation associated with Case 1, however, for the expected range of true $|\beta| < 1$. Of particular note, Case 2 can generate an inverse relationship even in the absence of any true relationship, while Case 1 cannot.

Case 3: Measurement error in independent variable correlated with true value of independent variable

$$\upsilon = \alpha X^* + \iota$$

where ι is uncorrelated with the error-free explanatory variable and the error term in equation 1. This, by substitution, implies that:

$$X = (1 + \alpha)X^* + \iota$$

Letting variance of $X^* = \rho x_*^2$ and variance of $\iota = \rho\iota_*^2$, OLS estimation of the relationship in Equation (3) using self-reported land area results in the following parameter:[13]

$$\beta^{OLS} = \frac{\beta(1+\alpha)\rho x_*^2}{(1+\alpha)^2\rho x_*^2 + \rho\iota^2} - \frac{\alpha(1+\alpha)\rho x_*^2}{(1+\alpha)^2\rho x_*^2 + \rho\iota^2} \tag{6}$$

This is the case of measurement error in land area assuming that we have a precise measure of production.

Equation (6) is a generic representation of the consequences of NCME in explanatory variables, including those which can also appear in the left-hand side of regressions if the same variable is used to construct the dependent variable (as in the case of yields). The first term in Equation (6) reflects special cases where the explanatory variable only appears in the right-hand side, and this expression simplifies further to the usual attenuation bias if the measurement error associated with the explanatory variable is classical ($\alpha = 0$).

The second term in Equation (6) arises if and only if the explanatory variable (plot size in our context) also appears in the left-hand side of the estimation, as is true in the SPR literature because yield (i.e., output per unit area) is the dependent variable of interest. This whole term disappears if the measurement error behaves classically ($\alpha = 0$). This is consistent with the fact that classical measurement errors in dependent variables are wholly

[13]To see this, consider $\beta^{OLS} = \frac{cov(Y^* - X, X)}{var(X)} = \frac{cov((\beta+1)X^* + \varepsilon - ((1+\alpha)X^* + \iota), ((1+\alpha)X^* + \iota)}{var((1+\alpha)X^* + \iota)}$ such that $\beta^{OLS} = \frac{(\beta-\alpha)(1+\alpha)\rho x_*^2}{(1+\alpha)^2\rho x_*^2 + \rho\iota^2}$ and the resulting Equation (6)

captured by the regression residual.

Importantly, we cannot know *a priori* the direction of bias associated with self-reported land area measurement in Equation 6. Indeed, we cannot even determine the direction of bias associated with the first term, even when land area only appears in the right-hand side of the Equation (Gibson and Kim (2010)). The direction of bias in the first term mainly depends on the relationship between the variances of self-reported and true area measurements as well as on the size (and sign) of the correlation between the measurement error and true area of land.

Intuitively, there are cases where self-reported land measurement can be expected to have lower variance than the true area measure, for example, if rounding is the main source of measurement error. In these cases, OLS estimation using self-reported farm size will overestimate the inverse relationship if the difference between the two variances is large enough relative to the negative correlation between the measurement error and true area of land. However, the second term in Equation (6) renders ambiguous the overall effect of inaccurate land area measurement.

Case 4: Measurement errors in both dependent and independent variables are correlated

$$cov(u, v) = \pi \neq 0$$

Using analogous substitutions and reformulation, we can show that OLS estimation of the size-productivity relationship using both self-reported measures yields the following identity:[14]

$$\beta^{OLS} = \frac{\beta(1+\alpha)\rho x_*^2}{(1+\alpha)^2\rho x_*^2 + \rho\iota^2} - \frac{\alpha(1+\alpha)\rho x_*^2}{(1+\alpha)^2\rho x_*^2 + \rho\iota^2} + \frac{\lambda\rho x_*^2}{(1+\alpha)^2\rho x_*^2 + \rho\iota^2} + \frac{\pi}{(1+\alpha)^2\rho x_*^2 + \rho\iota^2} \quad (7)$$

Equation (7) is a very general representation encompassing various types of classical and non-classical measurement errors as well as those affecting the dependent and independent variables of interest. For example, the standard attenuation bias associated with classical measurement errors in the explanatory variable of interest (size in our case) can be shown by setting $\lambda = \alpha = \pi = 0$. Similarly, we can show that ignoring measurement error in self-reported production and correlations between both types of measurement errors ($\pi = 0$) results in the special case of Equation (6).

[14]To see this, consider $\beta^{OLS} = \frac{cov(Y-X,X)}{var(X)} = \frac{cov((\beta+1)X^*+\varepsilon+u-((1+\alpha)X^*+\iota),X^*+v)}{var((1+\alpha)X^*+\iota)} = \frac{(\beta-\alpha)(1+\alpha)\rho x_*^2+\lambda\rho x_*^2+\pi}{(1+\alpha)^2\rho x_*^2+\rho\iota^2}$ and the resulting Equation (7)

Again, we are unable to sign the bias in Equation (7). However, the following insights emerge about the resulting SPR estimate. First, even in the absence of correlation in measurement errors ($\pi = 0$) , the fact that both size and productivity suffer from non-classical measurement error ($\alpha \neq 0, \lambda \neq 0$) implies that correcting for measurement errors in one of the variables does not ensure unbiased estimates of the SPR.

Second, if we correct for measurement error in one of our metrics, for example for plot size measurement, Equation (7) reduces to Equation (6) where the inverse relationship between plot-size and productivity would be inflated because of the usually negative correlation between measurement errors in production and true plot size. The resulting bias in the inverse relationship can be more consequential (i.e., greater in magnitude) than ignoring both types of measurement errors, so correcting one measurement error may aggravate the inferential problem, not resolve it. This can be expected for cases where the correlation between measurement errors (the last term in Equation (7)) is positive and strong enough to dampen part of the overestimation in the inverse relationship caused by the third term, a point to which we return below.

Finally, we can assess the relative effects of the different types of measurement errors on the parameter estimate. For example, assuming that there is no statistically significant relationship between farm size and productivity ($\beta = 0$), the first term in Equation (7) disappears. Then, with similar correlations between measurement errors and true area ($\alpha = \lambda$), NCME in self-reported production can generate a spuriously negative SPR estimate (through the third term in Equation (7), while the positive correlation between measurement errors may generate a spuriously positive one. This suggests that in the presence of correlation between measurement errors, the strength of this correlation is a key parameter that may define the direction and size of the bias in the SPR estimate, another point to which we return below.

Overall, the generic analytical expression in Equation (7) refines and qualifies recent studies arguing that measurement error in size or production spuriously generates the standard inverse SPR (Gourlay et al. (2017); Desiere and Jolliffe (2018); Carletto et al. (2013); Holden et al. (2013); Carletto et al. (2015)). Equation (7) highlights the intricacies through which measurement errors in cultivated area and crop output affect this oft-explored hypothesis. Our analytical framework shows that predicting the direction of bias associated with self-reported size and production is more complex than the existing literature suggests. Adding more covariates to Equation (1) also complicates the prediction of the direction of the biases, particularly if these covariates are correlated with cultivated area or the measurement errors, as will commonly be true for agricultural inputs such as labor, fertilizer and machinery use.

3.2 When should NCME be corrected ?

Note that the general expressions in Equation (7) apply to any OLS estimate involving NCME in outcome or explanatory variables, or both, as well as for cases where these measurement errors may be correlated. Table 1 summarizes the key analytical findings. These results underscore some more general insights as to when ignoring known measurement errors will be superior to correcting just one variables measurement error when correcting the second variables measurement error is infeasible, a 'second-best' type of estimation strategy. More specifically, we now use Equations (6) and (7) to illustrate numerically the parameter space over which the 'second best' option of ignoring known measurement error is likely to reduce bias compared to the seemingly-best-feasible option of correcting the known measurement error in a single variable.

Define relative bias, RB, in the SPR estimate as the bias in β arising when one corrects neither mismeasured variable minus the bias arising when one corrects just the one variable. Then, $RB > (<)0$ indicates that correcting one mismeasured variable reduces (increases) bias in the parameter estimate of interest. We start with the simplest case, in which we assume no true correlation between our dependent and explanatory variables ($\beta = 0$), unit variance of the mismeasured and correctly measured explanatory variables ($\rho x_*^2 = \rho x^2 = 1$), and negative and equal correlations between measurement errors and true measures ($\alpha = \lambda = -0.5$), assumptions that appear broadly consistent with our data (on which, more below). Under these assumptions, we can use Equation (7) to compute RB when the econometrician has the option to correct measurement error only in the explanatory variable. The left-hand panel of Figure 1 depicts RB as a function of , the correlation between measurement errors. The right-hand panel shows the same function when the negative correlation between measurement errors and true values is stronger ($\alpha = \lambda = -0.8$). Both figures show that there exists a significant parameter space – the region where $RB < 0$ – over which the second-best approach of not correcting for measurement error outperforms incomplete correction.

Careful examination of the two panels reveals two important points. First, the second best approach can dominate no matter the sign of the correlation between the measurement errors. Second, even if the measurement errors are perfectly correlated, as in the case of $\pi = 1$, with $\alpha = \lambda = -0.8$ in the right-hand panel, correcting the one measurement error might still not be superior.

Figure 2 replicates the exercise but now for RB when the econometrician has the option to correct measurement error only in the dependent variable. The same basic pattern emerges but with one important difference: the parameter space over which the second best approach of not correcting for measurement error dominates is appreciably smaller. The key point is that correcting NCME in the dependent variable is more likely bias-reducing than is correcting measurement error in the key explanatory variable, contrary to the case of classical measurement error.

Figures 3 - 4 offer three-dimensional representations of the same phenomena, relaxing some key assumptions imposed in Figures 1 - 2. In Figure 3, we show the effects on RB of correcting only measurement error in land area while also varying the correlation between measurement error in area and true area. Figure 4 offers a similar comparison when the option is to correct measurement error in output rather than area. Beyond reinforcing the findings from Figures 1 and 2, these plots add important nuance, such as the importance of the correlation between the measurement error in and the true value of the independent variable. If $\alpha \geq 0$, it is almost always better to correct even just one source of NCME; the second best approach is rarely better, especially if one can correct NCME in the dependent variable. The main driver, however appears to be the correlation among measurement errors, with moderately positive levels, $0 < \pi < 1$, most typically associated with a second best result that ignoring NCME beats correcting just one of the two sources of NCME. And correcting for NCME in the dependent variable is, all else equal, more consequential than correcting for NCME in the explanatory variable.

4 Characterizing measurement error among wheat producers in Ethiopia

Following the above analytical framework, we now empirically investigate how NCME affect the SPR amongst wheat farmers in Ethiopia. Our sample consists of 504 farmers, randomly selected from 36 villages (*kebeles*) spanning 18 districts (*woredas*) of the Ethiopian wheat belt.[15] Farmers were interviewed in February and March 2014, a few months after the *meher* season harvest.[16] The survey instrument covered standard household characteristics, along with detailed data on one wheat plot randomly selected amongst all wheat plots for those farmers cultivating wheat on multiple plots.

In addition to farmers recall data, the originality of this data lies with the collection of accurate measurement for both wheat harvest and plot size, at the time of harvest, in November-December 2013. For this, we relied on experts from the Central Statistical Agency (CSA) of Ethiopia to accurately measure the plot area using the compass-and-rope method, and to measure harvest through a crop-cut exercise by which one random subplot (4 meters 4 meters) is harvested, and the resulting crop output weighed. Of the 504 sampled farmers,

[15] See Abate et al. (2015) for detailed discussion on the sampling design.

[16] *Meher* is the long (main) rainy and production season in Ethiopia.

crop-cut wheat production was successfully measured on 382 plots.[17]

Table 2 presents the summary statistics of main household and plot level characteristics. The first six rows provide alternative measures of plot size and production while the remaining rows report household and plot characteristics. In particular, we consider detailed plot level characteristics that might confound accurate measurement of plot size and production. As shown in the top of the table, there are important discrepancies between self-reported and objective measures of land area and production.

Table 2 shows that sampled plots have, on average, about nine corners, indicating that precise measurement of such plots using scientific methods can also be difficult. Nonetheless, the closure error is one percent, on average.[18] About 40 and 60 percent of the sample households used standard units for reporting their plot size and production, respectively. One can argue that the use of standard units (e.g., kg or ha) may introduce considerable errors since these measurements might not be commonly used in some rural areas. On the other hand, local measurement units are likely to vary between regions, villages and even farmers. For this reason, we control for these measurement units in our empirical characterizations of measurement errors.

In the remainder of this section, we use these data to explore the extent and nature of measurement errors for both production and plot size, and the consequences of those NCME on estimates of the SPR.

4.1 Self-reported plot size

Farmers self-reported estimates and traversing (also known as compass-and-rope) are the two conventional methods of measuring the surface of plot size. With the advent of new technologies, there are now alternative ways of measuring plot size i.e., GPS and remote sensing (see Carletto et al. (2015) and Carletto et al. (2016) for detailed discussion on these methods).

While measuring plot size through farmers self-report in household surveys is the least costly, the obtained measures can be subject to considerable measurement errors. First, self-reported plot sizes are commonly based on traditional units whose conversion factor varies across regions and hence can introduce meaningful errors. For example, farmers in Ethiopia

[17]Crop-cuts could not be measured for the remaining 122 plots for three reasons. First, seven of the farmers had no wheat plot during the 2013 *meher* season. Second, five farmers could not be identified by anyone in their respective *kebeles* at the time of the household survey. Third, the remaining 110 farmers harvested their wheat plots early before the crop-cut survey. There were no refusals. In Table A1, we show that these nonresponses are not systematic and hence uncorrelated with the household and plot characteristics.

[18]Closure error is the shortest line of unknown length and direction connecting the initial and final station of the polygon or traverse. When the closing error is larger than 3 percent of the perimeter of the polygon, repeating the measurement procedure is highly recommended (Casley and Kumar (1988)).

commonly measure and report land areas in oxen days, but that measure will necessarily vary with weather conditions, slope of the plot, drainage and texture of soils, animal breed and condition, etc. Second, farmers rounding of area units (rounding to a half day, a full day, or two days for instance) can generate meaningful error, the extent of which may vary proportionally with actual size. Last, farmers may strategically report lower landholding, a means to avoid state taxes or qualify for social programs.

Compass-and-rope is the most reliable method to accurately measure land area (Keita and Carfagna (2009); Fermont and Benson (2011); Carletto et al. (2015, 2016)). Compared to GPS-based area measurement, the compass-and-rope method is expensive, though some argue that the level of accuracy is worth the extra time and cost (Diskin (1997)), in particular because GPS-based area measurement may be imprecise for smaller plots (Schoning et al. (2005); Keita and Carfagna (2009); Fermont and Benson (2011); Carletto et al. (2015)).

Figure 5 reports the error in self-reported plot size (darker columns), by categories of plot size as measured by compass-and-rope. Farmers tend to over-estimate plot size by 150% on average, for the smaller plots. As plot size increase, the level of error decreases with farmers accurately reporting plot size (on average) for those plots ranging between 0.375 and 0.75 hectares. Larger plots tend to be under-estimated, however, by a factor of 25%, on average, for plots larger than 1 hectare. These differences are statistically significant at the bottom and top of the distribution, and non-significant towards the middle where differences are negligible.[19]

Next, we explore potential sources of mismeasurement in self-reported plot sizes. Figure 6, panel (a) plots the self-reported plot size onto actual plot size measured by the compass-and- rope (CR) method. Several observations are in order. First, it is clear that the majority of observations lie above the 45 line, indicating a clear tendency for farmers to over-estimate the size of their plot as compared to what is obtained from CR method. Second, coordinates based on the CR method appear more smoothly distributed than coordinates from self-reports which display significant heaping on values that correspond to the conversion factor between the common local unit and hectare (e.g. 1/2 *oxen day*=0.125 ha; 1 *oxen day*=0.25 ha). Accordingly, rounding appears to be a potentially important source of measurement error, with larger proportional consequences for smaller plots.

Other aspects of local context may also contribute to these errors. In Ethiopia, mis-measurement of plot sizes could emanate in part from the traditional measurement units of land itself. Oxen days (*timad*) is the most common unit of area measurement and can be subject to a wide range of errors, including biases from differences in length of working hours and traction capacity of oxen and in weather conditions, as well as plot characteristics (e.g.,

[19]Table A2 in the Appendix provides further details on this distribution.

slope, soil texture and drainage, etc.). Moreover, some of those same plot characteristics (along with shape, fertility, and ownership of the plot) and household characteristics can affect farmers estimation of plot size.

Column (1) of Table 3 reports correlates of measurement error in self-reported farm size, expressed as differences in logarithmic values of self-reported plot size and CR measurement ($ln(self - reported) - ln(CR)$).[20] To facilitate comparison of estimates, we restrict our sample to those plots for which crop-cuts are available.[21] Table 3 provides estimates for both unconditional relationships within *kebeles* (as per out theoretical framework) in odd-numbered columns, as well as estimations using a set of household- and plot-level controls in even numbered columns. In columns (1) and (2), results point to a negative correlation between measurement error and true plot size (as measured by compass-and-rope method). The magnitude of the correlation between measurement error and true farm size is larger than those reported by Carletto et al. (2013) and Carletto et al. (2015). One potential explanation for these differences could be related to the land area measure we are using in this paper. Carletto et al. (2013) and Carletto et al. (2015), as well as other previous studies investigating measurement error in plot size, use GPS-based land area measurement, which might be susceptible to some systematic measurement error (especially large, proportionately, for smaller plots), while we are using a method commonly considered as the most accurate method to estimate land area. While characteristics of the household head (such as age and gender) do not appear correlated with the error, we find evidence that farmers with larger total landholding tend to over-estimate the size of their individual plot, while those with lower fertility plots tend to have a more accurate assessment of the size of these plots.

4.2 Self-reported wheat production

Crop-cuts and farmer self-reported estimates are the two methods most often used to measure production in developing countries. The crop-cut method is based on harvesting one or multiple random subplots in each plot. The method involves randomly locating a sub-plot(s) prior to the harvest and the subplot(s) will be harvested by survey enumerators at the time of maximum crop maturity. Then, the harvest is processed (e.g., dried) and weighed. Total plot level production is then estimated by extrapolating the sampled crop production. One notes

[20]Due to the skewed distribution of some of our variables (e.g., plot size), we also re-estimate the above regressions that characterize measurement errors using the inverse hyperbolic sine transformation of our main variables of interest. The inverse hyperbolic sine transformation better handles extreme values than the commonly used log transformation (Burbidge et al. (1988)). In our case, it also overcomes potential expansion of the heterogeneity of the distribution of biases for values between 0 and 1 due to the log transformation. Results based on inverse hyperbolic sine transformation are however similar in sign, significance and magnitude as those presented throughout this paper.

[21]Results based on the full sample are almost identical and available upon request.

that this extrapolation may introduce errors due mainly to variations in the productivity of plot parts (e.g., interior vs. periphery or edge).[22] However, one may account for crop-cut distance to edges to minimize such problems, as we do in all presented estimates.

Crop-cuts are commonly regarded as the most reliable and unbiased method for estimating crop production (Fermont and Benson (2011)). However, obtaining production estimates through crop-cuts can be costly; it is both a time- and labor-intensive undertaking. Crop production estimates based on farmers self-report are therefore most common in agricultural surveys, including those incorporated in standard household surveys. Recently, high-resolution satellite imagery-based remote sensing techniques are also being used to estimate crop yield, with some promising results (e.g., Lobell et al. (2015); Gourlay et al. (2017)).

Figure 5 reports the error in self-reported production (lighter columns), by categories of plot size, as measured by compass and rope. Farmers tend to over-estimate (or at least over-report) their wheat harvest, although the bias appears much more pronounced for smaller plots. We find average over-estimation of 250% for plots smaller than 0.125 hectares, and 150% for plots between 0.125 and 0.250 hectares. These biases significantly decrease as plots become larger, albeit remaining positive and statistically significant.[23]

Panel (b) of Figure 6 further confirms farmers general tendency to over-estimate their production, with most observations lying above the 45 line. We do find clear evidence of heaping (as was the case in panel (a)), mainly due to the fact that farmers report their production estimates in bags of 50 to 100 kg each, such that several bags are collected for each plot, reducing the scope for heaping on a limited number of categories.

In Columns (3) and (4) of Table 3, we present correlates of measurement error in production, conditional on plot size. Results confirm that measurement error in self-reported production is negatively correlated with true measure of farm size (compass-and-rope). These correlations are much higher than those reported in Gourlay et al. (2017) and Desiere and Jolliffe (2018). As shown in our analytical framework, this correlation between measurement error in the outcome variable and true explanatory variable induces overestimation of the inverse relationship. We do not however uncover significant correlation between measurement error in production and the introduced household, farm and plot characteristics.

In Columns (5) and (6) of Table 3 we show that measurement error in self-reported production is strongly and negatively correlated with crop-cut production, suggesting the type of mean-reverting measurement error documented in earnings (Bound and Krueger (1991))

[22]For example, previous agronomic studies indicate that the periphery of a plot is often more productive than its interior (Little and Jackson (1978); Barchia and Cooper (1996); Ward et al. (2016)). More recently, Bevis and Barrett (2017) argue that this could be one explanation for the inverse size-productivity relationship. We explore that hypothesis below.

[23]See Appendix Table A3 for further details on this distribution.

and consumption (Gibson et al. (2015)). Following our analytical framework in Section 3, this may lead to underestimation of the inverse relationship. We find that measurement error in self-reported production is also correlated with soil quality.

4.3 Correlated measurement errors

Taken together, evidence thus far give strong support to the presence of non-classical measurement errors in both production and plot size, highlighting the strong negative relationship between both errors and plot size, and with actual production. Following the discussion in our analytical section, Columns (7) and (8) of Table 3 reports estimates of correlates between measurement error in production and measurement error in plot size, for a given plot size and other household and plot characteristics. The correlation is large in magnitude: a one percent increase in measurement error in plot size is associated with a 0.37 percent increase in measurement error in production.

To summarize the analysis of measurement errors in self-reported wheat production and plot size, we find empirical support for each of the four cases that analytical Section 3 noted may lead to biased estimates of the SPR: (i) NCME in self-reported production caused by negative correlation between measurement error and true (i.e., crop-cut) production; (ii) correlation between NCME in self-reported production and true plot size; (iii) NCME in self-reported plot size, caused by negative correlation between the bias in plot size and its true value; and (iv) positive correlation between measurement errors in self-reported production and plot size. Because these introduce several opposing biases simultaneously, the net effect of these measurement errors on the SPR parameter estimate of interest is ambiguous. Columns (1), (2), (3) and (4) of Table 5 summarizes the findings of Table 3, in line with our analytical framework.

5 Measurement errors and the estimated size – productivity relationship

This section presents estimates of the plot size – productivity relationship, under various combinations of measurement errors in plot size and production. For sake of comparability, we follow the commonly used OLS estimation presented in Equation (3):

$$Y^* - X^* = \beta_1 X^* + Z'\tau_1 + \varepsilon_1 \tag{8}$$

where production (Y^*) and plot size (X^*) are both expressed in logs and measured without systematic error. Z is the same vector of village, household and plot-level characteristics that we relied on in Table 3 and ε_1 is a mean zero error term. Equation (8) is our benchmark estimation, immune to NCME because we estimate it using crop-cut measurement for Y^* , and compass–and–rope method for X^*.

To investigate how measurement error in production and/or plot size affect the estimated β parameter, we run the following three alternative specifications, where Y and X are farmers (log-transformed) self-reported production and plot size, respectively:

$$Y - X^* = \beta_2 X^* + Z'\tau_2 + \varepsilon_2 \tag{9}$$

$$Y^* - X = \beta_3 X + Z'\tau_3 + \varepsilon_3 \tag{10}$$

$$Y - X = \beta_4 X + Z'\tau_4 + \varepsilon_4 \tag{11}$$

Empirical results associated with Equations (8) - (11) are presented in Table (4). These estimates are based on our unconditional regressions (odd-numbered columns) as well as controling for a full set of covariates (even-numbered columns). For sake of comparability across estimations, we limit the sample to those plots with crop-cut estimates, although similar results are obtained upon using full sample when feasible. Columns (1) and (2) reports benchmark estimates associated with Equation (8). Controling for a number of characteristics - and in particular soil type - leads to a negative estimated β_1 parameter, relatively small in magnitude, and statistically insignificant, pointing to the absence of clear relationship between plot size and productivity amongst wheat farmers in Ethiopia. Our proxy for the edge effect that Bevis and Barrett (2017) hypothesize could explain the inverse SPR, distance of crop-cut from the edge, is also statistically insignificant. In what follows, we compare the parameter estimates from Equations (9), (10) and (11) against the null benchmark of $\beta_1 = 0$.

Results in columns (3) and (4) (corresponding to Equation (9)) show a large, negative and statistically significant estimated β_2 parameter. The effect of measurement error in production therefore appears to substantially over-estimate the inverse size productivity relationship. But recall from the previous section that measurement error in plot size is correlated with crop-cut production as well as true plot size. In our analytical section, this situation corresponds to the combination of Cases 1 and 2 wherein measurement error may lead to over or under-estimation of the SPR depending on the relative magnitudes of δ and λ. In our data, the effect of production mismeasurement correlated with true plot size appears (Case 2) to dominate that of its correlation with true production (Case 1), as predicted. Overall, our results suggest that using self-reported production leads to substantial overestimation of the inverse relationship. This corroborates recent findings by

Gourlay et al. (2017) and Desiere and Jolliffe (2018), which similarly show that self-reported production measures can generate an estimated inverse SPR even when none exists.

Columns (5) and (6) present the estimation results of Equation (10) where production is correctly measured but measurement error in plot size is negatively correlated with true plot size as established in the previous section. In our analytical framework, we showed that such mismeasurement may have ambiguous consequence in estimating the relationship between plot size and productivity. Accordingly, the direction of bias associated with measurement in plot size depends on the relationship between the variance of self-reported and true area measurements as well as on the size (and sign) of the correlation between the measurement error and true area of land. Our descriptive statistics (Table 2) indicate that variance of the self-reported plot size is smaller than that of the true area, implying a negative correlation between measurement error in plot size and true land area measure. Thus, we may expect OLS estimation using self-reported plot size to overestimate the inverse relationship. This is supported by our results in Column (6), where the estimated β_3 parameter is large in magnitude, negative in sign and statistically significant. It is also consistent with the pattern reported by Carletto et al. (2015) but in contrast to the results in Carletto et al. (2013). The consequences of measurement error in plot size may therefore vary across contexts, sources and empirical features of measurement errors, again highlighting the cost of inaccurate land measurements.

Finally, Columns (7) and (8)) reports estimation results of Equation (11), where both plot size and production are measured with error, that is, using self-reported plot size and production. The estimated parameter β_4 suggests a significant inverse relationship between plot size and productivity. However, the magnitude of this inverse relationship is less than half the magnitude of those in Columns (2) and (3), implying that the two sources of measurement error have somewhat offsetting effects on the bias in the estimate of the SPR parameter. This is consistent with our analytical expression in Equation (7), showing that positive correlation of measurement errors in the dependent and independent variables may cancel out part of the bias due to measurement error in the dependent or independent variable(s). In such a situation, ignoring both types of measurement errors appear to bias the parameter of interest less than does controlling for either source of measurement error alone. This underscores the threat of partial correction of multiple, correlated, non-classical measurement errors and the 'second-best' inference result we emphasized earlier. Table 5 summarizes the key empirical relationships considering the alternative empirical scenarios.

Comparing the other estimates associated with the other explanatory variables in our regressions, we also observe some importance differences among Columns (4), (6) and (8) of Table 4, in other parameter estimates of interest. For example, measures of soil quality (soil fertility and soil color) are significantly associated with productivity when one uses correct measures of plot size and productivity, while this is not the case when using self-reported measures because soil quality indicators are correlated with the measurement errors in production and plot size (see Columns (1) and (3) of Table 3). This is consistent with previous arguments that omitted attributes, including unobservable soil quality, may contribute to the disputed inverse size-productivity relationship (Benjamin (1995); Assunção and Braido (2007)). Similarly, some plot characteristics (number of corners and crop-cut distance to the edge) appear to be significant only when we use crop-cut production along with self-reported plot size (Column (3) of Table 4). These spurious correlations between productivity and plot characteristics are potentially driven by farmers misperception of plot size and associated endogenous investments, consistent with the behavioral mechanisms hypothesized by Bevis and Barrett (2017). More generally, these pieces of evidence suggest that the implication of NCME in size and production may go beyond the inverse relationship and hence affect other relationships and inferences.

6 Concluding remarks

We analytically investigate correlated non-classical measurement errors (NCME) in both dependent and independent variables within a standard regression framework. We set up a generic analytical framework in which both dependent and explanatory variables can suffer from NCME and these errors can be correlated. We show that the signs and magnitude of resulting biases are analytically ambiguous and depend on several parameters characterizing measurement errors in these variables as well as the relationship under investigation. We also show that accounting for measurement error in only one of the variables may worsen the bias in estimated parameters.

We use this framework to shed further light on the longstanding policy debate about the relationship between plot size and agricultural productivity. This relationship has considerable implications for agricultural development policy: previous findings of an inverse relationship have often been invoked to support land reform programs. However, most previous empirical studies rely on farmer self-reports of output and area cultivated, with considerable room for NCME. And while recent studies have attempted to correct biases on either one of the variables (e.g., through GPS devices for area cultivated, or crop-cuts for production), none to our knowledge has investigated the relationship more generally, by addressing measurement issues on both sides of the equation, nor explored the implications for incomplete correction for correlated NCME.

We rely on a unique dataset combining self-reported and gold standard measurements of both agricultural output and area cultivated in Ethiopia. These data enable us to empirically validate our analytical and numerical results, showing that the inverse size-productivity relationship that we find in the self-reported data vanishes with more accurate measures. We also find that fixing measurement error in just one of the variables does not solve the problem and may effectively worsen bias in the parameter estimate of interest. These findings carry strong implications, not only for work that relies on conventional survey data, but also for a far broader array of studies that incompletely correct for measurement errors, which may prove inferior to a 'second best' approach that uses multiple variables measured with error. These findings are relevant to many economic applications and estimation problems involving multiple error-ridden variables. It may also be relevant to aggregate metrics constructed from multiple variables suffering from competing sources and patterns of bias.[24]

7 Acknowledgments

This paper benefited from comments by Marc Bellemare, Leah Bevis, Chris Boone, Brian Dillon, John Gibson, Kalle Hirvonen, Alemayehu Seyoum Taffesse and seminar participants at the African Development Bank and the Ethiopian Development Research Institute. Any remaining errors are the authors' sole responsibility.

[24] For example, Arthi et al. (2018) show that aggregating households labor use involves competing biases, over-reporting at the extensive margin of labor use and under-reporting at the intensive margin, with these errors ultimately cancelling each other out to minimize aggregate bias.

References

Abate, G. T., A. de Brauw, N. Minot, and T. Bernard (2015). The impact of the use of new technologies on farmers wheat yield in ethiopia: Evidence from a randomized controlled trial. Technical report, International Food Policy Research Institute (IFPRI).

Ali, N., K. Kahn, P. Patel, and J.-m. Gorelick (2009). Moving the learning forward: From incremental to transformational impact on empowering smallholder farmers and women.

Arthi, V., K. Beegle, J. De Weerdt, and A. Palacios-López (2018). Not your average job: Measuring farm labor in tanzania. *Journal of Development Economics 130*, 160–172.

Assunção, J. J. and L. H. Braido (2007). Testing household-specific explanations for the inverse productivity relationship. *American journal of agricultural economics 89*(4), 980–990.

Baird, S. and B. Özler (2012). Examining the reliability of self-reported data on school participation. *Journal of Development Economics 98*(1), 89–93.

Barchia, I. and J. Cooper (1996). Edge row effects on yields in an oat trial. *Australian journal of experimental agriculture 36*(4), 459–461.

Barrett, C. B. (1996). On price risk and the inverse farm size-productivity relationship. *Journal of Development Economics 51*(2), 193–215.

Barrett, C. B., M. F. Bellemare, and J. Y. Hou (2010). Reconsidering conventional explanations of the inverse productivity–size relationship. *World Development 38*(1), 88–97.

Beegle, K., J. De Weerdt, J. Friedman, and J. Gibson (2012). Methods of household consumption measurement through surveys: Experimental results from tanzania. *Journal of development Economics 98*(1), 3–18.

Benjamin, D. (1995). Can unobserved land quality explain the inverse productivity relationship? *Journal of Development Economics 46*(1), 51–84.

Bevis, L. E. and C. B. Barrett (2017). Close to the edge: High productivity at plot peripheries and the inverse size-productivity relationship.

Borjas, G. J. (1980). The relationship between wages and weekly hours of work: The role of division bias. *The Journal of Human Resources 15*(3), 409–423.

Bound, J., C. Brown, G. J. Duncan, and W. L. Rodgers (1994). Evidence on the validity of cross-sectional and longitudinal labor market data. *Journal of Labor Economics 12*(3), 345–368.

Bound, J., C. Brown, and N. Mathiowetz (2001). Measurement error in survey data. *Handbook of econometrics 5*, 3705–3843.

Bound, J. and A. B. Krueger (1991). The extent of measurement error in longitudinal earnings data: Do two wrongs make a right? *Journal of Labor Economics 9*(1), 1–24.

Burbidge, J. B., L. Magee, and A. L. Robb (1988). Alternative transformations to handle extreme values of the dependent variable. *Journal of the American Statistical Association 83*(401), 123–127.

Carletto, C., S. Gourlay, S. Murray, A. Zezza, et al. (2016). Cheaper, faster, and more than good enough: is gps the new gold standard in land area measurement? Technical report, The World Bank.

Carletto, C., S. Gourlay, and P. Winters (2015). From guesstimates to gpstimates: Land area measurement and implications for agricultural analysis. *Journal of African Economies 24*(5), 593–628.

Carletto, C., S. Savastano, and A. Zezza (2013). Fact or artifact: The impact of measurement errors on the farm size–productivity relationship. *Journal of Development Economics 103*, 254–261.

Casley, D. J. and K. Kumar (1988). *The collection, analysis and use of monitoring and evaluation data*. The World Bank.

Chao, L.-W., H. Szrek, K. Peltzer, S. Ramlagan, P. Fleming, R. Leite, J. Magerman, G. B. Ngwenya, N. S. Pereira, and J. Behrman (2012). A comparison of epi sampling, probability sampling, and compact segment sampling methods for micro and small enterprises. *Journal of development economics 98*(1), 94–107.

Das, J., J. Hammer, and C. Sánchez-Paramo (2012). The impact of recall periods on reported morbidity and health seeking behavior. *Journal of Development Economics 98*(1), 76–88.

David, I. P. (1978). Non-sampling errors in agricultural surveys-review, current findings and suggestions for future research.

De Groote, H. and O. Traoré (2005). The cost of accuracy in crop area estimation. *Agricultural Systems 84*(1), 21–38.

De Janvry, A. and E. Sadoulet (2000). Rural poverty in latin america: Determinants and exit paths. *Food Policy 25*(4), 389–409.

Desiere, S. and D. Jolliffe (2018). Land productivity and plot size: Is measurement error driving the inverse relationship? *Journal of Development Economics 130*, 84–98.

Diewert, W. E. and K. J. Fox (1999). Can measurement error explain the productivity paradox? *The Canadian Journal of Economics/Revue canadienne d'Economique 32*(2), 251–280.

Diskin, P. (1997). Agricultural productivity indicators measurement guide.

Feder, G. (1985). The relation between farm size and farm productivity: The role of family labor, supervision and credit constraints. *Journal of development economics 18*(2-3), 297–313.

Fermont, A. and T. Benson (2011). Estimating yield of food crops grown by smallholder farmers. *International Food Policy Research Institute, Washington DC*, 1–68.

French, E. (2004). The labor supply response to (mismeasured but) predictable wage changes. *The Review of Economics and Statistics 86*(2), 602–613.

Gibson, J., K. Beegle, J. De Weerdt, and J. Friedman (2015). What does variation in survey design reveal about the nature of measurement errors in household consumption? *Oxford Bulletin of Economics and Statistics 77*(3), 466–474.

Gibson, J. and B. Kim (2010). Non-classical measurement error in long-term retrospective recall surveys. *Oxford Bulletin of Economics and Statistics 72*(5), 687–695.

Gottschalk, P. and M. Huynh (2010). Are earnings inequality and mobility overstated? the impact of nonclassical measurement error. *The Review of Economics and Statistics 92*(2), 302–315.

Gourlay, S., T. Kilic, and D. Lobell (2017). Could the debate be over? errors in farmer-reported production and their implications for the inverse scale-productivity relationship in uganda.

Holden, S. T., M. Fisher, et al. (2013). Can area measurement error explain the inverse farm size productivity relationship. *Centre for Land Tenure Studies Working Paper 12*(13), 18–29.

Howard, J. A., A. Said, D. Molla, P. Diskin, and S. Bogale (1995). Towards increased domestic cereals production in ethiopia: Using a commodity systems approach to evaluate strategic constraints and opportunities.

Hyslop, D. R. and G. W. Imbens (2001). Bias from classical and other forms of measurement error. *Journal of Business & Economic Statistics 19*(4), 475–481.

Keita, N. and E. Carfagna (2009). Use of modern geo-positioning devices in agricultural censuses and surveys: Use of gps for crop area measurement. In *Bulletin of the International Statistical Institute, the 57th Session, 2009, Proceedings, Special Topics Contributed Paper Meetings (STCPM22), Durban*.

Kim, B. and G. Solon (2005). Implications of mean-reverting measurement error for longitudinal studies of wages and employment. *The Review of Economics and Statistics 87*(1), 193–196.

Larsen, A. F., D. D. Headey, and W. A. Masters (2017). *Misreporting month of birth: Implications for nutrition research*, Volume 1617. Intl Food Policy Res Inst.

Lipsey, R. G. and K. Lancaster (1956). The general theory of second best. *The Review of Economic Studies 24*(1), 11–32.

Little, T. M. H. and F. Jackson (1978). Agricultural experimentation: design and analysis. Technical report.

Lobell, D. B., D. Thau, C. Seifert, E. Engle, and B. Little (2015). A scalable satellite-based crop yield mapper. *Remote Sensing of Environment 164*, 324–333.

McKenzie, D. and M. Rosenzweig (2012). Preface for symposium on measurement and survey design.

Naschold, F. and C. B. Barrett (2011). Do short-term observed income changes overstate structural economic mobility? *Oxford Bulletin of Economics and Statistics 73*(5), 705–717.

Ozler, B. (2013). Economists have experiments figured out. whats next?(hint: Its measurement). *Development Impact Blog, http://blogs. worldbank. org/impactevaluations/economists-have-experiments-figured-out-what-s-next-hint-it-s-measurement*.

Schoning, P., J. Apuuli, E. Menyha, and E. Zake-Muwanga (2005). Handheld gps equipment for agricultural statistics surveys: Experiments on area-measurement and georeferencing of holdings done during fieldwork for the uganda pilot census of agriculture. *Statistics Norway Report 29*.

Sen, A. K. (1966). Peasants and dualism with or without surplus labor. *Journal of political Economy 74*(5), 425–450.

Ward, B., B. E. Roe, M. T. Batte, et al. (2016). Potential profitability of strip intercropping with corn and soybeans. *Journal of the ASFMRA*.

Tables and Figures

Figure 1: Relative bias in SPR, where $RB > (<)0$ implies correcting the mismeasured explanatory variable reduces (increases) bias when measurement error remains in the dependent variable.

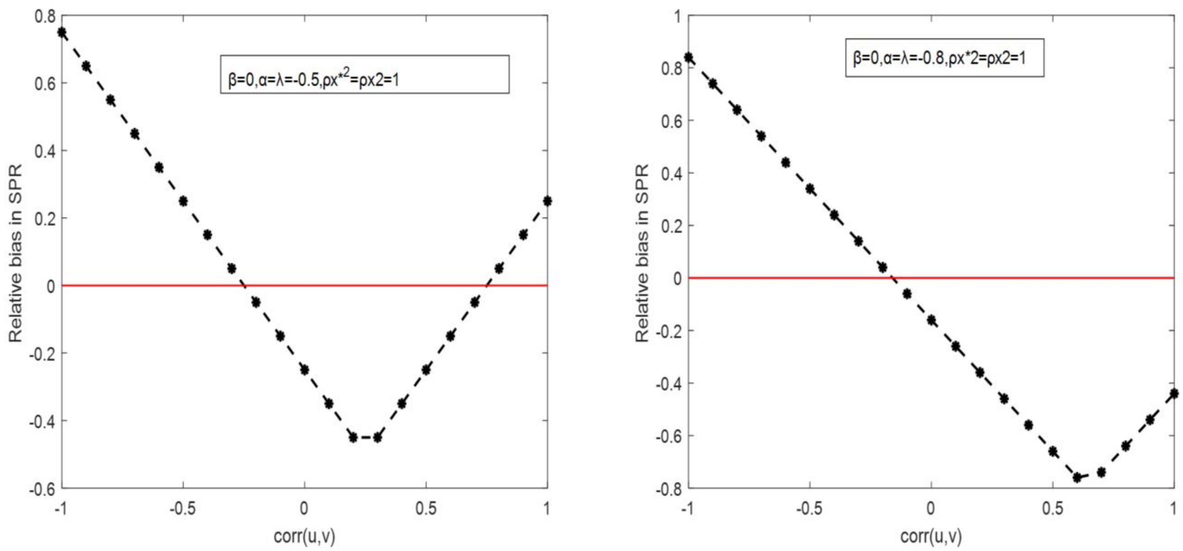

Figure 2: Relative bias in SPR, where $RB > (<)0$ implies correcting the mismeasured dependent variable reduces (increases) bias when measurement error remains in the explanatory variable.

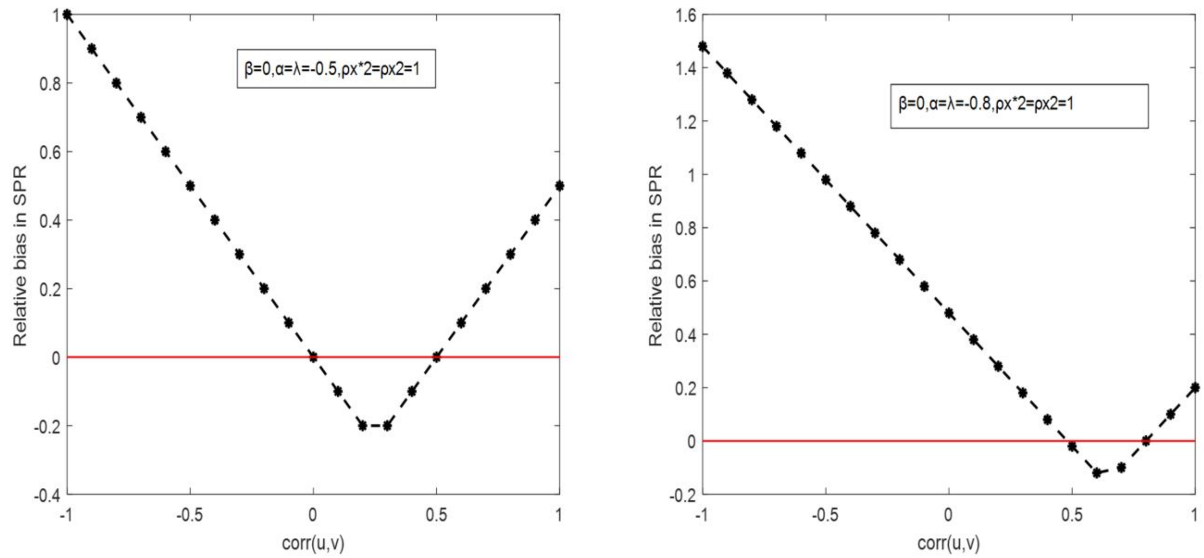

Figure 3: Relative bias in SPR, where $RB > (<)0$ implies correcting the mismeasured explanatory variable reduces (increases) bias when measurement error remains in the dependent variable.

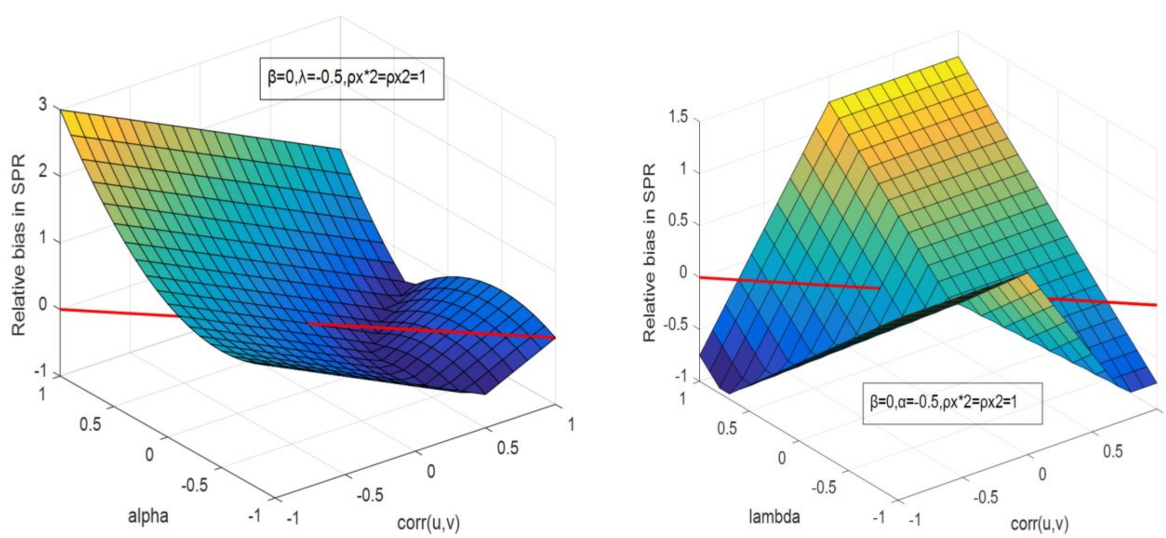

Figure 4: Relative bias in SPR, where $RB > (<) 0$ implies correcting the mismeasured dependent variable reduces (increases) bias when measurement error remains in the explanatory variable.

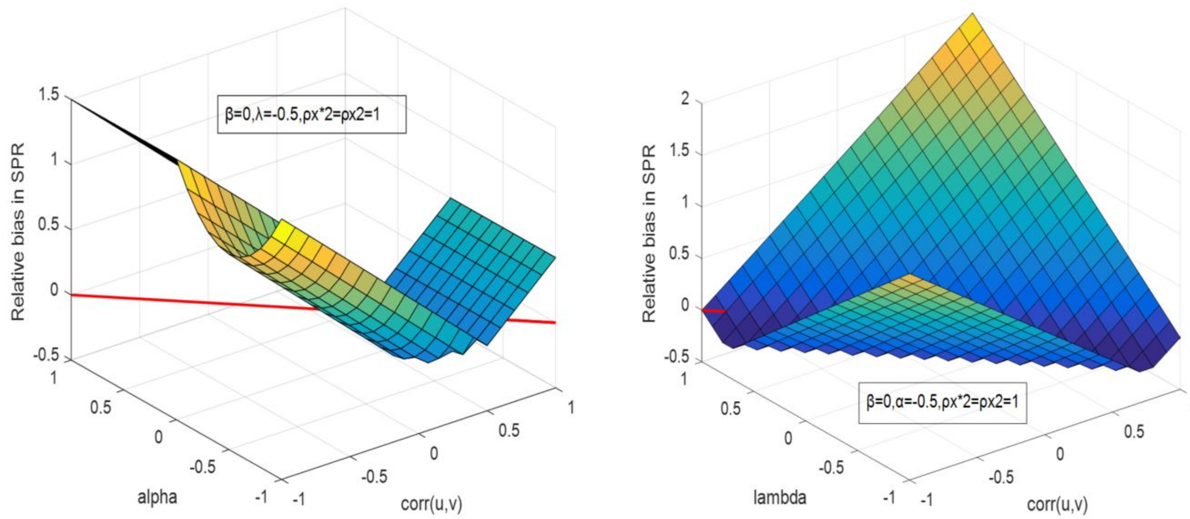

Figure 5: Measurement error in plot size and production, as a function of accurate plot size

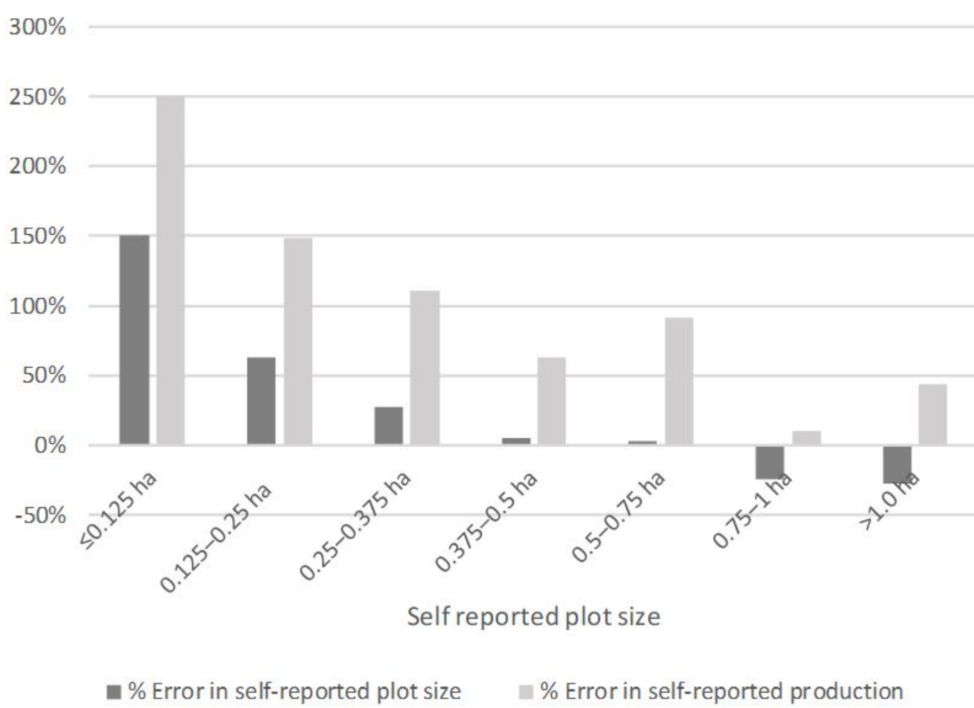

Figure 6: True vs self-reported plot size and production

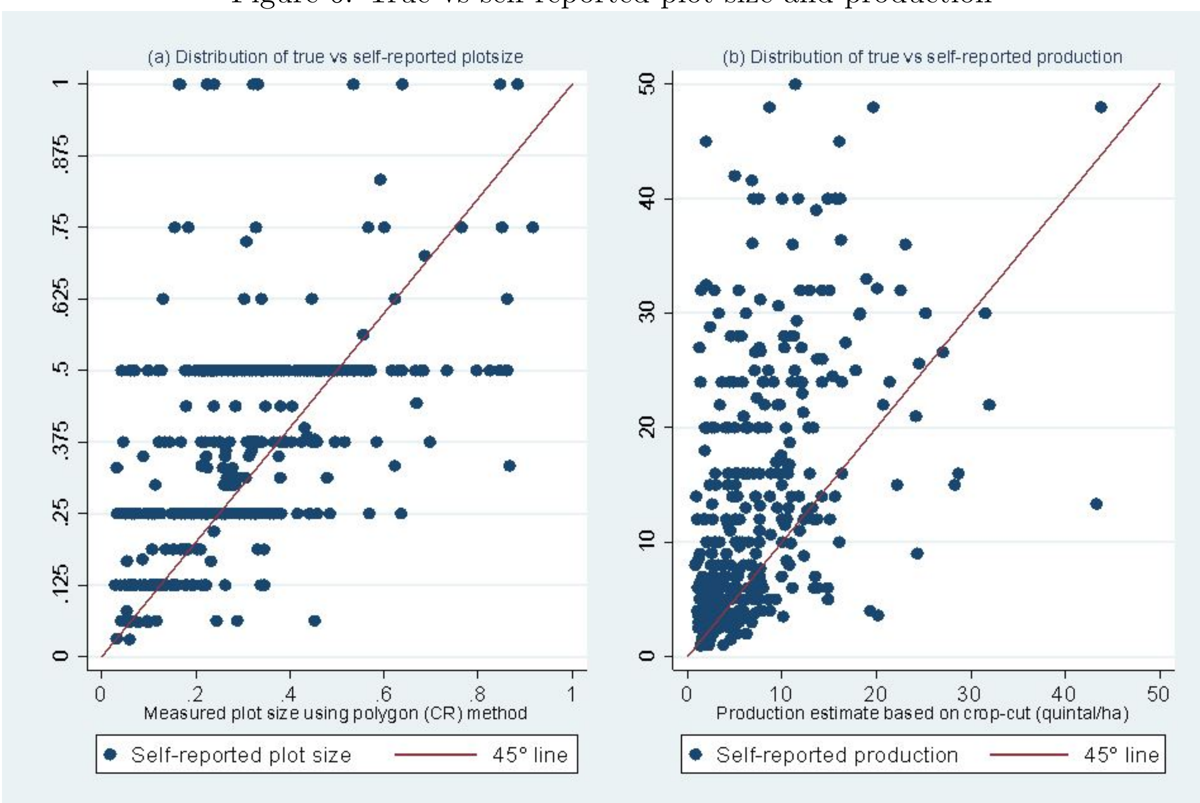

Table 1: Summary of Analytical Results

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
Errors in both	$*$	< 0	< 0	0	$\beta(1+\alpha)\phi - \alpha(1+\alpha)\phi - \lambda\phi$	Ambiguous
Correlated errors in both	$*$	< 0	< 0	> 0	$\beta(1+\alpha)\phi - \alpha(1+\alpha)\phi - \lambda\phi + \frac{\pi}{\rho x_*^2}$	Ambiguous

Notes: we rely on our data and empirical analysis to get an insight of the sign of the key parameters of interest.

$\phi = var(X^*)/var(X) = \frac{\rho x_*^2}{(1+\alpha)^2 \rho x_*^2 + \rho\lambda^2}$

$*$: value of these parameters can be zero or negative.

SPR stands for the size-productivity relationship and ISPR for inverse size-productivity relationship.

Table 2: Summary statistics

Variable	Description	Mean	Std. Dev.	Min	Max	Obs.
Area (SR)	Self-reported area size (ha)	0.42	0.36	0.03	4	488
Area (CR)	Measured area using compass and rope method	0.37	0.39	0.03	3.8	483
Production (SR)	Self-reported production for reference plot (qt.)	21.05	19.18	0.5	120	488
Production (CC)	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	96	488
Yield (CC)	Measured (production/area) using crop-cut, (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	77	488
Gender of HH head	Gender of the household head	0.86	0.34	0	1	488
HH size	Number of household members	6.79	2.39	1	16	488
Literacy of HH head	=1 if the household head is literate	0.64	0.48	0	1	488
No. of corners	Number of corners of the reference plot	8.74	4.88	4	23	484
Closure error	Closure error in plot area measurement	1.09	0.89	0.02	4.5	483
Area unit	=1 if farmers used ha for SR area measurement	0.39	0.49	0	1	488
Total owned area	Total farm land owned by sample farmers	2.31	2.14	0	20	488
Crop-cut to edge	Distance between crop-cut and closest plot edge (meters)	25.83	18.57	1.4	148	374
Production unit	=1 if farmers used kg for SR production measurement	0.59	0.49	0	1	488
Total wheat produced	Total wheat production during 2013/14 *meher*	46.64	75.26	0.95	755	488
High fertility	=1 if the fertility of the reference plot is high	0.44	0.49	0	1	488
Medium fertility	=1 if the fertility of the reference plot is medium	0.49	0.5	0	1	488
Poor fertility	=1 if the fertility of the reference plot is poor	0.07	0.26	0	1	488
Red soil	=1 if the color of the reference plot is red	0.26	0.44	0	1	488
Black soil	=1 if the color of the reference plot is black	0.54	0.49	0	1	488
Grey/sand soil	=1 if the color of the reference plot is grey or sandy	0.20	0.4	0	1	488
Distance to plot	Walking time between dwelling and the plot (in minutes)	30.98	9.94	0	120	488
Plot ownership	=1 if the reference plot owned by the HH	0.82	0.38	0	1	488

Table 3: Correlates of measurement errors

	ln area ratio		ln production ratio		ln production ratio		ln production ratio	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln plot size (CR)	–0.550***	–0.532***	–0.596***	–0.564***			–0.397***	–0.363***
	(0.045)	(0.045)	(0.073)	(0.079)			(0.109)	(0.110)
ln production (CC)					–0.656***	–0.671***		
					(0.055)	(0.054)		
ln (land area bias)							0.362**	0.377***
							(0.132)	(0.131)
Medium soil fertility		–0.096*		–0.060		–0.151**		–0.024
		(0.054)		(0.087)		(0.070)		(0.075)
High soil fertility		–0.237**		–0.074		–0.308***		0.015
		(0.089)		(0.106)		(0.082)		(0.121)
Black soil		–0.124		0.120		–0.035		0.166
		(0.089)		(0.127)		(0.101)		(0.125)
Grey or sandy soil		–0.004		0.258*		0.232		0.259*
		(0.080)		(0.150)		(0.145)		(0.147)
Distance to home		0.003		0.003		0.002		0.002
		(0.003)		(0.005)		(0.004)		(0.004)
Distance to edge		0.001		0.000		0.000		–0.000
		(0.002)		(0.003)		(0.003)		(0.003)
Number of corners		–0.002		–0.013		–0.002		–0.013
		(0.007)		(0.011)		(0.010)		(0.011)
Plot is owned by hh		–0.037		–0.103		–0.080		–0.089
		(0.072)		(0.093)		(0.080)		(0.095)
Observations	365	360	365	360	365	360	365	360
R2	0.46	0.52	0.50	0.51	0.61	0.63	0.53	0.55

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Note: Standard errors are clustered at *kebele* level and given in parentheses. High and red are reference categories for soil fertility and color, respectively. Other unreported controls in even numbers include the household head's age (linearly and squared), her gender, her education level, household size and total farm size. All estimates include *kebele*-level fixed effects.

Table 4: Plot size productivity relationship

	ln(crop-cut production/ compass-and-rope plot size)		ln(self-reported production/ compass-and-rope plot size)		ln(crop-cut production/ self-reported plot size)		ln(self-reported production/ self-reported plot size)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln plot size (CR)	–0.247***	–0.104	–0.679***	–0.668***				
	(0.060)	(0.062)	(0.079)	(0.072)				
ln plot size (self-reported)					–0.410***	–0.579***	–0.154**	–0.187**
					(0.067)	(0.079)	(0.062)	(0.077)
Medium soil fertility		–0.137***		–0.197***		–0.092		–0.063
		(0.049)		(0.066)		(0.084)		(0.063)
High soil fertility		–0.355***		–0.429***		–0.169		–0.215*
		(0.115)		(0.087)		(0.135)		(0.107)
Black soil		–0.234***		–0.114		–0.148		–0.068
		(0.069)		(0.091)		(0.091)		(0.088)
Grey or sandy soil		–0.043		0.215		0.009		0.124
		(0.074)		(0.150)		(0.125)		(0.119)
Distance to home		–0.001		0.002		–0.001		0.001
		(0.002)		(0.004)		(0.004)		(0.003)
Distance to edge		–0.001		–0.001		0.011***		0.002
		(0.002)		(0.003)		(0.003)		(0.002)
Number of corners		0.012		–0.001		0.072***		0.013
		(0.011)		(0.010)		(0.012)		(0.010)
Plot is owned by hh		0.036		–0.067		0.037		–0.024
		(0.064)		(0.083)		(0.095)		(0.072)
Observations	365	360	365	360	365	360	365	360
R2	0.10	0.56	0.58	0.60	0.40	0.53	0.46	0.48

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Note: Standard errors are clustered at *kebele* level and given in parentheses. High and red are reference categories for soil fertility and color, respectively. Other unreported controls in even numbers include the household head's age (linearly and squared), her gender, her education level, household size and total farm size. All estimates include *kebele*-level fixed effects.

Table 5: Summary of estimation results

Source of non-classical measurement error	Key empirically estimated parameters				Estimated SPR	Relative implication on the SPR
	δ	λ	α	π		
No error	NA	NA	NA	NA	0.104 (0.062)	Insignificant ISPR estimated
Error in production	-0.671*** (0.054)	-0.564*** (0.079)	NA	NA	-0.668*** (0.072)	Strongest ISPR estimated
Error in production	NA	NA	-0.532*** (0.045)	NA	-0.579*** (0.079)	Strong ISPR estimated
Error in both	-0.671*** (0.054)	-0.564*** (0.079)	-0.532*** (0.045)	> 0	-0.187*** (0.077)	Weakest ISPR estimated

Notes: we extracted the above estimates and standard errors (given in parenthesis) from our conditional regressions associated with Equations (8)-(11). NA refers that these parameters are either not relevant or not empirically estimated. SPR stands for the size-productivity relationship while ISPR represents the inverse size-productivity relationship.

Appendix

Table A1: Characterizing non-response in crop-cut survey

	Crop-cut (1=yes)
Age of hh head	0.001
	(0.008)
Age square	0.000
	(0.000)
Gender of HH head	-0.065
	(0.039)
Size of HH	-0.002
	(0.008)
Education of HH head	-0.023
	(0.039)
Total landholding size	0.006
	(0.006)
Soil fertility	
Medium	0.036
	(0.022)
Poor	-0.006
	(0.046)
Soil color	
Black	-0.003
	(0.039)
Grey or Sandy	-0.020
	(0.043)
Distance from home	0.000
	(0.001)
Own plot (1=yes)	0.018
	(0.035)
Constant	0.822* * *
	(0.167)
Observations	488
R-squared	0.692

*** p<0.01, ** p<0.05, * p<0.1

Kebele-level fixed effects included.

Std. errors clustered at *kebele* level in parenthesis.

Table A2: Discrepancy between compass and ropes and self-reported plot size

Plot-size group (CR)	# Obs.	Self reported (SR) (1)	Compass and rope (CR) (2)	Bias: $SR - CR$ Bias: (1) − (2) (3)	%Bias: (3)/(2) (4)	Difference in means p-values (5)
≤0.125 ha	70	0.20	0.08	0.12	150%	0.000
0.125 − 0.25 ha	132	0.31	0.19	0.12	63%	0.000
0.25 − 0.375 ha	125	0.38	0.30	0.08	27%	0.000
0.375 − 0.5 ha	74	0.46	0.44	0.02	5%	0.350
0.5 − 0.75 ha	46	0.60	0.58	0.02	3%	0.783
0.75 − 1 ha	12	0.64	0.85	-0.21	-25%	0.005
>1.0 ha	24	1.22	1.70	-0.48	-28%	0.019
Total	483	0.42	0.37	0.05	14%	0.002

Note: CR refers compass-and-rope, while SR stands for self-reported farm size.

Table A3: Discrepancy between crop-cut and self-reported production

Plot-size group (CR)	# Obs.	Self reported (SR) (1)	Crop-cut (CC) (2)	Bias: $SR - CC$ Bias: (1) − (2) (3)	 %Bias: (3)/(2) (4)	Difference in means p-values (5)
$\leq$0.125 ha	59	9.1	2.6	6.5	250%	0.000
0.125 − 0.25 ha	108	13.9	5.6	8.3	148%	0.000
0.25 − 0.375 ha	87	16.3	7.7	8.6	111%	0.000
0.375 − 0.5 ha	50	19.1	11.7	7.4	63%	0.000
0.5 − 0.75 ha	33	26.1	13.6	12.5	91%	0.000
0.75 − 1 ha	9	24.2	21.8	2.3	10%	0.800
>1.0 ha	19	46.5	32.2	14.3	44%	0.064
Total	365	17.5	8.9	8.5	95%	0.000

Note: CC refers crop-cut and SR stands for self-report, while CR stands for compass-and-rope measurement of farm-size.